К ЧИТАТЕЛЮ

Мы далеко не лучше знаем то, что видим ежедневно.

Л. Мерсье

Как правило, наибольшего успеха добивается тот, кто располагает лучшей информацией.

Б. Дизраэли

Эта книга посвящена информатике, наверное самой интересной и самой популярной из наук.

Определяя предмет информатики, можно сказать, что она занимается построением информационных объектов, теорией, систематическим анализом, проектированием и реализацией алгоритмов, программ и планов, а также процессами трансформации и передачи информации.

Что же такое информация? Как информатика связана с кибернетикой? В чём суть теории информации? На эти и многие другие вопросы можно найти ответы на страницах тома «Информатика».

В наши дни информатика «начинает и выигрывает»: с ней потихоньку сливается экономика, она простёрла свою руку в психологию и управление, поиск в библиотеках всё больше заменяется поиском в Интернете.

Сама книга, которую вы держите, создавалась с использованием современных информационных технологий не только в процессе вёрстки и печати, но и на стадии подготовки и

Под компьютерной грамотностью подразумевается набор навыков: применение компьютера для набора и редактирования текста, составления таблиц, просмотра и редактирования фотографий, программирования; использование Интернета для взаимодействия с другими людьми и передачи информации.

обсуждения текстов и иллюстраций. Большинство фотографий делалось с применением новейшей цифровой фототехники, рисунки сканировались на компьютере. А сами авторы общались при помощи электронной почты. Большинство из них лично ни разу не встречались и знакомы только по виртуальной переписке. Объём данной корреспонденции наверняка в несколько раз превысил том энциклопедии!

Во многом по причине своей современности информатика — одна из немногих наук, необходимость изучения которой не требует аргументов. Всё, что связано с современным обществом и современным человеком, самое передовое, интересное и перспективное так или иначе имеет отношение к сфере информатики.

Хранение и обработка информации в цифровую эпоху резко упрощают деятельность человека. Цифровое представление текстов, звука, фотографий и фильмов не только обеспечивает долговременность и относительную дешевизну хранения, но и делает нужную информацию легкодоступной. Многотомная энциклопедия может разместиться на нескольких компакт-дисках, а найти нужную статью удастся за доли секунды.

Том «Информатика» подробно рассказывает о методах цифрового представления разных видов информации, её хранения и обработки. Читатель узнает и о том, как кодировалась информация в докомпьютерные времена. Подобные знания правомерно считать необходимым элементом культуры человека нового тысячелетия.

Существует определённое общественное согласие относительно того, чем должен отличаться образованный человек, где бы и чему бы он ни учился. Это отличие усматривают не в той или иной совокупности полученных знаний, а в умении решать реальные задачи и проблемы.

Так, чтобы быть современным человеком, вероятно, уже недостаточно знать четыре арифметических действия, помнить наизусть таблицу умножения и уметь складывать дроби. Технологии требуют хотя бы минимальных навыков по использованию окружающих нас телефонов, телевизоров, микроволновых печей. В гостинице или дома вряд ли вам будут разъяснять правила пользования выключателем или учить, как звонить по телефону с кнопками вместо диска. Умные приборы всё больше входят в наш быт, и обращение с ними должно достичь нормального автоматизма.

Большинство жителей развитых стран умеют пользоваться кредитными банковскими карточками. В России этот метод платежа пока не очень распространён. И часть людей, особенно старшего поколения, с трудом осваивают его. К трудной, почти недоступной для наших дедушек и бабушек науке относится и Интернет.

Между тем большой процент молодёжи активно пользуется Всемирной сетью. Интернету в томе посвящена отдельная глава. В ней рассказывается не только о приёмах использования Интернета, но и о его устройстве, технологии передачи данных и т. п. Такие сведения позволят справиться с практическими трудностями, причём затратив минимум времени и усилий.

Считается бесспорным, что каждому из нас надо знать географию, математику, химию, биологию хотя бы в рамках школьной программы. Сегодня уже слышны голоса учёных, утверждающих, что информатика — такая же метадисциплина, как философия или математика. Даже более базовая,

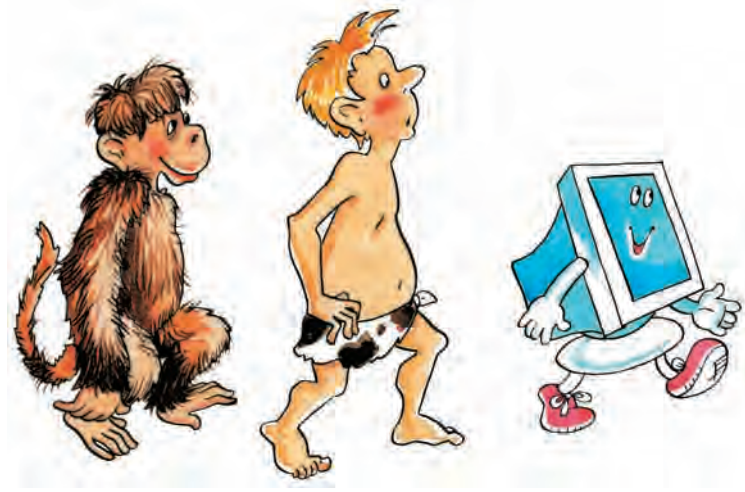
чем математика, так что правильнее сравнить её с арифметикой. Информатика совсем недавно появилась в школе как отдельный предмет. Ещё не утихли споры, с какого класса учить её и стоит ли вообще это делать. На страницах книги свою точку зрения высказывают авторы популярных школьных учебников по информатике.

Современный человек должен владеть компьютерной грамотностью, т. е. не просто знать, где находится кнопка включения домашней ЭВМ, но и использовать компьютер в работе и повседневной жизни. Ведь с помощью компьютера сегодня бухгалтер готовит баланс, лётчик управляет самолётом, музыкант занимается аранжировкой произведения, милиционер проверяет документы... Даже военный разведчик передаёт добытые данные в штаб при помощи ЭВМ, которую носит на поясе.

В томе «Информатика» много страниц посвящено компьютеру — универсальному изобретению человечества. Вы узнаете историю создания вычислительных машин с древних веков до наших дней. Молодая информатика имеет богатую родословную. Ряд предшествующих её формированию гениальных изобретений и открытий не потеряли своего значения и поныне.

Но информатика — это не только «про компьютер», так же как арифметика — не «про калькулятор», а музыка — «не про пианино».

Информационные технологии меняют общество и нас самих, но и они подвержены непрерывному усовершенствованию. Часть из них преобразуется, часть уходит в прошлое. Непрерывно появляются новые, внедряясь в доселе не охваченные сферы жизнедеятельности человека и в каких-то областях отстраняя нас от выполнения привычных функций. Громоздкие вычисления уже редко производят вручную, как редко кто



чертит тушью чертежи на ватмане. Недалеко время, когда информационные технологии будут преобладать повсеместно. Однако у всякого прогресса есть как положительные, так и отрицательные стороны. Ядерные исследования дали человечеству не только недорогой вид энергии, но и страшное оружие — атомную бомбу. Немалые проблемы может создать широкомасштабное использование ЭВМ и компьютерных технологий. На страницах тома идёт речь о том, с чем в этой связи мы столкнёмся в XXI веке, намечены и возможные пути преодоления отрицательных последствий.

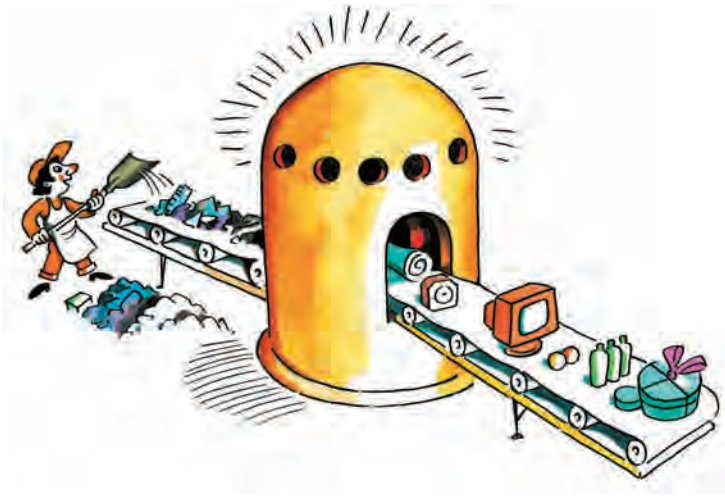
Даже объёмистый том слишком мал, чтобы вместить основные знания об информатике и её разнообразных аспектах. Так же как и о человеке, о науке можно рассказывать бесконечно, открывая всё новые грани. Кроме того, пока создаётся книга, возникают новые идеи, претерпевает изменения существующая техника, учёные совершают рывок вперёд. Однако фундаментальные знания, которые составляют основу тома, вряд ли устареют.

А о новейших достижениях в сфере ЭВМ скоро расскажет очередной том «Энциклопедии для детей», который будет называться «Компьютер».

ИНФОРМАЦИЯ И ЧЕЛОВЕК

КОМПЬЮТЕР — УНИВЕРСАЛЬНЫЙ ИНСТРУМЕНТ ДЛЯ ОБРАБОТКИ ИНФОРМАЦИИ

Эпоха глобальных информационных технологий только началась. Значительная часть создаваемой человечеством информации уже представлена в цифровом виде и может быть обработана на компьютере и передана в любую точку Земли по информационным сетям. Накопленная за предыдущие века информация (тексты, картинки, фильмы или аудиозаписи и т. п.) постепенно переводится в цифровую форму и поступает в те же информационные сети. Её легко сохранять, она не искажается при копировании. Хотя физические носители информации и не вечны, своевременное копирование, дублирование, хранение информации в разных узлах компьютерной сети позволяют не бояться локальных катастроф — поломок компьютеров, пожаров. Любые операции по обработке информации, от рутинных до творческих, проводятся с помощью компьютера. Ещё несколько лет назад игра в шахматы считалась исключительно творческим занятием, а сегодня становится ясно, что для компьютера она ненамного сложнее крестиков-ноликов.



Там, где должен стоять компьютер, он уже стоит или будет стоять в ближайшее время. Современный компьютер практически достиг предела возможностей. Текст, хранимый в компьютере, не в состоянии прочитать его владелец даже за всю свою жизнь. Математическая общественность всего мира производит текстов за год меньше, чем хранится в памяти современного компьютера.

Взрослые используют компьютеры для работы, дети — для учёбы. Деловое письмо составляется на компьютере, хранится в архиве, посылается адресату электронной почтой и лишь в исключительных случаях печатается на бумаге.

Книги уже сегодня создаются в цифровом виде. Такие «электронные книги» могут быть скопированы по компьютерной Сети и прочитаны с монитора персонального компьютера или специального карманного компьютера. Если сегодня человек привык идти за книгой в магазин или заказывать её по почте, то завтра, заплатив (с помощью компьютера) издательству, он получит право прочесть книгу с экрана и напечатать её в одном экземпляре на собственном принтере или сможет скопировать книгу с компьютера автора, взяв обязательство заплатить непосредственно ему (если книга понравится).

Современный компьютер хранит и бесконечное количество статических изображений — картинок и фотографий. Ещё распространены традиционные оптические фотоаппараты, регистрирующие изображение на светочувствительной плёнке, но уже активно используются цифровые фотокамеры, без плёнки (и вообще без движущихся частей) фиксирующие изображение сразу в цифровом виде и записывающие его в электронную память фотокамеры. Отснятые кадры можно переписать на персональный компьютер, посмотреть на экране, напечатать на принтере или отредактировать с помощью специальной программы и поместить в компьютерный семейный фотоальбом. Традиционный фотоаппарат (просуществовавший 150 лет, практически не меняясь) скоро станет таким же музейным экспонатом, как механические арифмометры и ламповые радиоприёмники.

Подобная судьба ждёт технологии киносъёмки и записи звука. Новые цифровые технологии здесь также теснят старые аналоговые. На переход

ОСНОВНЫЕ НАУЧНЫЕ И ТЕХНОЛОГИЧЕСКИЕ ДОСТИЖЕНИЯ XIX–XX ВЕКОВ

- Идея создания компьютера — универсального устройства для обработки информации.
- Идея его реализации на электронных схемах с использованием цифрового, двоичного кодирования любой информации.
- Создание стандартов представления любой информации в двоичном виде.
- Разработка языков программирования для записи алгоритмов обработки информации и автоматизация преобразования написанных человеком алгоритмов в двоичные программы.
- Создание общемировой Сети связанных между собой компьютеров.
- Технология микроэлектроники (реализация электронных схем на одном кристалле, при которой электронные схемы с миллионами и миллиардами переключательных элементов могут изготоваляться миллионными тиражами).
- Технология дискретизации (цифровое представление любых видов информации). Аналоговые, непрерывные методы записи, передачи и обработки информации заменяются цифровыми, искажение информации при хранении и передаче исключено, информация может считываться и копироваться неограниченное количество раз.
- Технология цифровых коммуникаций, которую используют все современные виды связи для создания общемировой Сети.

Функционально неграмотные люди существуют и сейчас. Например, человек, который пришёл на вокзал и не может разобраться в расписании поездов. Точно так же существует и электронная функциональная неграмотность. Социальный статус определяется не тем, насколько человек владеет современными информационными технологиями, а тем, насколько он способен адаптироваться к новым технологиям.

к цифровым видеокамерам и магнитофонам без движущихся частей потребуется несколько лет. Но пока компьютеры еще не могут хранить и передавать бесконечное количество фильмов, традиционные кассеты в домашнем видеомагнитофоне заменены на вращающийся диск.

Переход к новым информационным технологиям совершается в условиях конкуренции различных позиций, технологий, стандартов, протоколов программного обеспечения и оборудования. Формы представления и методы доступа к информации неизбежно будут отличаться большим разнообразием. Для того чтобы успешно ориентироваться в быстро меняющемся мире, базовые знания об информационных инфраструктурах и практические навыки работы с ними придётся поддерживать и обновлять в течение всей жизни.

Полноценный член общества нового тысячелетия должен каждодневно и эффективно взаимодействовать с компьютерными информационными сетями, от локального до планетарного уровня. Степень умения работать с информационными инфраструктурами станет во многом определять социальный статус индивидуума.

ЧТО ТАКОЕ ИНФОРМАЦИЯ

С конца XX столетия слово «информация» кажется столь широко употребляемым, сколь и неопределённым. В «Философском энциклопедическом словаре» даётся четыре его значения:

«Информация — общенаучное понятие, включающее обмен сведениями между людьми, человеком и автоматом, автоматом и автоматом; обмен сигналами в животном и растительном мире; передачу признаков от клетки к клетке, от организма к организму (генетическая информация)».

«Большой энциклопедический словарь», 2001 г.

• сообщение, осведомление о положении дел, сведения о чём-либо, передаваемые людьми;

• уменьшаемая, снимаемая неопределённость в результате получения сообщения;

• сообщение, неразрывно связанное с управлением, сигналы в единстве синтетической, семантической и прагматической характеристик;

• отражение разнообразия в любых объектах и процессах неживой и живой природы.

После того как американский учёный и инженер Клод Шеннон в 1948 г. опубликовал работу об основах математической теории свя-

АТОМЫ ИНФОРМАЦИИ

Информация, каким бы способом она не была обретена, всегда фиксирует отличия и изменения. «Поздно» имеет смысл только потому, что существует «рано», весна наступает после зимы. Мы мыслим противопоставлениями, и это фундаментальное свойство человеческого мышления. Ведь в природе не существует никакого конкретного момента, когда весна сменяет зиму. Это человек, не пассивно воспринимая мир, а извлекая из него информацию, делит непрерывную реальность на элементы с помощью подразумеваемых порогов, границ, критических точек, до которых было что-то одно, а после стало что-то другое. Любое «да» имеет смысл лишь в сопоставлении с «нет», любое «больше» — в сопоставлении с «меньше», «война» — с «миром», «дворцы» — с «хижинами»...

Итак, каждый раз, создавая информацию или передавая её, мы делаем цельное дискретным, вычленяем элементы. В некоторых случаях дискретность очевидна: включена сирена — не включена, подал руку — не подал. Такие оппозиции, устроенные по принципу «или — или», называются бинарными. В этих случаях одно состояние естественным, очевидным образом отделено от другого, различие между ними бросается в глаза (в случае с сиреной — в уши). С наступлением весны всё гораздо сложнее. В мире много процессов, не содержащих явных порогов. Эти пороги устанавливаем мы сами, вводя дискретность своей волей. «Совсем зарос, — говорит мама, — стричься пора». — «Не совсем, в глаза же не лезут, — отвечает сын. — Ещё так похожу». Здесь разные ситуации разграничиваются по количеству или степени признака. Допустим, степень «лохматости»: лысый — нулевая, коротко стриженный — первая, волосы немного отросли — вторая, лезут в глаза — третья, заплетаются в косичку — четвёртая и т. д. Так организуется градуальная оппозиция. Способ деления непрерывности бывает различным. Понятно, что, если выделять градуальные оппозиции произвольно, это не даст выполнить важнейшее условие передачи информации — точность. Представим себе, что не существует единого календаря, а каждый решает сам, ориентируясь на собственные ощущения, когда, по его мнению, пришла весна, а когда кончится сегодня и наступит завтра. Создание общих для всех и потому обеспечивающих точность передачи информации знаковых систем — задача, которую не раз приходилось решать человечеству. Без таких систем, представляющих непрерывное прерывным, невозможно представить современную жизнь, ведь это нотная грамота, денежная система, система счёта времени, географические координаты, шкала температур и многое другое.

Фердинанд де Соссюр.

Впервые фундаментальное значение противопоставления выяснил основатель современной лингвистики Фердинанд де Соссюр (1857—1913), изучавший язык как знаковую систему. Именно он показал, что ни один знак не имеет смысла сам по себе, он существует исключительно в оппозиции (противопоставлении) с другими знаками. Невозможен светофор с одним постоянно горящим фонарём, невозможен алфавит из одной буквы. Даже когда кажется, что знак всего один, в действительности это не так. Например, сирена — звуковой знак, означающий «внимание» или «опасность». Но если подумать, становится понятно, что и здесь есть противопоставление: наличие звука — отсутствие звука. Молчание сирены тоже значимо, оно выражает смысл «всё как обычно», «ничего особенного не происходит». Знаки, у которых внешняя форма состоит в отсутствии объекта, называются нулевыми. Примеры нулевого знака — пробел (пауза) между словами, негорящие лампочки на приборе, непротянутая в знак неуважения рука.

зи — теории информации, она нашла применение почти во всех областях науки, где играет роль передача информации в самом общем смысле этого слова. Примерно в те же годы появились первые компьютеры, ставшие позднее универсальным инструментом для обработки информации. Постепенно учение об информации проникло из кибернетики и информатики в физику, химию, биологию, философию, логику и т. д. Понятие «информация» превратилось в фундаментальное понятие науки и рассматривается под различными углами зрения в многочисленных трудах учёных. Предлагаются разнообразные его определения.

«Информация — это особая совокупность сведений, первичным источником которых является опыт».

А. А. Красовский, Г. С. Поспелов, 1961 г.

Как не запутаться в таком разнообразии? Имеем ли мы дело с одним или несколькими понятиями? И есть ли между ними какие-либо точки соприкосновения? Постараемся разобраться.

В первоначальном и наиболее узком смысле информация — атрибут разумных существ, людей. Это какие-либо сведения, данные, факты, которые, будучи получены из опыта, наблюдения или путём размышления, зафиксированы в материальной форме для сообщения другому существу или самому себе. При таком понимании любая информация с неизбежностью содержит два компонента — содержательный и материальный. То есть, во-первых, она должна быть понятна тем, для кого предназначена (содержательный компонент), во-вторых, представлена на том или ином физическом носителе, например на бумаге (материальный компонент).

Ясно, что одну и ту же информацию можно представить совершенно по-разному (например, на русском или на английском языке) и на разных носителях (в книге, на магнитной ленте).

Но если мы оставим в стороне смысловой компонент информации и в то же время абстрагируемся от любого конкретного материального её носителя, то всё ещё остаётся нечто, что может быть предметом исследования. Это — формальный компонент. Он включает в себя идею кодирования как таковую, т. е. безотносительно к смыслу кодируемого и независимо от физических свойств носителя информации. Формальный компонент принимает во внимание лишь количественные, частотно-вероятные и структурные свойства информации. Например, информация может быть вероятной и маловероятной; составной, т. е. содержать не один, а несколько независимых фактов.

На самом деле в итоговом кодировании «1» не содержится информации «У вас родился мальчик», но мы можем восстановить её на основе дополнительной информации «Сообщение из роддома».

Формальные свойства информации можно исследовать математическими методами — так родилась теория информации. Замечательно же то, что результаты данных исследований равно применимы к любым процессам и явлениям, в том числе и не относящимся к информационным в узком смысле этого слова, лишь бы только они обладали формальными свойствами информационных процессов. Поэтому совершенно логично, что сегодня слово «информация» имеет гораздо более широкое значение. Некоторые исследователи даже связывают понятие информации с разнообразием в природе и обществе, с упорядоченностью вообще.

КОЛИЧЕСТВО ИНФОРМАЦИИ

Чтобы научиться измерять какую-то величину, нужно, во-первых, выбрать единицу измерения, эталон, а во-вторых, описать процедуру сведения любой величины к эталону.

Если попытаться таким же образом определить меру информации, то выяснится, что сделать это можно различными способами.

Рассмотрим простейший случай. Допустим, информация — любая последовательность нулей и единиц. Тогда единицей информации логично считать одно место (разряд) в такой последовательности, а количеством информации — длину последовательности.

Например, сообщение из роддома: «У вас родился мальчик». Если кодировать каждую букву байтом, как это принято в компьютере, то всё сообщение займёт 21 байт, или 21·8 = = 168 бит. Можно несколько уменьшить это значение, ограничившись сообщением: «У вас сын» или просто «Сын», что даст соответственно 72 и 24 бит. Если же заранее договориться, что 1 означает «сын», а 0 — «дочь», то всё сообщение уложится в 1 бит.

Теперь понятен главный недостаток такого подхода: количество информации определяется скорее не содержанием самой информации, а используемым кодом. В зависимости от того, насколько оптимален или избыточен код, количество информации в одном и том же сообщении будет меньше либо больше.

Попробуем пойти другим путём. Прежде всего абстрагируемся от конкретного способа представления той или иной информации. Пусть ин-

ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ

Для хранения, преобразования и передачи информации важнее её различные представления и методы их использования, чем строгое определение.

Информация обязательно должна быть зашифрована в виде знаков, символов, кодов. Кодирование может производиться много раз и с различными целями: обеспечить передачу по тем или иным каналам связи, уменьшить нагрузку на какой-либо из них, сохранить конфиденциальность информации и т. д. Перевод на другой язык — то же кодирование, осуществляемое для того, чтобы информацию мог воспринять человек, незнакомый с языком, т. е. с исходным представлением информации.

В последнее время широкое распространение получила цифровая телефония. Любой человек может купить карточку с персональным кодом и позвонить в другой город или даже в другую страну. Во время разговора происходят следующие информационные процессы. Сначала информация шифруется в мозгу человека последовательно в виде образов, мыслей, речи. С помощью телефонного устройства колебания (речь) преобразуются в электрические импульсы. На компьютерных станциях-коммутаторах аналоговые электрические импульсы трансформируются в цифровой код, который обрабатывается при помощи специальных алгоритмов, чтобы ускорить его передачу. Потом код передаётся по каналам связи, ещё раз претерпевая кодирование, на этот раз для того, чтобы обеспечить его корректную доставку получателю. На принимающей станции происходит обратный процесс (раскодирование): цифровой код преобразуется в аналоговый электрический импульс, затем в телефоне — в звуковые колебания, которые наконец воспринимаются вторым собеседником.

В такой длинной цепочке изменялось представление информации, но целевое содержание (передаваемая информация) осталось неизменным.

В мире существуют многообразные виды информации. Предложим классификацию, основанную на восприятии человека.

1. Визуальная (зрительная) информация. Такую информацию, конечно, понимает не только человек, но и животные. В отличие от животных, в основном усваивающих всего лишь отражение внешнего мира, человек смог записать, сохранить визуальную информацию сначала в виде наскальных рисунков, потом иконописи, живописи, фотографии, видео. Прогресс идёт вперёд, XXI век — это время создания и широкого внедрения технологии записи и воспроизведения не плоского, а уже настоящего трёхмерного изображения, на которое можно посмотреть сбоку, сверху, его можно обойти и увидеть, что происходит с другой стороны.

Появились фотостудии, где уже предлагают сделать голографическую, псевдотрёхмерную фотографию родных, домашних питомцев.

2. Акустическая информация (речь, звук). Практически все животные в той или иной степени способны воспринимать звук: они предупреждают друг друга об опасности, устанавливают свой статус в сообществе, возможно, даже согласовывают действия при совместной охоте. Но только человек с помощью речи выражает чувства, мысли, до некоторой степени заменяет визуальное восприятие информации. Если рассказчик красноречиво, ярко, образно опишет что-то, то слушатель вообразит изображаемый предмет, словно он его видит. На компьютере речь бывает представлена как любой другой звук или музыка либо разбита на отдельные звуки — фонемы или даже целые слова.

3. Символьная информация (текст, письменность). С помощью письменности человек обрёл нечто подобное бессмертию: мысли, однажды занесённые на папирусные свитки, пергамен, бумагу, на различные цифровые носители, при условии правильного хранения будут существовать почти вечно.

Любая подобная классификация будет неполной — всегда есть такое представление информации, которое не подходит под составленную схему. К чему отнести биологическую информацию в ДНК? Или напряжения в земной коре, свидетельствующие о глубинных тектонических процессах? Или излучение далёких звёзд, по которому можно определить их вид, состав, процессы, идущие в них?

Поэтому классифицировать информацию необходимо в зависимости от использования. То ли это информация, воспринимаемая человеком, то ли информация, обрабатываемая компьютером, то ли многовековые знания человечества — везде необходима своя целевая классификация.

«Информация — это содержание воздействия, его величина, изменение в пространстве и времени, используемое как средство связи сложных систем».

Н. М. Амосов, 1963 г.

формация (сообщение) передаётся от отправителя А к получателю В. Независимо от реальной длины сообщения и способа его кодирования будем считать, что получатель узнает из него что-то новое относительно какого-то факта (степень неопределённости относительно данного факта уменьшилась). Поэтому допустим, что количество информации есть не что иное, как уменьшение степени неопределённости поступления информации, которое произошло в результате получения этой информации. Это ещё не определение количества информации. Сказанное только главная идея. Далее, уточняя вышеизложенное, шаг за шагом мы попытаемся вывести настоящее определение.

Шаг 1. *Неопределённость из двух равнозначных вариантов.* Например, родится сын или дочь, выпадет «орёл» или «решка». Если в результате получения информации неопределённость исчезает (нам сообщили, какой именно из двух вариантов имеет место), будем считать, что получена одна единица информации (единица измерения информации называется битом).

Шаг 2. *Неопределённость из N равнозначных вариантов.* Например, при игре в своеобразные кости. Интуитивно понятно, что в данном случае неопределённость выше. Но насколько, вот в чём вопрос! Допустим, $N = 8$. Тогда однозначный выбор из восьми вариантов можно сделать, получив ответ («да» или «нет») всего на три вопроса. Первым вопросом мы сокращаем число вариантов с восьми до четырёх, вторым — с четырёх до двух, третьим — до одного. После каждого ответа мы получаем 1 бит информации. Значит, всего 3 бит.

В более общем случае, когда $N = 2^k$, мы получаем k бит. То есть количество информации, необходимое для устранения неопределённости из нескольких равнозначных вариантов, равно $k = \log_2 N$.

Шаг 3. *Неопределённость из двух неравнозначных вариантов.* Пусть дано восемь равнозначных вариантов. Нам важно только знать, какой из них (первый или какой-либо другой) был принят. Происходит выбор из двух вариантов, причём неравнозначных. Допустим, ответ «Да, первый». Логично считать, что, как и на шаге 2, мы получили 3 бит информации, так как изначальная неопределённость из восьми вариантов свелась к одному варианту.

Предположим теперь, что ответ «Нет, не первый». Хотя неопределённость полностью и не устранена, тем не менее она уменьшилась с восьми вариантов до семи. Значит, сколько-то информации мы всё же получили, допустим, x бит.

Для устранения оставшейся неопределённости из семи вариантов нам нужно ещё $\log_2 7$ бит информации. Вместе они дают 3 бит информации, так как сокращают неопре-

«Понятие количества информации совершенно естественно связывается с классическим понятием статистической механики — энтропии. Как количество информации в системе есть мера организованности системы, точно так же энтропия есть мера неорганизованности системы. Одно равно другому, взятому с обратным знаком».

Н. Винер. «Кибернетика»

делённость с восьми до одного варианта. Поэтому $x + \log_2 7 = 3$ или $x = 3 - \log_2 7$.

Итак, сообщение «Да, первый» дало нам $3 = \log_2 8$ бит информации. Сообщение же «Нет, не первый», сократившее число вариантов с 8 до 7, дало $3 - \log_2 7 = \log_2(8/7)$ бит информации.

В общем случае, когда существует N равнозначных вариантов и есть сообщение о том, что имеет место один из k ($k < N$) вариантов, причём неизвестно, какой именно, мы получим $\log_2(N/k)$ бит информации.

Из теории вероятностей известно, что k/N — вероятность p данного события. Поэтому окончательно получаем $\log_2(1/p) = -\log_2 p$.

Количество информации в сообщении является в определённом смысле оптимальной длиной сообщения.

СОЗДАЮТ ЛИ ИНФОРМАЦИЮ КОМПЬЮТЕРЫ

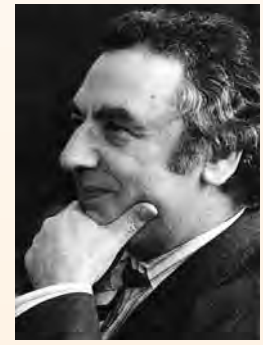
Математическая теория информации, как и все математические теории, кажется безупречной с точки зрения логики и к тому же прекрасно «работает» в самых разных областях. И тем не менее она не лишена внутреннего противоречия.

Рассмотрим такой парадокс. Предположим, кто-либо получает следующее сообщение: «Дважды два — четыре». Это весьма содержательное сообщение, часть таблицы умножения. И тем не менее с формальной точки зрения оно не содержит никакой информации. Действительно, вероятность того, что дважды два равно именно четырём, а не пяти, не шести и так далее, — единица: $\log_2 1 = 0$.

Следовательно, в сообщении содержится 0 бит информации. Точно так же с формальной точки зрения количество информации, заключённой в любой математической теореме, любом законе физики, любом корректном рассуждении или доказательстве, равно нулю.

С каждым годом компьютеры становятся всё «умнее и умнее». Они

СМЫСЛОВАЯ ИНФОРМАЦИЯ



Ю. А. Шрейдер.

В теории связи количество информации оценивается с точки зрения затрат на передачу сообщений. Для работы приёмно-передающих устройств первостепенен не смысл сигналов, а их статические свойства: разнообразие сигналов, избыточность и т. п. Для человека же существенное значение имеют именно смысл передаваемого сообщения и получаемые при этом знания, т. е. *смысловая информация*.

На основе смысловой информации вырабатывается и закрепляется в мозгу соответствие между будущими символами или образами на входе и необходимыми действиями. Высшие животные не могли бы существовать, если бы у них не было достаточного соответствия между сигналами ситуаций внешней среды (запах хищника, запах добычи и т. д.) и целесообразными в этих ситуациях действиями. Причём если у насекомых такое соответствие является в основном врождённым и поведение их формируется на базе врождённых инстинктов, то у млекопитающих оно пополняется в течение всей жизни, и получаемая информация через органы чувств носит во многом смысловой характер, идёт на пополнение запаса соответствий в их памяти. У человека смысловая информация идёт на пополнение соответствий не только между внешними сигналами и действиями, но и между понятиями.

Таким образом, общей чертой смысловой информации является то, что она изменяет запас сведений, запас соответствий у получателя информации. Первоначальный (и меняющийся в ходе поступления информации) запас соответствий можно представить себе как некоторый обобщённый словарь или справочник, который Ю. А. Шрейдер предложил называть *тезаурусом* (*греч*. «тезаурус» — «сокровище»; имеются в виду словари, где даны не только значения слов, но и связи между ними). В качестве меры количества смысловой информации естественно взять изменение тезауруса приёмника под действием поступившей информации.

Количество информации зависит от приёмника, от его тезауруса. Если тезаурус приёмника слишком беден, количество информации в сообщении оказывается вообще равным нулю. Например, в сообщении «Это учебник по высшей математике для первого курса вуза» трёхлетний ребёнок совсем ничего не поймёт, а школьник старших классов уже кое-что выделит. Максимальную же информацию извлечёт, очевидно, студент того курса, для которого учебник и предназначен. Любопытно, что по мере дальнейшего расширения тезауруса приёмника воспринимаемая информация начинает уменьшаться. Уже у студента второго курса при чтении учебника для первого курса изменение тезауруса будет меньше, а для человека, хорошо знающего высшую математику, этот учебник не несёт никакой информации.

В области искусства и литературы методы теории информации станут плодотворными лишь в том случае, когда будут разработаны способы оценки количества информации (не только статистической, но и семантической, смысловой), а также методы оценки изменения тезауруса человека (в том числе и его эмоционального тезауруса) под действием произведений искусства.

Информация (*лат.* informatio — «разъяснение», «изложение», «осведомлённость») — одно из наиболее общих понятий науки, обозначающее некоторые сведения, совокупность каких-либо специальных знаний и т. п.

«Энциклопедия кибернетики», 1974 г.

мастерски играют в шахматы, предсказывают погоду. Известен случай, когда с помощью компьютера была доказана теорема (см. дополнительный очерк «Задача о четырёх красках» в томе «Математика» «Энциклопедии для детей»). Кажется очевидным, что компьютеры производят информацию. Однако вся информация, содержащаяся, к примеру, в решении дифференциального уравнения, на самом деле уже имеется в его записи и начальных условиях, внесённых в машину оператором. Машина лишь «проявила» эту скрытую информацию, перевела её в другую, более пригодную для нас форму.

Аналогично, в речи кажется естественной фраза «электростанция производит электроэнергию». Но из курса физики известен закон сохранения энергии, и, строго говоря, электростанция не производит никакой энергии, а лишь преобразует её из одной формы в другую.

ТЕОРИЯ ИНФОРМАЦИИ

Передача сообщений является одним из важнейших видов деятельности человека. Изобретение телеграфа, телефона и Интернета изменило само существование человечества. Если проанализировать процесс передачи сообщений, то, несмотря на всё многообразие (от разговора до компьютерных команд), выделяют два основных этапа. Сначала сообщение переписывается в том виде, в котором его удобнее передать (хранить), затем оно дописывается, для того чтобы никакие помехи при передаче не повлияли на качество приёма. Первый этап называется кодированием источника сообщений, второй — помехоустойчивым кодированием.

Если взять в качестве примера обычный текст, набранный на компьютере, то буквам алфавита, как, впрочем, и знакам препинания, и цифрам, можно поставить в соответствие определённые наборы нулей и единиц. Эти наборы удобно рассматривать как слова, состоящие из комбинаций двух знаков — нуля и единицы. (Разряды таких двоичных слов принято называть битами.) Для записи всех букв русского языка, цифр, знаков препинания хватило бы и 7 бит (7 бит позволяют записать $2^7 = 128$ различных сообщений), используется же 8, с дополнительным проверочным битом, позволяющим обнаружить многие (но не все) ошибки, произошедшие при передаче (записи, хранении) данной буквы. Более впечатляющим является пример записи музыки на CD, когда даже царапины почти не влияют на качество воспроизведения (это стало возможным благодаря помехоустойчивому кодированию).

Математическая теория передачи сообщений, или теория информации, была сформулирована в работах знаменитого американского учёного и инженера Клода Элвуда Шеннона (1916—2001).

Шеннон широко использовал идеи и методы теории вероятностей и ввёл

важнейшее понятие энтропии как количества информации.

Один из наиболее популярных примеров кодирования источника — азбука (или код) Морзе, применяемая в телеграфии. Здесь с разными буквами сопоставляются различные двоичные слова (тире и точки заменены на нули и единицы), буквам с большой частотой появления (таким, как «А», «О») соответствуют короткие слова, а буквам с малой частотой употребления (как «Ы») — длинные. Это позволило Морзе сделать среднюю длину двоичных слов близкой к минимально возможной. А какова минимально возможная средняя длина? Следуя Шеннону, можно рассмотреть вероятностную постановку этой задачи. Пусть P_i — частота (вероятность) появления в словах i-й буквы алфавита. Для простоты предположим, что i-я буква алфавита появляется в словах независимо от предыдущих букв (конечно, в любом естественном языке, в частности русском, существуют определённые правила). Буквам алфавита ставятся в соответствие двоичные слова. Например, при соотнесении с буквой «Т» слово (0), «Ы» — слово (1), «Я» — слово (01) получится следующая неразрешимая ситуация. Из начала некоторой двоичной записи в виде (01) не известно, что же закодировано — слово «Я» или слово «ТЫ».

Наиболее простой способ обеспечить однозначное декодирование — это рассматривать только *префиксные коды,* когда никакое кодовое слово не является началом другого кодового слова. Приведённый выше пример несомненно не является префиксным кодом. Простейший префиксный код — это *равномерный код,* где все слова одной и той же длины. Чуть более сложный пример префиксного кода с неравной длиной кодовых слов:

$$A \to 0;\ B \to 10;\ C \to 110;\ D \to 111.$$

Чтобы декодировать двоичную последовательность, составленную из слов этого кода, следует начать с первого символа и оставить только те кодовые слова, для которых он является началом. Если таких слов несколько, то придётся рассмотреть первые два символа и сократить список возможных слов, сохранив опять лишь слова, начинающиеся с данных символов, и т. д. В конце получится единственное кодовое слово, совпадающее с несколькими первыми битами (не более трёх в примере), а все остальные кодовые слова будут на этих позициях отличаться от принятых битов. Теперь эти биты «вычёркиваются», и процедура продолжается дальше. Так, для двоичной последовательности 100111 получится:

$$100111 = (10)\,0111 = (B)(0)111 = (B)(A)(111) = (B)(A)(D) = BAD.$$

Клод Элвуд Шеннон.

Как сравнивать различные кодирования (однозначные) по эффективности, что выбрать в качестве меры эффективности?

Шеннон предложил рассматривать среднюю длину кодирования (среднюю длину кодовых слов) $E(L) = p_1 L_1 + ... + p_m L_m$, где L_i — длина i-го слова, и считать кодирование оптимальным с минимально возможной средней длиной. Ясно, что неважно, как называются буквы *А, Б*... или *X, Y*..., а важно только, сколько букв и вероятности P_i их появления, т. е. всё определяется распределением вероятностей $P = \{p_1, ..., p_m\}$. Множество букв (алфавит) с заданным распределением вероятностей называют источником сообщений.

Для алфавита, состоящего из четырёх букв (*A, B, C, D*), и распределения вероятностей: $P(A) = 1/2$, $P(B) = 1/4$, $P(C) = P(D) = 1/8$. Естественно, что $P(A) + P(B) + P(C) + P(D) = 1/2 + 1/4 + 1/8 + 1/8 = 1$. Средняя длина кода, описанного в предыдущем примере, $E(L) = p_1 L_1 + ... + p_m L_m$ равна 7/4, что меньше 2 бит, необходимых при равномерном кодировании.

Распределению вероятностей $P = \{p_1, ..., p_m\}$ сопоставляется число $H(P) = p_1 \log_2(1/p_1) + ... + p_m \log_2(1/p_m)$, называемое энтропией источника и служащее мерой неопределённости. Чаще всего в этой формуле используются логарифмы по основанию 2, и тогда количество информации, необходимой для восстановления значения источника (т. е. устранения неопределённости), измеряется в битах. Один бит соответствует количеству информации, получаемой из ответа на вопрос с двумя равновероятными ответами.

Первая теорема Шеннона — теорема кодирования для канала без шума утверждает, что для любого кодирования (с однозначным декодированием) его средняя длина $E(L)$ не меньше энтропии $H(P)$, а с другой стороны, существуют кодирования со средней длиной, сколь угодно близкой к $H(P)$. Тем самым теорема означает, что сообщения от источника могут быть закодированы в двоичной форме со средней длиной представления, незначительно большей, чем энтропия источника, и что «экономнее» энтропии никакой источник сообщений представить нельзя. Величину $r(L) = E(L) - H(P)$ называют избыточностью кодирования L, тем самым $0 < (L) < 1$ для однобуквенных кодировок.

Для побуквенных кодировок нельзя утверждать, что избыточность кодирования можно сделать сколь угодно малой, а гарантировать лишь то, что избыточность кодирования можно сделать меньше 1. Для получения малой избыточности кодирования переходят от побуквенных кодировок к блочным, когда кодируются блоки из n букв алфавита. Пусть для простоты анализа буквы источника появляются независимо. Тогда вероятность появления блока $A = (a_1, \dots, a_n)$ равна $P(A) = p(a_1) \cdot \dots \cdot p(a_n)$, а энтропия ансамбля блоков букв равна $H(P^n) = nH(P)$. Следовательно, для блочных кодирований средняя длина кода на букву (текста) не меньше $H(P)$, а с другой стороны, она меньше, чем

$$\frac{1}{n}H((P^n) + 1) = H(P) + \frac{1}{n}.$$

Тем самым при увеличении длины кодируемых блоков средняя длина кодирования может стать очень близкой к энтропии источника, а избыточность кодирования — достаточно малой. Алгоритмы кодирования также называют алгоритмами сжатия информации, поскольку они позволяют представлять данные более экономно, устраняя имеющуюся в данных естественную избыточность.

До сих пор рассматривался вопрос о кодировании источника информации (сообщений), т. е. экономном представлении информации в удобном виде (например, двоичном), предполагая, что закодированная информация не подвергается каким-то искажениям — шума нет. Это и объясняет название первой теоремы Шеннона — «теорема кодирования для канала без шума». Но, конечно, в большинстве реальных ситуаций шум есть, иногда столь незначительный, что им можно пренебречь, но весьма часто и такой, который приводит к изменению выбранной нами формы представления (кодирования). Тогда уже нельзя гарантировать однозначность декодирования. Или «декодирование» будет однозначным, но приведёт к неверному (не тому, которое посылалось) сообщению.

Так, для слова BAD изменение первого бита его двоичного представления (одиночная ошибка) 100111 приведёт к 000111 = = (0)(0)(0)(111) = AAAD.

Для формального описания задачи исправления ошибок Шеннон ввёл понятие *канал передачи информации*. У канала есть вход, куда поступают символы из некоторого алфавита X (другие символы канал просто не воспринимает). Канал рассматривается как «чёрный ящик», т. е. неважно (или неизвестно), что происходит внутри канала, сообщается только результат — выход y (y принадлежит некоторому алфавиту Y). Следовательно, канал задаётся некоторым законом того, что на выходе канала будет y, если на входе было x (*условные распределения вероятностей* $P(Y = y \mid X = x)$). Одним из важнейших примеров является *двоичный симметричный канал* (ДСК), где вход и выход двоичные ($X = Y = \{0,1\}$) и вероятность того, что $y = x$ (вероятность безошибочной передачи двоичного символа), равна q, т. е. вероятность ошибки $p = 1 - q$. Простейший способ надёжно передавать сообщения по такому каналу — повторять символы по нескольку раз (как во время телефонного разговора, если собеседник не расслышал слово).

Пусть $p = 0{,}1$ и каждый двоичный символ повторяется $2m + 1$ раз (например, семь). Тогда приёмник принимает решение, что передавался тот символ, который чаще (не менее чем $4 = m + 1$ раз) появлялся на выходе

КОДЫ ХЭММИНГА

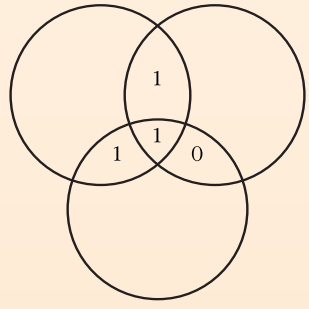
История кодов, исправляющих ошибки, началась чуть более полувека назад, когда молодой американский учёный Ричард Уэсли Хэмминг решал задачи на одной из первых вычислительных машин. Компьютеры того времени были ненадёжны и часто ошибались. Применяемый уже тогда простой метод обнаружения ошибок (проверка на чётность, каждому машинному слову из 0 и 1 приписывается дополнительный избыточный бит, равный 0, если в слове чётное число единиц, и 1 — если нечётное) позволял обслуживающему персоналу перезапустить вычислительный процесс. Но по выходным дням, когда у Хэмминга было основное машинное время, компьютер никто не обслуживал, и в результате много вычислений проводилось напрасно. Тогда ему пришла мысль, что если коды могут обнаруживать ошибки автоматически, то должны быть и коды, способные исправлять ошибки.

Так появился теперь хорошо известный [7,4]-код Хэмминга, исправляющий одиночные ошибки. Этот код позволяет представить (закодировать) 4 бит информации с помощью 7 бит таким образом, что любая одиночная ошибка в 7-битном представлении будет исправлена, в результате чего удастся правильно восстановить исходные 4 бит.

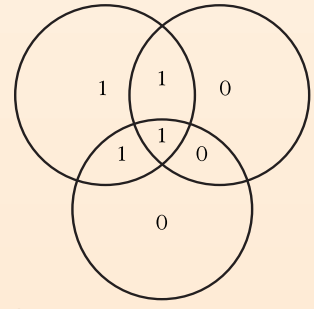
Так как 4 бит информации представляют $2^4 = 16$ различных сообщений, то код Хэмминга состоит из 16 так называемых кодовых слов длины 7. Свойство исправления одиночных ошибок означает, что искажение любого бита в каком-то кодовом слове и любого (или ни одного) бита в другом не может привести к одному и тому же 7-битному слову, имеется множество («пространство») из $2^7 = 128$ 7-битных слов, и в нём подмножество (код) из 16 слов. Вокруг каждого кодового слова есть «облако» («шар») из 7 слов, полученных в результате одиночных ошибок, «облака» («шары») не должны пересекаться. Последнее условие исправления одиночных ошибок равносильно тому, что любые два кодовых слова должны отличаться более чем в двух позициях. Так как в каждом шаре 8 точек (7 плюс центр), а всего слов в пространстве 128, то можно разместить не более чем 128/8 = 16 непересекающихся шаров, т. е. *код Хэмминга оптимален*.

Код Хэмминга задаётся как множество решений однородной (т. е. правая часть равна 0) системы из четырёх линейных уравнений с семью неизвестными $x_1, \ldots, x_7$, у которой столбцы матрицы коэффициентов пробегают все 7 ненулевых столбцов из 0 и 1, а равенство нулю понимается как равенство по модулю два (т. е. результат арифметических действий, как с обычными целыми числами, должен быть чётным числом).

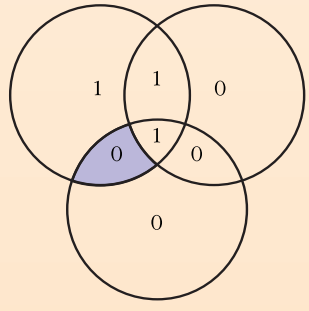
Пусть $x = (x_1, \ldots, x_7)$ — кодовое слово, а $y = (y_1, \ldots, y_7)$ — принятое (из канала) слово, отличающееся от x не более чем в одной позиции. *Декодирование*, т. е. восстановление x по y, происходит следующим образом. Сначала вычисляется *синдром* $s(y) = (s_1, \ldots, s_4)$ слова y, как значения правых частей четырёх уравнений, в которые подставлены y_i (вместо x_i). Затем ищется такая позиция i, где соответствующий столбец коэффициентов уравнений совпадает с синдромом и исправляется (т. е.



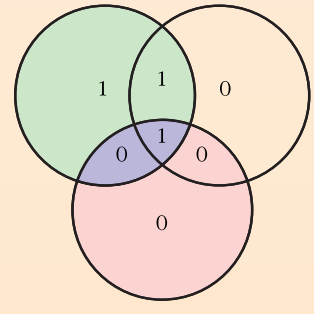
Биты инфомации ханятся в четыёх областях, обазованных пеесечением тёх окужностей.



Записываются ти бита чётности по павилу : число единиц каждой окужности должно быть чётным.



Возникла ошибка устанимого типа (выделено цветом).



Ошибка легко обнаужится.

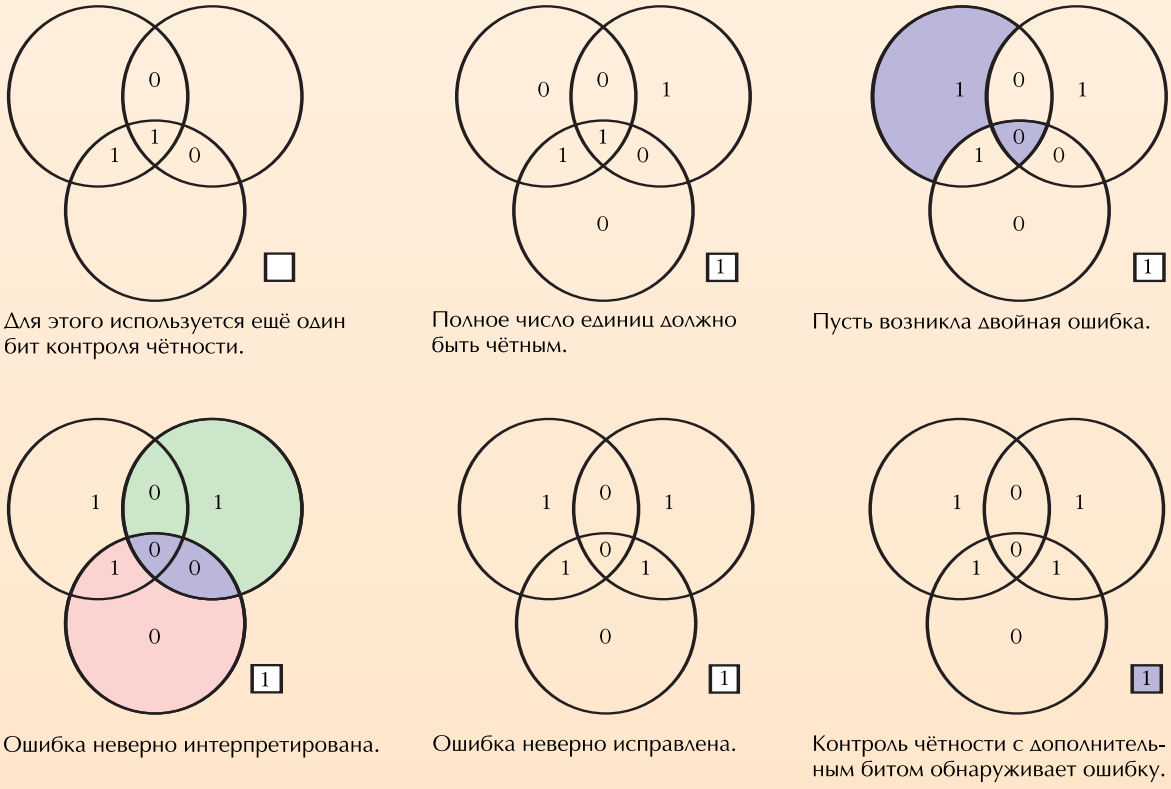
Устранение одиночных ошибок (диаграммы Венна).

инвертируется) i-я позиция в слове y. Если же синдром равен 0, то одиночной ошибки не было: $x = y$.

Эта конструкция легко переносится на случай, когда слова имеют длину $n = 2^r - 1$, а число кодируемых битов $k = n - r$. Получаемый код Хэмминга позволяет исправлять одиночные ошибки и оптимален. Для длин кода, отличных от $2^r - 1$, построение оптимального двоичного кода с исправлением одиночных ошибок является сложной математической задачей, не решённой до конца. Известно, что максимальное число слов в коде Хэмминга равно: 16 для $n = 7$ (1950 г.); 20 для $n = 8$ и 40 для $n = 9$ (1980 г.); 72 для $n = 10$ (1999 г.), для $n = 11$ ответ неизвестен (однако он существует для n от 12 до 15).

Ещё более сложным и интересным является вопрос о построении кодов, исправляющих многократные ошибки. В настоящее время существует несколько классов подобных кодов, широко и успешно применяемых на практике: БЧХ, Рида — Соломона, Рида — Маллера и др. Наиболее популярны коды Рида — Соломона, которые исправляют многократные ошибки блоковой структуры (например, в словах из 128 байт исправляются любые пять ошибочных байтов). Эти коды являются оптимальными, так как для исправления ошибочных блоков им требуется избыточность в виде $r = 2t$ проверочных блоков. В частности, коды Рида — Соломона используются в CD-проигрывателях и обеспечивают высокое качество воспроизведения музыки. Ещё пример применения кодов с исправлением ошибок: передача данных как из космоса, так и по обычному телефонному каналу (например, при сеансовом доступе в Интернет), мобильная связь, хранение данных в компьютере и многое другое.

Ричард Уэсли Хэмминг.



Обнаружение двойной ошибки.

«В наш век возрастающей дифференциации человеческих знаний Клод Шеннон является исключительным примером соединения глубины отвлечённой математической мысли с широким и в то же время совершенно конкретным пониманием больших проблем техники. Его в равной мере можно считать одним из первых математиков и одним из первых инженеров последних десятилетий».

А. Н. Колмогоров

канала. При этом вероятность ошибки декодирования стремится к нулю с ростом m, правда довольно медленно.

Основной недостаток данного метода заключается в том, что его КПД = $= 1/(2m + 1)$, называемый скоростью передачи, также стремится к нулю. На первый взгляд кажется, будто этот недостаток «врождённый» и требование стремления к нулю вероятности ошибки декодирования приводит к тому, что у соответствующих методов передачи информации скорость передачи будет стремиться к нулю.

Вторая теорема Шеннона говорит, что это не так и для любой величины R, меньшей, чем *пропускная способность канала* C, можно передавать сообщения со скоростью передачи R и вероятностью ошибки меньше любой заранее заданной сколь угодно малой величины. Для двоичного симметричного канала пропускная способность $C = p \log_2 p + q \log_2 q$, и она больше 0 для любой вероятности ошибки, отличной от 1/2. В частности, пропускная способность ДСК с вероятностью ошибки $p = 0{,}1$ равна примерно 1/2. Это означает, что существуют *коды, исправляющие ошибки*, кодирующие блок из k бит информации в кодовый блок из примерно $2k$ бит, передаваемый по ДСК. Поэтому даже в результате ошибок в канале вероятность «перепутать» один кодовый блок с другим очень мала.

Шеннон доказал свою теорему, не предъявив такие коды, что было бы естественным, а показав, что случайно выбранный код почти обязательно обладает требуемым свойством.

И до сих пор, несмотря на интенсивные исследования, параметры явно построенных кодов остались хуже, чем у случайных кодов. Тем не менее коды, исправляющие ошибки, нашли очень широкое применение в технике — от надёжной передачи данных с межпланетных станций до бытовых CD-проигрывателей.

КЛОД ЭЛВУД ШЕННОН

Клод Элвуд Шеннон.

Теория информации является одним из редких примеров научных дисциплин, о которых точно известно время их создания. Данная наука была описана в эпохальной работе Клода Элвуда Шеннона «Математическая теория связи» (1948 г.). В этой работе Шеннон не только ввёл новые понятия, такие, как «количество информации», «энтропия», «канал связи» и его пропускная способность, и показал теоретические пределы передачи информации, но и во многом предопределил дальнейшее развитие теории информации на несколько десятилетий вперёд.

Клод Элвуд Шеннон родился 30 апреля 1916 г. в городке Гейлорд (США). Окончив в 1936 г. Мичиганский университет, Шеннон получил степень бакалавра как математик и как электроинженер. Двумя годами позже он стал магистром, защитив диссертацию на тему «Символический анализ релейных и переключательных схем», в которой впервые показал, что анализ таких схем (прообраз современной компьютерной техники) может быть проведён с помощью математического аппарата символической (булевой) логики.

Одновременно аналогичные результаты получили В. И. Шестаков и А. Накасима.

Довольно неожиданно Шеннон обратился к генетике и защитил в 1940 г. диссертацию по математике (на степень доктора философии) на тему «Алгебра теоретической генетики». К сожалению, эта работа не бы-

ла вовремя опубликована и увидела свет лишь в конце XX в.

В начале Второй мировой войны Шеннон поступает работать в математическую лабораторию компании AT&T Bell Telephones, одновременно он является советником национально-исследовательского комитета Министерства обороны США. В Bell Labs он трудится вплоть до 1972 г. За эти годы учёный создал теорию информации, заложил основы теории управляющих систем и, написав всего одну статью по криптографии, превратил этот предмет из искусства в науку.

В 1956 г. Клод Шеннон избран членом Национальной академии наук США и Американской академии искусств и наук, а с 1957 г. становится профессором электротехники и математики Массачусетсского технологического института.

Элизабет Шеннон у памятника мужу в Гейлорде.

Клод Шеннон не был стереотипным «кабинетным учёным». Его всегда интересовала возможность создания «шахматной машины» или «шахматной программы». Приехав в Москву в 60-х гг. он добился того, чтобы сыграть партию с чемпионом мира Михаилом Ботвинником. Он сочетал таланты циркового артиста (мог жонглировать, катаясь на одноколёсном велосипеде) и умельца (сам смастерил несколько автоматов, среди них — жонглирующаю машина).

Около двух десятилетий учёные разных стран (но в основном СССР и США) обобщали результаты, полученные Шенноном, и давали им строгое математическое обоснование. Следует особо отметить работы А. Я. Хинчина, И. М. Гельфанда, А. Н. Колмогорова, А. М. Яглома, Р. Л. Добрушина и М. С. Пинскера.

При этом обнаружилось, что идеи теории информации оказались полезными и для самой математики. Так, использование понятия энтропии привело А. Н. Колмогорова к решению проблемы изоморфизма динамических систем, а также подтолкнуло его к решению 13-й проблемы Гилберта (о представимости непрерывных функций от многих переменных в виде суперпозиции непрерывных функций от меньшего числа переменных).

Теория информации Шеннона базировалась на теории вероятностей. Во второй половине 60-х гг. А. Н. Колмогоров создал алгоритмическую теорию информации, позволившую по-иному взглянуть на такое основополагающее понятие, как «случайность».

Выдающемуся учёному Клоду Шеннону посчастливилось перешагнуть в XXI в. Он прожил почти 85 лет и скончался 24 февраля 2001 г. в Медфорде, штат Массачусетс.

А. Я. Хинчин.

И. М. Гельфанд.

А. Н. Колмогоров.

КОДИРОВАНИЕ ИНФОРМАЦИИ

ОТ РИСУНКА К БУКВЕ

Первые начертания, смысл которых нам понятен, — изображения животных на стенах пещер, сделанные нашими далёкими предками десятки тысяч лет назад. Однако между ними и письменностью — огромная дистанция.

Для передачи информации изначально использовались так называемые предметные знаки: ракушки, камешки, кости и перья животных, узелки и т. д. Такой знак может, например, указывать на наличие целого класса аналогичных предметов (именно на этом принципе строится заполнение витрин магазинов, когда выставляются только некоторые предметы из ассортимента). В другом случае предметный знак выступает как символ понятия, предмета или целой ситуации. Так, в одном из африканских племён в качестве знака оповещения о казни осуждённому посылалось яйцо попугая.

Наиболее типичные функции таких знаков связаны с обозначением принадлежности, статуса, собственности, с магией и со счётом. Забор — один из первых знаков собственности, корона или скипетр — знаки власти и положения в обществе, хрустальный шар — символ ясновидения, а камешки давно превратились в костяшки счётов.

Вероятно, именно со счётом были связаны знаки, переходные от предметов к изображению. Это зарубки на палке, которыми обозначался долг. Когда нанесённые на какую-либо поверхность значки используются уже не только для счёта, но и для передачи развёрнутых сообщений, возникает идеографическое, т. е. картинно-изобразительное, письмо, нечто среднее между рисунком и знаковым письмом. Картинка в целом могла изображать некую ситуацию, как в наскальной живописи, но в ней уже выделялись элементы, не столько изображающие, сколько обозначающие отдельные элементы ситуации.

Закрепление отдельных знаков за определёнными понятиями, возможность их использования для описа-

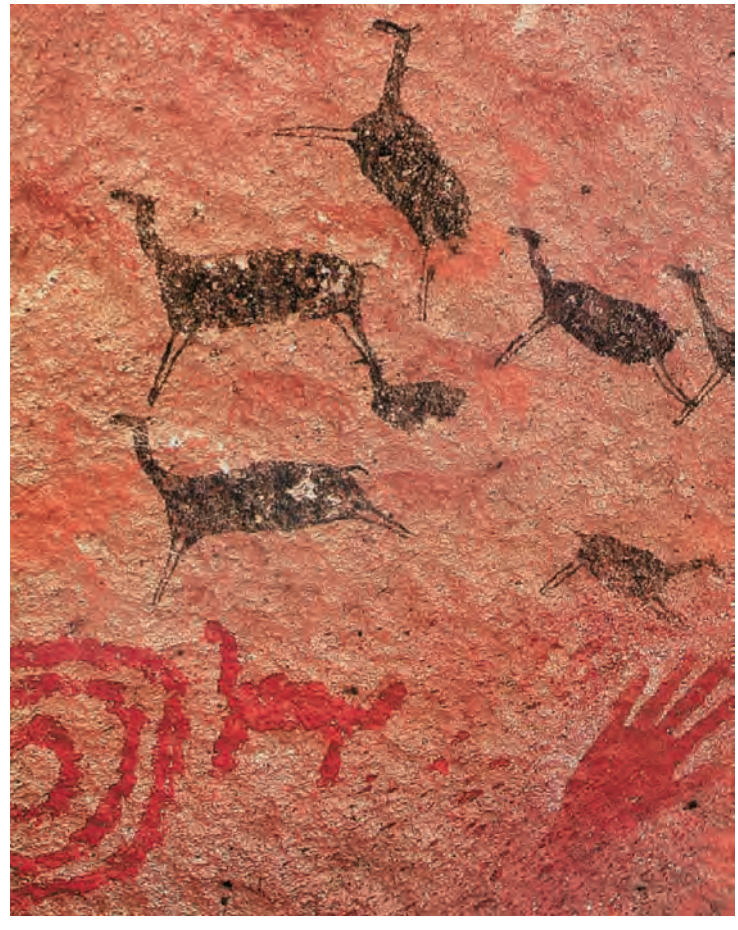
ния разнообразных ситуаций означают переход к словесному, или иероглифическому, письму. В нём начертанный знак обозначает слово. При этом начертание всё больше схематизировалось, теряло сходство с первоначальным изображением, ассоциировалось уже не с самим предметом, а со словом. Происходил переход к письму, кодирующему уже не непосредственно действительность, а человеческую речь.

Возможности иероглифической письменности несопоставимо шире, чем идеографии. Ведь далеко не всё, о чём думают и что считают важным сообщить друг другу люди, можно буквально изобразить. Знаки же, которые больше не привязаны к изображению, а напрямую передают слова языка, позволяют передать даже самый общий, отвлечённый смысл. Не только отдельный «старик», но и вообще «старость», не только определённый «холодный день», но и вообще «зима» как время года. Смыслы можно комбинировать. Например, знаки со значением «вода» и «падать» вместе могут обозначать «водопад», подобно тому как эти смыслы сложены в русском слове.

Иероглифические письменности — это богатые, развитые знаковые системы, которыми пользовались многие великие цивилизации и пользуются до сих пор (например, в китайском языке).

Наверное, иероглифическая письменность существовала бы вечно, если бы на земле был только один язык. Именно ситуация столкновения с иноязычными словами выявила ограничение иероглифической письменности. У иероглифа, как знака, нет двойного членения, он отсылает к слову целиком, а не передаёт его по звукам. И как же в таких условиях обозначить новое слово? Например, имя иностранца? Если этот иностранец становился правителем, захватив данную территорию, или иностранка становилась супругой правителя, такие события не могли не найти отражения в летописях. Чтобы решить столь непростую задачу, древние писцы стали произвольно делить иностранное имя на части, похожие по звучанию на те или иные слова родного языка, и передавать при написании соответствующими иероглифами. Будь, например, современная русская письменность иероглифической, нам пришлось бы передавать фамилию одного из американских президентов Билла Клинтона изображениями колышка и ноты: «клин» плюс «тон». А чтобы читатели понимали, что в данном случае иероглифы использованы, так сказать, не всерьёз, а лишь как части ребуса, ставились особые значки, которые указывали: «Воспринимай не смысл, а только звучание».

Последний шаг к письменности в современном понимании тоже был связан с взаимодействием цивилизаций и культур. Финикийцы, купцы и мореплаватели, приспособили для своих нужд позаимствованные у египтян иероглифы, упростив их начертание и не вдаваясь в исходное значение



ИЩЕМ КЛАД

Сколько надо сделать шагов?
На север десять и десять, на восток пять и пять,
на юг два и два, на запад один и один и потом вниз...

А. Конан Дойл. «Обряд дома Месгрейвов»

Все, наверное, помнят, как по этому указанию из старинного документа был найден клад. Этот текст представляет собой задание системы координат для поиска точки в пространстве. Путь определяется использованием географических координат — север, юг, запад, восток — и меры длины — шага. Ориентирование по солнцу — один из наиболее известных и удобных способов определения положения в пространстве. Древние люди, путешествуя по суше и по воде, прекрасно умели определять своё местоположение, ориентируясь по солнцу, луне, звёздам. На таких приёмах до появления компаса строилась вся навигация.

Сохранить и передать информацию о пути, о незнакомых берегах и странах и даже о местах, где спрятаны сокровища, помогают карты. Если нам нужно объяснить кому-то дорогу, легче нарисовать план, чем рассказывать на словах.

Первые карты появились, по-видимому, в древнем Двуречье: известны вавилонские карты на глиняных табличках, относящиеся к XXIII в. до н. э. В Средние века моряки использовали так называемые портоланы — карты морских торговых путей с нанесением «роз компаса», имевших вид лучей, расходящихся из одного центра, и указывавших проекции кратчайших путей из одной точки к другой. Дальнейшие достижения в развитии картографии связаны с введением Герардом Меркатором (из Фландрии) метода цилиндрической проекции, устанавливавшего систематическое соотношение между изображениями на сферической и плоской поверхностях. Использование проекции позволяло перевести географические координаты — широту и долготу — со сферической поверхности на плоскость с эффектом наименьшего искажения.

Современная карта — это универсальный указатель, кодирующий пути в пространстве. Геометрическое подобие реальных очертаний физических объектов и их контуров на карте создаётся масштабированием. Используется также цвет, который носит изобразительный характер (синий для водоёмов, зелёный для лесов и равнин) или чисто символический, например на политической карте мира. Кроме того, на карте используются собственно символьные знаки и обозначения, указывающие населённые пункты, дороги, места залегания полезных ископаемых или зоны обитания животных.

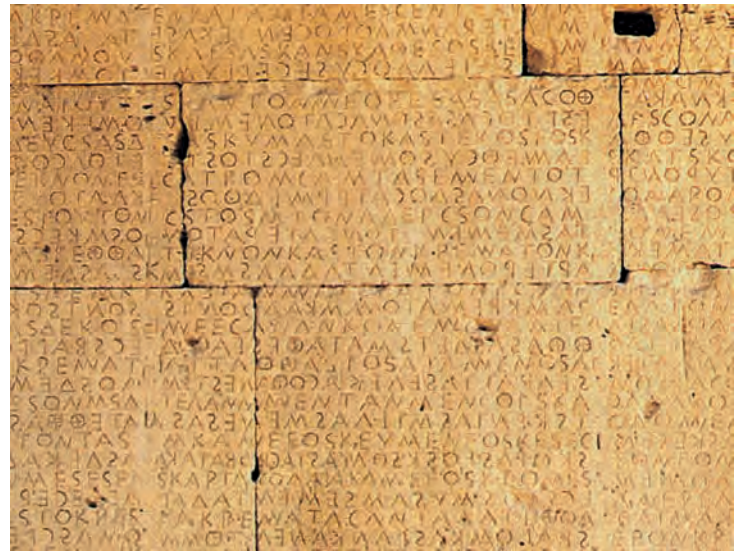
этих значков, а используя их для передачи слогов собственного языка. Так появилось слоговое фонетическое письмо. В нём отдельный значок передавал сочетание гласного и согласного, а знаки исчислялись уже не тысячами, как в иероглифическом письме, а десятками.

Наконец, греки, переняв письменность финикийцев, сделали последний, решающий шаг: они отделили гласные от согласных и стали передавать одним значком — буквой — один звук. Именно греки создали знакомое всем нам звуко-буквенное, или алфавитное, письмо, явившееся заключительным этапом в общей истории письма. По такому пути шли все исторические письменности.

Развитие знаковых систем записи — путь утраты изобразительного принципа, внешнего подобия и замена его функциональным подобием. На этом пути можно выделить два наиболее важных, революционных момента. Первый — создание иероглифа, особого письменного знака для обозначения слова, второй — разложение значимой единицы — слова — на незначимые элементы — слоги или звуки, обозначаемые отдельными зна-

Сэмюэл Финли Бриз Морзе.

ками. Эволюция письма даёт преимущество — сокращение кода от тысяч иероглифов до двух-трёх десятков букв. Предел сокращения кода — два знака. Ближе всего к этому решению подошли Сэмюэл Морзе и Луи Брайль. Морзе зашифровал буквы латинского алфавита комбинациями из двух элементарных графических знаков (и ещё одного — паузы) — точки и тире, а Брайль — выпуклыми точками или их отсутствием. Однако обратная сторона сокращения кода — увеличение текста. Фраза, для записи которой в иероглифической письменности потребуется три значка, в звуко-буквенной займёт почти целую строчку, а в записи азбукой Морзе — несколько строк.



Фрагмент древнегреческой надписи из Гортины (остров Крит). Середина V в. до н. э.

СЧИТАЕМ ВОРОН

В основе счёта лежит идея сопоставления, т. е. предметы счёта последовательно сопоставляются со счётными средствами: палочками, камешками, пальцами рук и пр. Именно пальцевый счёт лёг в основу самой распространённой системы счисления — десятичной. Ведь пальцы рук — наиболее удобное средство для счёта. Впрочем, к пальцам обращаются и другие системы, например двадцатеричная система древних галлов и грузин (видимо, учитывались и пальцы ног), пятеричная (считали по одной руке). Совсем необычная пальцевая система была у папуасов племени телефол на Новой Гвинее: они считали начиная от мизинца правой руки, переходя на локоть, плечо, доходили до кончика носа и симметрично переходили на левую руку — получалось 27 единиц счёта.

Допустим, мы сосчитали ворон. Как эту информацию передать? Можно, конечно, просто нарисовать предметы счёта, т. е. шесть ворон. Но лучше поступить более экономно: изобразить одну ворону, а рядом — шесть

ТОЧКИ ВМЕСТО БУКВ

Чем короче код, тем длиннее запись, и наоборот. Задачу найти оптимальное соотношение краткости кода и краткости записи пришлось решать при создании письменности для незрячих людей. Читать на ощупь привычные нам буквы очень неудобно, противопоставление составляющих их элементов, легко воспринимаемое зрением, оказывается недостаточным для осязания либо буквы приходится делать настолько крупными, что запись и хранение сколько-нибудь длинного текста превращаются в настоящую проблему.

В 1824 г. 15-летний Луи Брайль, ослепший в раннем детстве и очень страдавший от невозможности самостоятельно читать, придумал использовать для обозначения букв комбинации из шести позиций (2·3), часть из которых может быть выпуклой. Таких комбинаций, как легко подсчитать, 64, чего вполне достаточно, чтобы передать все буквы алфавита, знаки препинания и цифры. Выпуклые точки наносятся на плотную бумагу и легко распознаются на ощупь. Более того, с помощью этой азбуки слепой человек и сам может писать, выдавливая точки острым пером.

В системе Брайля имелся особый код-признак числа и отменяющий его код-признак буквы, поскольку для удобства цифры кодировались теми же знаками, что и первые десять букв. Это были первые в истории коды переключения.

Позже для удобства и сокращения длины записи комбинациям, оставшимся незанятыми, присвоили значения наиболее употребительных сочетаний букв.

Кроме того, если знак отделяется от соседних большим интервалом, он обозначает не букву, а часто встречающееся слово: you вместо y, not вместо n и т. п.

пальцев (палочек). Именно так обычно и делали в древности. Вероятно, чёрточка изначально и изображала палец (исторически русские слова «палец» и «палка» одного корня), а может, это просто самый удобный для начертания знак. У некоторых народов (например, у майя) единица обозначалась кружком или точкой — тоже максимально простой знак, а возможно, изображение камешка.

В принципе имея знак для единицы, можно им одним и обходиться. Но такой способ, наверное, удобен, только чтобы считать дни в темнице. В прочих случаях громоздкость записи существенно затруднит работу с ней. Поэтому люди перешли к обозначению знаками целых множеств: пятёрок, десяток, дюжин и т. д. Например, римская цифра V, как считают исследователи, восходит к изображению ладони. Таким способом довольно удобно обозначать большие числа, имея в виду, что общее числовое значение получается при сложении этих величин. Для экономии вводят порядок в расположении знаков и наряду со сложением используют приём вычитания: допустим, единицы справа от знака, обозначающего десятки, прибавляются к нему, а слева — вычитаются. Так, в частности, устроена римская система: её основные знаки — идеограммы в виде букв латинского алфавита — I, V, X, L, C, M (1, 5, 10, 50, 100, 1000) и т. д. С помощью букв обозначали числа и греки, и русские (перенявшие у греков бóльшую часть алфавита), и армяне, и грузины — способ оказался очень удобным и получил широкое распространение. Понятно, что в какой-то момент букв может не хватить, но в те далёкие времена со столь большими числами работали редко.

Тот же важнейший шаг, который совершили финикийцы и греки для письменности, сделали индийцы и арабы для записи чисел. Они придумали нуль и позиционную систему записи чисел. Идея её внешне проста: два человека могут, используя пальцы рук, досчитать не до 20, а до 100, если договорятся, что первый будет считать единицы, а второй — десятки (трое досчитают не до 30, а аж до 1000 и т. д.).

СИСТЕМЫ СЧИСЛЕНИЯ

КАК СЧИТАЛИ В ВАВИЛОНИИ

Между реками Тигр и Евфрат, в могучем государстве Вавилонии, высокого развития достигли счёт и системы счисления (см. раздел «Старинные системы записи чисел» в томе «Математика» «Энциклопедии для детей»). Вавилонские числа являются комбинацией не двух, а трёх клинописных знаков: единицы — , десятки — и сотни — .

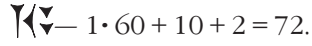
С помощью этих знаков можно записать любое число, используя принцип сложения или умножения:

 — $10 \cdot 100 = 1000$,

 — $10 + 1 = 11$.

Большие числа всегда предшествовали меньшим. Кроме этого способа записи чисел применялась также позиционная шестидесятеричная система. Знак единицы мог обозначаться как 1, 60, 60^2, а знак десятки — соответственно 10, $10 \cdot 60^2$ и т. д.,

в зависимости от порядка расположения:

 — $1 \cdot 60 + 10 + 2 = 72$.

Вавилоняне имели подобие знака «нуль» —  (два наклонных знака единицы), но, увы, этот знак не ставился в конце чисел.

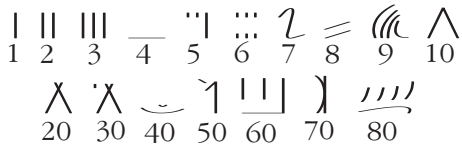
Они умели также пользоваться простыми и шестидесятеричными дробями (со знаменателями 60, 60^2 и т. д.), которые записывали так, как сейчас записывают десятичные дроби. При вычислениях вавилоняне использовали готовую таблицу умножения.

КАК СЧИТАЛИ В ЕГИПТЕ

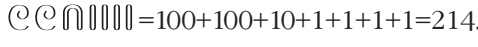
Главным источником современных знаний о египетской числовой системе является так называемый папирус Ахмеса, найденный в 1853 г. (или

папирус Райнда — по фамилии владельца, который приобрёл этот документ в 1858 г.).

Для записи чисел египтяне применяли иероглифы: *один* —  , *десять* — , *сто* —  , *тысяча* —  , и так до 10 млн —  . Затем иероглифическое письмо было упрощено иератическим (от *греч.* «иератикос» — «священный»):



Все остальные числа записывались из этих знаков при помощи операции сложения. Вначале писались числа высшего порядка, а затем низшего:



Умножение и деление египтяне производили путём последовательного удвоения чисел — особая роль отводилась двойке.

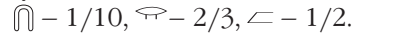
В примере 19 • 31 египтяне последовательно удваивали число 31. В правом столбце записывали результаты удвоения, в левом — соответствующую степень двойки:

1	31
2	62
4	124
8	248
16	496

Древнеегипетский папирус с математическими выкладками.

Затем отмечали вертикальными чёрточками строки левого столбца, из которых можно было сложить множитель (19 = 1 + 2 + 16), и складывали числа, стоящие в отмеченных строках справа (31 + 62 + 496 = 589).

Египетские дроби всегда имели в числителе единицу (исключение составляло 2/3; см. статью «Методы вычислений» в томе «Математика» «Энциклопедии для детей»). Дроби записывались как натуральные числа, только над ними ставилась точка, специальные знаки были для 1/2 и для 2/3:



КАК СЧИТАЛИ В РИМЕ

Римские числа общеизвестны, их сейчас можно увидеть во многих местах, например на циферблате кремлёвских курантов — главных часов России.

В римской системе счисления семь чисел обозначаются буквами: 1 — I, 5 — V, 10 — X, 50 — L, 100 — C, 500 — D, 1000 — M, а остальные числа записываются комбинациями этих знаков. Если числа в комбинации идут

в порядке от больших к меньшим, числа складываются:

$$XXI — 10 + 10 + 1 = 21,$$

$$MMV — 1000 + 1000 + 5 + 1 = 2005,$$

если от меньших к большим — значение числа вычитается из следующей буквы:

$$IV = 5 - 1 = 4.$$

Складывать и вычитать в такой системе удобно, но умножать и делить очень сложно. Римляне пользовались дробями со знаменателем 60 (как в Вавилонии) и со знаменателями 12, 24, 48. Освоение дробей требовалось для счёта денег, мер и весов. Римская монета асс чеканилась из меди, имела вес один фунт и делилась на 12 унций.

КАК СЧИТАЛИ ГРЕКИ

Греки применяли несколько способов записи чисел. При использовании ионической нумерации числа выражались буквами алфавита. Чтобы отличить число от слова, над буквами числа ставился специальный значок ⁓ — *титло*. Этот способ записи чисел применялся жителями Милета и Александрии. Афиняне для обозначения чисел пользовались первыми буквами слов — числительных:

Γ (Γεντε)	— пять,
Δ (Δεκα)	— десять,
Η (Ηκατον)	— сто,
Χ (Χιλιασ)	— тысяча,
Μ (Μυριασ)	— десять тысяч,
I, II, III, IIII	— соответственно 1, 2, 3, 4,
ΔΔIIII	— 10 + 10 + 10 + 4 = 34.

С помощью этих цифр житель Древней Греции мог записать любое число.

Великий греческий математик Диофант Александрийский записывал дроби примерно так, как принято сейчас: числитель над знаменателем, но без черты. Это был один из способов записи дробей в Древней Греции.

ИНДИЙСКАЯ НУМЕРАЦИЯ

Цифры, которыми пользуются сейчас, пришли из Индии. Европейские народы познакомились с ними благодаря арабам. Математик Леонардо Пизанский первым упоминает об арабских числах в 1202 г. В XVI в. новая нумерация получает широкое распространение; в Россию она попадает в XVII в. и в начале XVIII в. полностью вытесняет алфавитную систему. Десятичная позиционная система (арабская система) даёт принципиальную возможность записывать сколь угодно большие числа.

В десятичной системе десять цифр: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

Число зависит не только от цифры, но и от порядка расположения. Число 44 обозначает количество единиц и десятков, а цифра 0 указывает позицию цифры, например 40.

Особую роль играют число 10 и его степени: 10, 10^2, 10^3...

$$2005 = 5 + 0 \cdot 10 + 0 \cdot 100 + 2 \cdot 1000,$$

или

$$2005 = 5 + 0 \cdot 10^1 + 0 \cdot 10^2 + 2 \cdot 10^3,$$

или

$$2005 = 2 \cdot 10^3 + 0 \cdot 10^2 + 0 \cdot 10^1 + 5 \cdot 10^0.$$

Любое четырёхзначное число можно записать как

$$N = a_3 \cdot 10^3 + a_2 \cdot 10^2 + a_1 \cdot 10^1 + a_0 \cdot 10^0,$$

где a_0, a_1, a_2, a_3 — десятичные цифры числа (цифра a_3 отлична от 0). Число 10 в формуле называют *основанием* системы счисления. Существуют системы счисления с другим основанием — p.

Запись числа n в p-системе счисления:

$$n = a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p^1 + a_0 p^0,$$

где $a_i \in \{0, 1, 2, \ldots, p-1\}$ и $a_n \neq 0$.

Если в качестве p взять число 2, то выйдет двоичная система счисления.

ДВОИЧНОЕ КОДИРОВАНИЕ

Двоичная система счисления была придумана задолго до появления компьютеров. Ещё великий немецкий математик Готфрид Вильгельм Лейбниц увидел в двоичной системе особый скрытый смысл. Действительно, двоичная система предельно проста:

используется всего две цифры — 0 и 1;

таблица умножения состоит всего из трёх строк:

$$0 \cdot 0 = 0$$
$$0 \cdot 1 = 0$$
$$1 \cdot 1 = 1$$

К недостаткам двоичной системы можно отнести только «длинную» запись чисел (чем меньше в системе цифр, тем длиннее будет запись числа):

$$2002_{10} = 11111010010_2.$$

Однако этот досадный факт не помешал инженерам использовать двоичную систему при конструировании электронных схем. Перевести число из двоичной системы в привычную десятичную может не только вычислительная машина, но и человек.

Для удобства запишем двоичное число в таблицу поразрядно, пронумеруем разряды в следующей строке (справа налево, начиная с 0) и выпишем соответствующие степени двойки (справа налево с 0-й степени).

Для числа 11111010011_2 таблица выглядит так:

Число	1	1	1	1	1	0	1	0	0	1	1
N	10	9	8	7	6	5	4	3	2	1	0
2^n	1024	512	256	128	64	32	16	8	4	2	1

Единицы в первой строке таблицы указывают, какие степени двойки нужно сложить, чтобы получить число:

$$1024 + 512 + 256 + 128 + 64 + 16 + 2 + 1 = 2003,$$

т. е. $111110100112 = 200310$.

Алгоритм перевода числа в двоичную систему аналогичен предыдущему, но требует проведения операции деления с остатком. Число последовательно делят на два, выписывая результат в столбик, одновре-

менно записывают остатки (не пропуская нулевые). Ответом будет число, получившееся в правом столбце (снизу вверх). Для числа 413 это выглядит так:

413	1
206	0
103	1
51	1
25	1
12	0
6	0
3	1
1	1

Если теперь выписать получившееся во втором столбце (снизу вверх) число, оно и будет ответом:

$$110011101_2 = 413_{10}.$$

В переводе чисел из двоичной системы счисления в десятичную систему и обратно нет ничего сложного и магического. Каждый освоивший эту несложную науку может считать себя понимающим язык ЭВМ. Однако записывать в двоичной системе числа по-прежнему неудобно, так как запись занимает много места. И перевод в двоичную систему вряд ли удастся выполнить в уме. Поэтому стали использовать системы, родственные двоичной системе счисления, в которых запись числа на бумаге короче, чем в двоичной, а алгоритмы перевода не требуют сложных вычислений.

ВОСЬМЕРИЧНАЯ СИСТЕМА СЧИСЛЕНИЯ

В этой системе восемь цифр: 0, 1, 2, 3, 4, 5, 6, 7. Цифра 1, записанная в младшем разряде, означает — как и в десятичном числе — просто единицу, а в следующем разряде она означает 8_{10}, в следующем — 64_{10} и т. д.

Число 100_8 — это 64_{10}, а число $635_8 = 6 \cdot 64 + 3 \cdot 8 + 5 = 413_{10}$. При вычислениях числа указаны в десятичной системе и признак десятичной системы 10 опущен.

Компьютер PDP 11/70 использовал восьмеричную кодировку.

Перевод из восьмеричной системы в десятичную по сложности вычислений мало отличается от перевода из двоичной. А вот перевод из восьмеричной системы в двоичную очень прост. Достаточно составить таблицу триад (по три цифры)

$0_8 - 000_2$	$4_8 - 100_2$
$1_8 - 001_2$	$5_8 - 101_2$
$2_8 - 010_2$	$6_8 - 110_2$
$3_8 - 011_2$	$7_8 - 111_2$

и запомнить её как таблицу умножения.

При переводе восьмеричного числа в двоичное заменяют каждую восьмеричную цифру на соответствующую триаду из таблицы. В качестве примера переведём число 611_8 в двоичную систему:

$611_8 = 110\ 001\ 001_2$
(6_8 заменили на 110_2,
а 1_8 — на 001_2).

В ЭВМ VAX использовалась шестнадцатеричная кодировка.

Для обратной операции, т. е. для перевода из двоичной в восьмеричную систему, двоичное число разбивают на триады (справа налево), потом заменяют каждую группу одной восьмеричной цифрой. Допустим,

$$101\ 111\ 111_2 = 577_8.$$

Если первой полноценной триады не получается (не хватает цифр), то её дополняют ведущими нулями:

$$1\ 111\ 000_2 = 001\ 111\ 000_2 = 170_8.$$

ШЕСТНАДЦАТЕРИЧНАЯ СИСТЕМА СЧИСЛЕНИЯ

Запись числа в восьмеричной системе достаточно компактна, но ещё компактнее она получается в шестнадцатеричной системе. Для первых десяти цифр используют привычные цифры: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, а для остальных шести цифр — первые буквы латинского алфавита:

A — 10,
B — 11,
C — 12,
D — 13,
E — 14,
F — 15.

Как и в восьмеричной системе, цифра 1, записанная в младшем разряде, означает единицу. В следующем разряде та же цифра 1 означает 16_{10}, в следующем — 256_{10} и т. д. Цифра F, записанная в младшем разряде, означает 15_{10}, в следующем разряде — $15_{10} \cdot 16_{10}$ и т. д. Например,

$$100_{16} = 256_{10},$$

$$1AF_{16} = 1 \cdot 16^2 + 10 \cdot 16 + 15 = 431_{10}.$$

Перевод из шестнадцатеричной системы в двоичную и обратно аналогичен действиям с восьмеричной системой. Надо опять запомнить таблицу перевода, где вместо триад стоят тетрады (четвёрки):

$0_{16} - 0000_2$	$8_{16} - 1000_2$
$1_{16} - 0001_2$	$9_{16} - 1001_2$
$2_{16} - 0010_2$	$A_{16} - 1010_2$
$3_{16} - 0011_2$	$B_{16} - 1011_2$
$4_{16} - 0100_2$	$C_{16} - 1100_2$
$5_{16} - 0101_2$	$D_{16} - 1101_2$
$6_{16} - 0110_2$	$E_{16} - 1110_2$
$7_{16} - 0111_2$	$F_{16} - 1111_2$

Следовательно,

$61A_{16} = 110\ 0001\ 1010_2$
(6_{16} заменили на 0110_2,
1_{16} — на 0001_2, A — на 1010_2).

При совершении обратной операции двоичное число надо разбить на четвёрки цифр справа налево (если первой четвёрке не хватает цифр, то её дополняют ведущими нулями) и потом заменить каждую группу одной шестнадцатеричной цифрой:

$$11\ 1110\ 0111_2 = 0011\ 1110\ 0111_2 = 3E7_{16}.$$

БИТЫ И БАЙТЫ

Символы двоичной системы счисления — 0 и 1 называют двоичными цифрами или битами (от *англ.* binary digit — «двоичная цифра»).

У электронных схем для обработки двоичных кодов есть только два устойчивых состояния — «есть сигнал»/«нет сигнала» (высокое напряжение/низкое напряжение).

Подобные схемы нетрудно реализовать технически. Поэтому инженеров двоичное кодирование информации привлекает, несмотря на то что числа получаются слишком длинными. Компьютеру легче иметь дело с большим числом простых элементов, чем с небольшим числом сложных.

При хранении информации в памяти компьютера каждый бит хранится в одном разряде памяти. Разряды объединяются в ячейки памяти фиксированного размера — 8, 16 и 32 разряда и носят специальные названия «байт», «слово» и «двойное слово». Фактически это является стандартом для современных персональных компьютеров.

При кодировании чисел важно помнить, что, например, число 1000_{10} не удастся поместить в один байт, поскольку его двоичная запись содержит десять цифр (1111101000_2), что больше размера байта. И конечно, существуют большие числа, не помещающиеся даже в двойное слово.

Такие числа можно разместить в большем числе байтов, однако компьютер умеет быстро производить вычисления с числами, которые помещаются в байты и слова (иногда двойные и четверные слова), и если число большое, то операции с ним будут распадаться на операции с байтами, словами и т. д.

По аналогии в одном байте можно закодировать 256 различных символов какого-либо текста. На первый взгляд это много: буквы русского алфавита, прописные и строчные, знаки препинания, десять цифр, ещё латинский алфавит. Однако если потребуется ввести ещё какой-нибудь алфавит, то, скорее всего, места уже не хватит. Тогда символ текста будет кодироваться не байтом, а словом, что автоматически увеличит объём хранимой информации в два раза.

Как уже говорилось, при двоичном кодировании в компьютере нельзя записать ничего, кроме двоичных цифр. В одном байте можно хранить одно неотрицательное целое число от 0 до 255, а требуется ещё придумать кодирование отрицательных чисел. Например, в одном байте можно хранить 128 отрицательных целых чисел (от −128 до −1), 0 и 127 положительных чисел от 1 до 127. Всего же чисел попрежнему останется 256, как и закодированных символов.

Нуль будет кодироваться комбинацией из восьми нулей: 00000000.

Положительные числа от 1 до 127 кодируются, как обычно, в двоичной системе, при этом слева дописывают *незначащие* нули:

1 — 00000001
2 — 00000010
3 — 00000011
…
126 — 01111110
127 — 01111111

Отрицательное число запишется в *дополнительном коде*. Например, отрицательное число −5 хранится как двоичная запись числа 256 − 5 ($256 = 2^8$), т. е. дополнение к 5 до 256:

11111011

Поскольку −5 + 256 >127, в старшем разряде двоичного числа будет единица, сигнализирующая, что число отрицательное, поэтому старший разряд ячейки обычно называют *знаковым*. Так кодируются отрицательные числа −1, −2, −3, … , −127, −128.

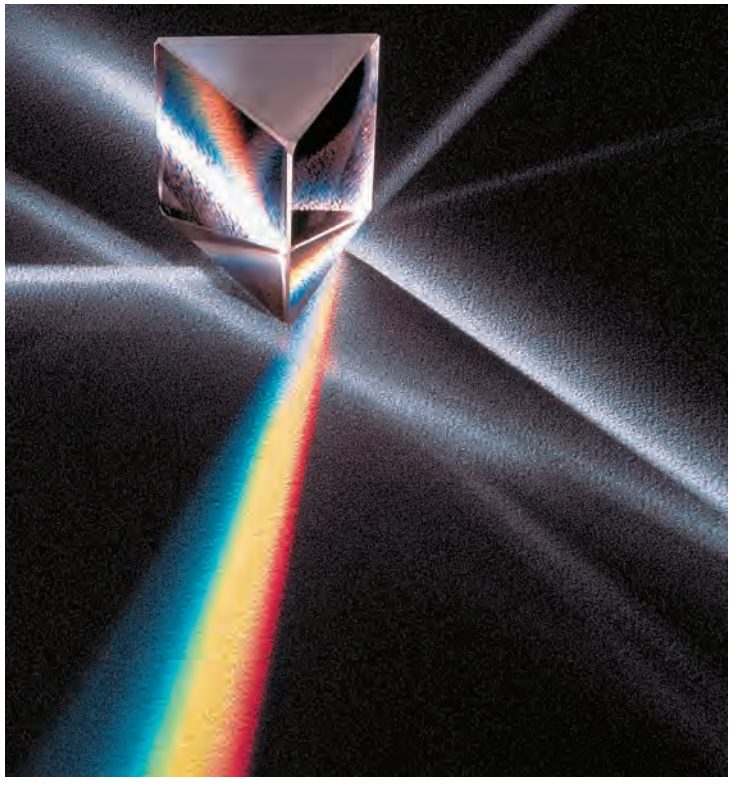
КОДИРОВАНИЕ ИЗОБРАЖЕНИЙ

Человек получает бóльшую часть информации из окружающего мира с помощью зрения. Глаз человека ощущает свет только тогда, когда на сетчатку попадает лучистая энергия диапазона 360—750 нм, т. е. видимая часть спектра электромагнитных колебаний.

Предметы воспринимаются человеком благодаря их цветности (вместе цветовой тон и насыщенность цвета) и различной яркости. Чем дальше удалён предмет от источника света, тем менее ярким он воспринимается. Яркость предмета зависит также от его отражательной способности и формы. Белый свет — самый сложный, он объединяет излучения разных длин волн. Каждый предмет отражает его по-своему, что и объясняет цветовое различие между ними.

Цвет можно оценить по количеству энергии, приходящейся на каждую длину волны, — по *спектральной характеристике*.

Призма и спектральное разложение света.



Однако применима и визуальная оценка цвета. В 1756 г. выдающийся русский учёный Михаил Васильевич Ломоносов (1711—1765) впервые высказал мысль, что для воспроизведения любого цвета в природе достаточно смешать в определённых пропорциях три основных цвета: красный, зелёный, синий. Теория *трёхкомпонентности* цвета, опираясь на многочисленные опыты, утверждает, что в зрительной системе человека возникают нервные возбуждения трёх типов, каждое из которых независимо от остальных. Многообразие цветовых ощущений при зрительном восприятии можно объяснить многообразием комбинаций трёх типов возбуждений. Эта теория лежит в основе большинства практических методов воспроизведения цветных изображений: в фотографии, телевидении, кино и т. п.

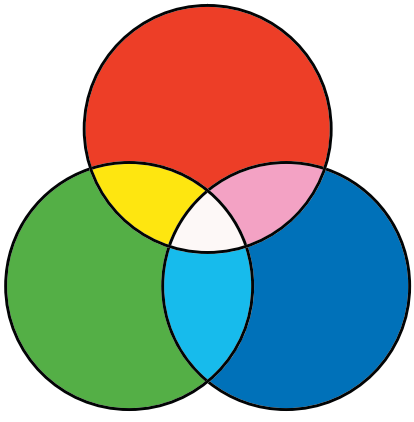
Компьютерное кодирование изображений также построено на этой теории. Картинка разбивается вертикальными и горизонтальными линиями на маленькие прямоугольники. Полученная матрица прямоугольников называется растром, а элементы матрицы — пикселями (от *англ.* picture's element — «элемент изображения»). Цвет каждого пикселя представлен тройкой значений интенсивности трёх основных цветов. Чем больше битов выделено для каждого основного цвета, тем большую гамму цветов можно хранить про каждый элемент изображения. В стандарте, называемом true color (реальный цвет), на каждую точку растра тратится 3 байта, по 1 байту на каждый основной цвет. Таким образом, 256 уровней яркости красного цвета, 256 уровней яркости зелёного и 256 уровней яркости синего дают вместе примерно 16,7 млн различных цветовых оттенков, это превосходит способность человеческого глаза к цветовосприятию.

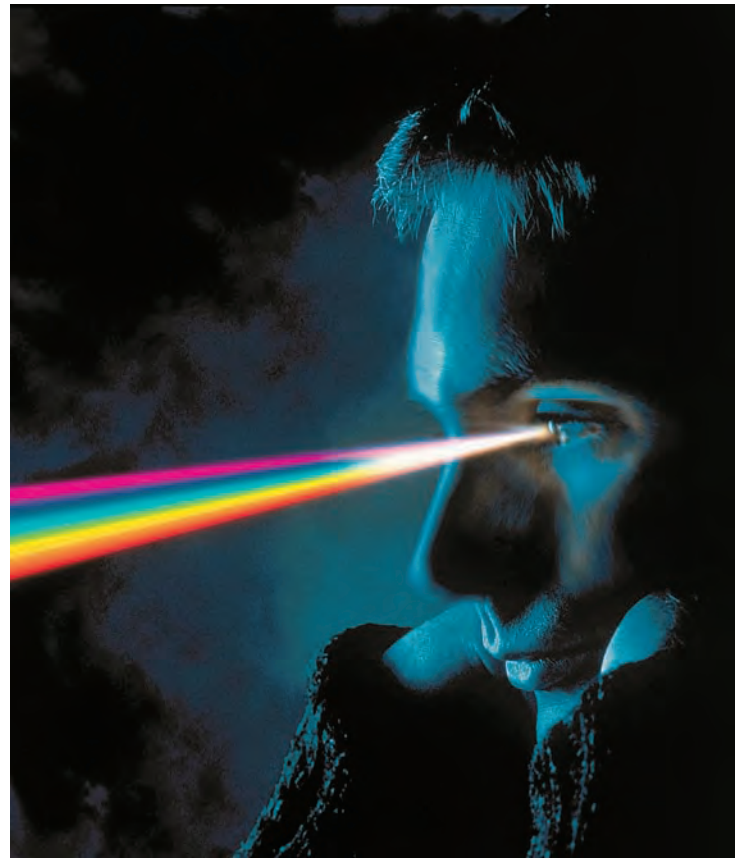
Чтобы хранить всю картинку, достаточно записывать в некотором порядке матрицу значений цветов пикселей, например слева направо и

сверху вниз. Часть информации о картинке при таком кодировании потеряется. Потери будут тем меньше, чем мельче пиксели. В современных компьютерных мониторах с диагональю 15—17 дюймов разумный компромисс между качеством и размером элементов картинки на экране обеспечивает растр от 768•1024 точки (бытовой видеомагнитофон — 250 линий на кадр, т. е. 250 точек по вертикали; отечественное телевидение стандарта SECAM — 625 линий на кадр).

При кодировании изображений не всегда используются именно три основных цвета. При печати картинок (и при рисовании) на бумаге если смешать красную и зелёную краски, то получится не жёлтый цвет, а коричневый. Это происходит потому, что краски сами по себе не отражают света, а только поглощают некоторые цвета из падающего на них светового потока. Поэтому в качестве основных применяются другие цвета: голубой, сиреневый и жёлтый, а метод кодирования цвета называется CMY (от *англ.* cyan — «голубой», magenta — «сиреневый» и yellow — «жёлтый»). При этом красный цвет получается как сумма сиреневого и жёлтого, а зелёный — как сумма жёлтого и голубого.

Так как визуальное восприятие цвета является трёхкомпонентным, то всякий цвет может быть задан не менее чем тремя параметрами. Во многих программах обработки изображения используется более приближённая к человеческому восприятию схема кодирования HSV (от *англ.* hue — «цветовой тон», saturation — «насыщенность» и value — «величина», «яркость»).





Цветовой тон — близость цвета к тем или иным цветам спектра. Наблюдая разложенный в спектр белый свет, глаз человека видит семь участков спектра, резко различающихся по зрительному ощущению: красный, оранжевый, жёлтый, зелёный, голубой, синий, фиолетовый. При внимательном рассмотрении человек улавливает гораздо больше цветов, только по цветовому тону глаз способен различить около 200 цветов. Цветовой тон характеризуется длиной волны того спектрального излучения, к которому ближе всего данный цвет.

Насыщенность — степень выраженности тона в цвете, т. е. насколько к данному спектральному цвету примешан одинаковый с ним по яркости белый цвет. Если к чистому спектральному цвету добавить белый

Три числа описывают цвет точки – это могут быть и RGB или три другие. Кодировка цвета, принятая в телевидении, описывает формулы для перевода из RGB к YC_bC_r:
$Y = 0{,}299R + 0{,}587G + 0{,}114B$,
$C_b = 0{,}564(B - Y)$,
$C_r = 0{,}649(R - Y)$.

Часто используется обозначение HSB (brightness) вместо HSV.

цвет, то при неизменной общей яркости насыщенность спектрального цвета уменьшится. Интересно, что в природе мало что (или кто) отличается насыщенностью цвета (возможно, лишь тропическая флора и фауна). Искусственные краски, например изображение на мониторе, обладают более высокой насыщенностью и поэтому так радуют человеческий глаз.

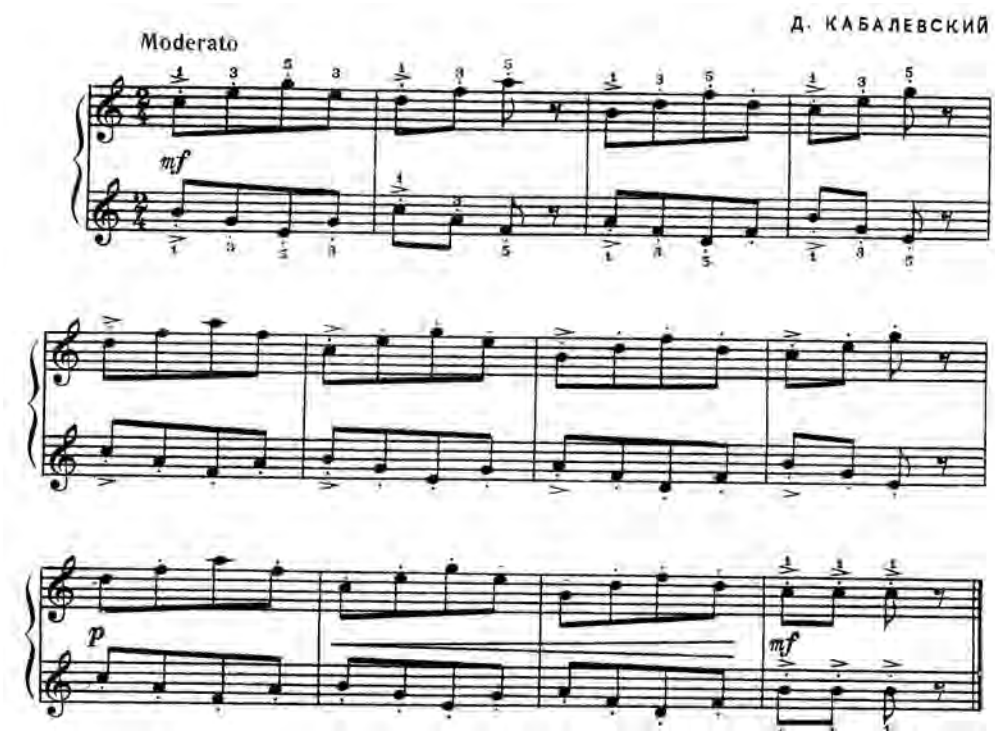
Яркость (светлота) определяется уровнем действующего на глаз излучения. Если одна часть объекта освещена прямым светом, а другая — рассеянным от того же источника, то цвета воспринимаются человеком поразному, несмотря на то что цветовой тон и насыщенность этих частей одинаковы.

Кодировка с помощью цветового тона, насыщенности и яркости позволяет легко получить из цветной картинки монохромную, или, как не совсем верно говорят, чёрно-белую, не меняя яркости объектов изображения.

КОДИРОВАНИЕ ЗВУКА И МУЗЫКИ

Как известно, звук представляет собой колебания воздуха. Амплитуда колебаний постоянно меняется, так как звук является непрерывным сигналом. При кодировании звука этот сигнал надо представить в виде последовательности нулей и единиц. Например, используя микрофон, звук можно превратить в колебания электрического тока, измеряя амплитуду колебаний через равные промежутки времени несколько десятков тысяч раз в секунду. Каждое измерение производится с ограниченной точностью и записывается в двоичном виде. Данный процесс называют *дискретизацией,* а устройство для его выполнения — *аналого-цифровым преобразователем (АЦП)*. Типичный восьмибитный *АЦП* преобразует напряжение в диапазоне от —500 мВ до 500 мВ в восьмиразрядные двоичные числа в диапазоне от -128_{10} до 127_{10}.

Обратный процесс — воспроизведение закодированного звука производится с помощью *цифроаналогового преобразователя (ЦАП)*. Тогда двоичные числа из диапазона от -128_{10} до 127_{10} преобразуются в напряжение из диапазона от −500 мВ до 500 мВ.



При работе со стереозвуком дискретизация проводится отдельно и независимо для левого и правого каналов.

Полученный ступенчатый сигнал сначала сглаживается посредством аналогового фильтра, а затем преобразуется в звук с помощью усилителя и динамика.

На качество воспроизведения закодированного звука в основном влияют два параметра — частота дискретизации и размер в битах, отводимый под запись значения амплитуды. Например, при записи на компакт-диски (CD) используются 16-разрядные значения, а частота дискретизации равна 44 032 Гц. Эти параметры обеспечивают превосходное качество звучания речи и музыки.

Выбор частоты дискретизации объясняется тем, что максимальная частота звука, которую способен слышать человек, не превосходит 22 кГц. Колебание с частотой 22 кГц при дискретизации с той же частотой неотличимо от тишины. Чтобы на каждом периоде дискретизации записывалось два значения, нужна вдвое большая частота дискретизации, а именно 44 кГц. Когда высокое качество не требуется, можно использовать меньшие частоты дискретизации: 11 кГц, 5,5 кГц и т. д.

Как и всякий звук, музыка является не чем иным, как звуковыми колебаниями, и, закодировав их достаточно точно, звук можно будет потом воспроизвести. Однако такой способ не позволяет композитору записать придуманное произведение, а музыканту — исполнить его, ни разу не услышав.

Русский композитор Ю. А. Шапорин писал: «Мне кажется, нет на свете другого искусства, которое так роднило людей, как музыка. Язык её понятен каждому...». Он имел в виду, что миллионы людей способны слушать музыку и наслаждаться ею, не зная ни названий музыкальных инструментов, ни принципов создания музыкальных произведений.

Выше частоты 5 кГц принцип умножения на два нарушается. Чтобы получить ощущение увеличения высоты на октаву, надо увеличить частоту почти в 10 раз. Даже люди с абсолютным музыкальным слухом затрудняются в определении нот для звуков с частотой выше 5 кГц. Это говорит о различии в механизмах восприятия высоты тона до 5 кГц и выше.

ЦВЕТА И ЗВУКИ

Мы выделяем семь цветов радуги и семь нот. Естественно, ни в звучащей мелодии, ни в цветовом спектре нет никаких реальных переходов, границ между отдельными звуками и цветами. Это условность, необходимая для того, чтобы музыку или цвет можно было кодировать и передавать информацию о них.

Рассмотрим, как разделяется на знаки, кодируется непрерывность музыкального звука. Музыкальный звук отличается от обычного шума наличием тона. Тон изменяется по высоте (фактически, частоте вибрации воздуха). В непрерывном изменении высоты выбирается некая условная начальная точка («до»). От неё отстраиваются другие точки, равно или соизмеримо отстоящие друг от друга (на условное расстояние 1 тон или 1/2 тона, которое является мерой изменения). Каждая из них имеет наименование — «ре», «ми» и т. д. В своём последовательном расположении они образуют звукоряд. Частоты вибраций воздуха членов звукоряда соотносятся как дробные числа (4/3, 3/2 и т. п.), что позволяет добиться ощущения гармонического созвучия в аккорде. Звукоряд повторяется в разных октавах при повышении или понижении частоты в два раза.

Для музыкальных звуков помимо высоты необходимой характеристикой является соотносительная длительность звучания. Её отмечают, соизмеряя с эталоном целой, которая последовательно делится на два: половинка, четверть, восьмая, шестнадцатая и т. д. (длительность целой условно приравнивалась к двум ударам человеческого сердца; основной же метрической единицей выступала четверть; именно по ней фиксировался так называемый музыкальный размер — длительность такта, начинающегося ударной долей и формирующего ритм).

Данная знаковая система помимо языковой имеет графическую форму кодирования — музыкальную грамоту, в которой с помощью особых символов обозначены высота и длительность звуков. Основной знак этой системы — нота в определённой позиции на нотном стане; декодировать высоту ноты по её положению помогает ключ (обычно скрипичный или басовый), задающий точку отсчёта. Нота — сложный знак, кодирующий и высоту звука, и его длительность.

Знаки певческой нотации, которые дали начало современным нотам, известны нам в довольно поздних византийских записях (с IX в.), хотя системы музыкальной записи существовали, по-видимому, значительно раньше. Так, известно, что у древних греков были буквенные системы для записи голоса и для записи игры на инструмен-

Выход был придуман давным-давно — *нотное письмо,* или *нотация* (от *лат.* notatio — «записывание», «обозначение»), — система графических знаков, применяемых для записи музыки. В начале Средних веков особыми знаками — невмами, состоящими из чёрточек, точек, запятых, — указывались отдельные звуки, группы звуков, ходы голоса вверх и вниз без точного определения высоты звуков, поэтому они могли лишь напоминать певцу уже известный ему напев. В XI в. итальянский музыкальный теоретик Гвидо д'Ареццо ввёл систему из четырёх нотных линий, явившуюся прототипом современного нотного стана. В начале линий помещались буквенные знаки (прообразы современных музыкальных ключей), определявшие высоту записанных звуков. В дальнейшем число линий довели до пяти, а невмы заменили нотами с квадратными головками. Впоследствии приняли нотацию, фиксировавшую как высоту, так и длительность звуков. С небольшими изменениями нотация дошла до наших дней.

Высота и тон звука устанавливаются частотой звукового колебания. Каждая нота имеет свою известную частоту. Сейчас основной считается нота «ля» первой октавы — 440 Гц (раньше её частота равнялась 435 Гц). Остальные ноты определяются относительно неё, ведь самым важным для слухового восприятия человека является не абсолютная частота, а отношение частот. Звуки, часто ты которых отличаются ровно в два раза, кажутся похожими, «родственными». Интервал частот от 262 Гц до $2 \cdot 262 = 524$ Гц разбивается на 12 частей; полученные 12 звуков называют первой октавой, семь из них знают все: *до, ре, ми, фа, соль, ля, си.* Удвоив частоты данных 12 звуков, получим вторую октаву, при следующем удвоении — третью.

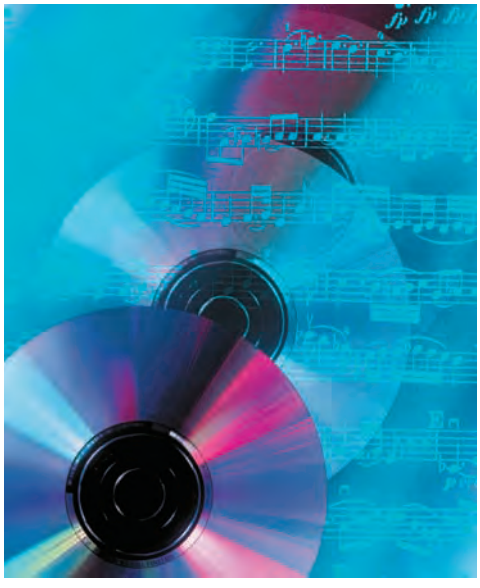
Нотная запись для музыканта представляет последовательность команд, например нажать клавишу и удержать её, нажать одновременно несколько клавиш, отпустить ранее нажатую клавишу и т. п. Если выписать все мысли-

мые команды, то получится система команд воображаемого исполнителя.

В 80-х гг. XX в. появились электронные музыкальные инструменты — синтезаторы, способные воспроизводить звуки многих существующих музыкальных инструментов, а также и абсолютно новые звуки. Фирмы, изготовляющие электронные музыкальные инструменты, стали выпускать *секвенсоры* — приборы, записывающие команды с клавиш, а затем их воспроизводящие. Но единого формата записи ещё не существовало.

В 1983 г. ведущие производители электронных музыкальных синтезаторов и производители компьютеров договорились о системе команд универсального синтезатора, о том, какими электрическими сигналами будут подаваться такие команды, и даже о разъёмах и кабелях, соединяющих компьютеры и синтезаторы. Это соглашение получило название *стандарта MIDI* (от *англ.* Musical Instrument Digital Interface — «описание цифрового музыкального инструмента»). MIDI стал удобным способом кодирования музыки.

Но программы разных секвенсоров по-прежнему имели различные форматы MIDI-файлов. У каждой фирмы имелся свой закрытый формат файла, который могла прочитать только она. В 1988 г. по предложению



тах. В старых тибетских молитвенных книгах пение обозначалось волнообразной линией, передававшей переливы голоса. Распространившиеся на Руси (с XI в.) знаки носили общее название «крюки» или «знамёна». Среди них помимо простых геометрических знаков — «крюков», «палок» — выделяли и более сложные элементы с характерными названиями: «чаша», «подчашие», «стопица», «сорочья ножка», «светлое облачко», «тёмное облачко» и др. Собственная система певческой нотации, также восходящая к византийской, использовалась и у армян; знаками её являлись так называемые хазы, имевшие преимущественно геометрический характер.

Древние системы нотации отмечали изменение тона, не указывая на конкретную высоту звука, т. е. они не столько записывали мелодию, сколько напоминали её, подсказывая исполнителю дальнейшие действия. Лишь после введения нотного стана и фиксации высоты звука стала возможна собственно запись мелодии в современном понимании.

Выделение цветов спектра является ещё более условным, чем выделение звуков. Цветовой спектр разделяется на отрезки, но обозначение приписывается не точке внутри отрезка, как в случае с музыкальными звуками, а всему отрезку. Всякому, говорящему по-русски, известна фраза, помогающая запомнить последовательность цветов радуги: «Каждый охотник желает знать, где сидит фазан». Однако для англичанина или француза цветов в радуге не семь, а шесть — не разделяются голубой и синий. Это не значит, что они хуже видят, при необходимости голубой цвет можно обозначить, например, по-английски как light blue, но как элемент системы, равноправный синему и красному, голубой цвет не воспринимается.

Во вьетнамском, а также в японском языках для обозначения синего, голубого и зелёного цветов вообще обходятся одним словом, а при необходимости уточняют оттенок, добавляя его название к основному. Так, во вьетнамском языке существует 42 общепринятых оттенка этого основного цвета — xanh, например xanh морской волны, xanh ростков риса. В папуасском языке тангма (остров Новая Гвинея) всего лишь два слова для обозначения основных цветов: одно — для тёплых цветов (жёлтый, оранжевый и красный), другое — для холодных (голубой, фиолетовый, зелёный).

Значимость основных цветообозначений (т. е. длина участка спектра) в разных языках будет различной. Цветообозначения — один из ярких примеров роли языка как универсального классификатора, кодирующего явления окружающего мира с помощью словесных обозначений.

В 1983 г. был представлен первый синтезатор с MIDI — Prophet 600, вскоре вышел MIDI-интерфейс PC с синтезатором Roland. К 1985 г. практически все электромузыкальные инструменты в мире имели MIDI-разъёмы.

фирмы Opcode приняли формат Standard MIDI File. Этот открытый формат файла стали поддерживать все производители программ, использующих MIDI наряду со своими собственными.

Запись музыкального произведения в формате MIDI — последовательность закодированных сообщений синтезатору, разделённых закодированными паузами. Сообщение может быть командой синтезатору (нажать или отпустить определённую клавишу, изменить высоту или тембр звучания и т. д.); описанием параметров воспроизведения (например, значение силы давления на клавишу); управляющим сообщением (например, команда включения полифонического режима, синхронизирующее сообщение и т. д.).

Однако при подобном кодировании нельзя записать вокальные произведения, ведь звучание голоса певца или хора не входит в систему команд синтезатора. В остальном такая система кодирования весьма удобна, так как запись очень компактна и соответствует человеческому слуховому восприятию.

КОДИРОВАНИЕ ФИЛЬМОВ

С 1895 года — года изобретения кинематографа, и до наших дней техническими средствами для производства фильмов являются киносъёмочный аппарат и киноплёнка.

Особенность зрения человека позволяет создавать иллюзию движения частой сменой кадров (более 15 раз в секунду), на которых изображены последовательные фазы движения. На этом принципе основаны и кино, и телевидение, и компьютерное кодирование фильмов.

Разработки видеокодировок привели к созданию VideoCD (форматов для видеокомпакт-дисков) и DVD. Накладно не только хранить «неупакованное» кино, но и воспроизводить его (и записывать, и передавать), так как для этого требуются скоростные каналы связи и высокопроизводительные компьютеры.

Цифровое кодирование (запись) изображения начали использовать совсем недавно, потому что не хватало ресурсов для реализации этой задачи традиционным способом. Для записи 1 с (25 кадров размером 1024 x 768 пикселей) цветного изображения без звука потребуется примерно 60 Мбайт (25 кадров · 1024 · 768 · 3 байт на точку = 58 982 400 байт). При этом будет обеспечено качество, близкое к телевизионному.

Легко подсчитать, что на запись двухчасового фильма потребуется более 400 Гбайт! Хранение такого объёма дорого даже в начале XXI в., несмотря на то что средний объём винчестера настольного компьютера превысил несколько десятков гигабайтов.

Объём записи не очень увеличится, если к изображению добавить звук. Как и для записи на компакт-диске, нужно около 170 кбайт для записи 1 с сопровождения.

Звук записывают независимо от изображения, но при этом важно добиться синхронности при воспроизведении.

Чтобы решить проблему большого объёма информации при записи фильмов, разрабатывают новые носители информации и придумывают хитроумные алгоритмы. Например, можно хранить не кадры, а изменения кадров. Тогда, если картинка статична, сокращается объём, требуемый для записи очередного кадра.

Однако главных успехов учёные добились, учитывая ещё одну особенность человеческого зрения: глаз более восприимчив к яркости отдельной точки, чем к её цветности. Поэтому они предложили на каждые четыре точки хранить по 4 байт яркости (на каждую точку разные) и по 2 байт цветности. Таким образом, необходимый объём сократился в два раза: вместо 12 байт ($12 = 4 \cdot 3$) занимали только 6 байт ($4 \cdot 1 + 2$). Качество записи ухудшилось, но это практически незаметно для человеческого глаза.

Синхронизацию звука и изображения при монтаже обеспечивает щелчок «хлопушки», на которой написаны номера сцены и дубля. Момент щелчка в начале съёмок очередного дубля хорошо виден (так же хорошо слышен звук «хлопушки») на плёнке. Это позволяет легко совместить его с началом записей изображения и звука.

КОДИРОВАНИЕ БУХГАЛТЕРСКОЙ ИНФОРМАЦИИ

Практически любая женщина является экспертом, когда речь заходит о деньгах. Как правило, именно жёны ведут семейный бюджет: планируют траты, учитывают расходы и доходы. Правильная семейная бухгалтерия требует особого внимания, записи вносятся каждый день, за год может быть исписана толстая тетрадь, в которой указано всё: когда, сколько и на что было истрачено.

По существу, тем же целям служит *бухгалтерская отчётность* на предприятии. От семейной бухгалтерии она отличается бóльшим объёмом разнотипной информации и наличием определённых правил ведения бухгалтерского учёта.

В бухгалтерии учитываются деньги и материальные ценности обязательно в денежном эквиваленте. Предприятие использует два типа капитала: собственный и заимствованный, привлечённый *(кредиторская задолженность).* Они и составляют экономические ресурсы предприятия:

Экономические ресурсы =
= Привлечённый капитал +
+ Собственный капитал.

Активы — это денежные средства, готовая продукция (товарные запасы), земельные участки, оборудование и даже нематериальные активы, например патенты и авторские права.

Пассивы отражают долги предприятия, товары, полученные в кредит, деньги, взятые взаймы, задолженности по зарплате сотрудникам предприятия и по налогам государству.

Экономические ресурсы в бухгалтерии называют активами, а привлечённый капитал — пассивами. Поэтому основное уравнение можно записать так:
Активы = Пассивы + + Капитал.

Активы		Пассивы	
Касса	50	Уставный капитал	85
Основные средства	01	Задолженность поставщикам	60
Расчётный счёт	51	Задолженность по оплате труда	70
Готовая продукция	40	Резервный капитал	80

В России сюда же для простоты включают и собственный капитал предприятия, куда входят, например, и уставные взносы владельцев предприятия. Важнейшим понятием бухгалтерии является *баланс*. Это двусторонняя таблица, левая часть которой представляет собой актив, состав и размещение хозяйственных средств предприятия, а правая часть — пассив (или пассив плюс капитал), где группируются все поступления. Сумма активов баланса должна совпадать с суммой пассивов, так как актив и пассив — это фактически две точки зрения на одни и те же средства.

Информация, необходимая для управления и учёта на предприятии, должна накапливаться и всегда быть под рукой у руководителя. Система хранения информации при бухгалтерском учёте состоит из *счетов*. Счёт создают на каждый вид актива, пассива или капитала, включая доходы и расходы. Соответственно счета используются для учёта активов или пассивов. Каждый счёт имеет свой шифр, который можно писать вместо названия счёта. В небольшой фирме достаточно иметь в плане несколько десятков счетов, в крупной корпорации их тысячи.

Когда бухгалтерский учёт ведётся вручную, на каждый счёт выделяют отдельную страницу или карточку.

И. Г. Тишбейн. Иоганн Вольфганг Гёте в римском предместье. 1787 г.

Великий немецкий поэт Иоганн Вольфганг Гёте с иронией относился к бухгалтерам, называя двойную бухгалтерию «одним из самых замечательных изобретений человеческого разума», хотя это высказывание и не лишено здравого смысла.

Система двойной записи основывается на принципе двойственности. Все экономические явления имеют два аспекта — увеличение и уменьшение, компенсирующие друг друга, как будто подчиняясь «закону сохранения».

В этой системе факт хозяйственной жизни должен быть зафиксирован как минимум дважды: по дебету одного счёта и кредиту другого — так, чтобы общая сумма по дебету уравновешивала общую сумму по кредиту.

Запись в таблице, похожей на букву «Т», включает название счёта, дебет (левая часть) и кредит (правая часть). Пусть в кассу поступают деньги (левая часть) и из неё производятся выплаты (правая часть):

Касса	
10 000	5000
3000	1200
1000	
14 000	6200
Сальдо	7800

Активные счета имеют дебетовое сальдо, которое показывает остаток средств, а пассивные счета — кредитовое сальдо, учитывающее все источники поступлений. По кредиту пассивных счетов отражается увеличение, по дебету — уменьшение этих источников.

Каждая хозяйственная операция показана в дебете одного и в кредите другого счёта в одинаковой сумме (проводка). Это и называется способом *двойной записи*. Между счетами возникает взаимосвязь: один счёт по дебету связан с другим счётом по кредиту. Например, при начислении из-

носа основных средств дебетуется счёт 80, а кредитуется счёт 02. Можно представить себе, что деньги вычитаются из прибыли и откладываются для будущего ремонта или замены изношенного оборудования.

При операциях суммы на счетах изменяются, однако равенство актива и пассива не нарушается. По каждому счёту ведётся ведомость, в которой фиксируются приход и расход средств. В конце каждого месяца на основании ведомостей составляются *журналы-ордера,* где учитываются суммарные приход и расход средств на каждый счёт или группу счетов. На основании журналов-ордеров в конце отчётного периода вносятся записи в *Главную книгу* и составляется *баланс предприятия.*

Это колоссальная по объёму работа: в бухгалтерии могут вестись сотни счетов, по каждому из них в течение месяца бывает несколько сотен и даже тысяч проводок.

Поддержка равенства актива и пассива баланса, заполнение журналов-ордеров требуют огромных усилий. Если добавить ещё учёт документов, на основании которых делаются проводки, составление платёжных поручений, ведомостей зарплаты, приходных и расходных ордеров, то станет понятным, что ведение бухгалтерского учёта — кропотливая и ответственная работа. Это именно та область, где компьютер способен помочь человеку. Какие же бухгалтерские задачи лучше поручить компьютеру?

Прежде всего, компьютер можно использовать как хранилище информации о движении средств. Про каждый счёт нужно помнить следующую информацию: шифр счёта, название счёта, тип (актив или пассив), остаток (текущее количество средств на счёте). Информацию о счетах помещают в таблицу. Пусть в кассе предприятия находится 11 тыс. рублей:

Шифр	Название	Тип	Остаток
50	Касса	Актив	11 000

Дебетуемый и кредитуемый счета не могут быть выбраны произвольно. Существуют строго определённые типы проводок, которыми бухгалтер имеет право пользоваться. Поэтому при кодировании бухгалтерской информации необходимо записать и этот список, например:

Тип	Название	Шифр дебета	Шифр кредита
1	Продажа готовой продукции	40	51
2	Снятие денег с расчётного счёта	51	50
3	Начисление износа основных средств	80	02

Слова «дебет» (*лат.* debet — «он должен») и «кредит» (*лат.* credit — «он верит») являются просто бухгалтерскими терминами для обозначения правой и левой частей счёта, а не для обозначения увеличения или уменьшения.

Система двойной записи возникла в эпоху Ренессанса. Ещё в 1494 г. францисканский монах Лука Пачоли, друг Леонардо да Винчи, дал первое описание двойной записи.

Лука Пачоли.

где *тип* — это своего рода шифр проводки, по которому из других таблиц можно ссылаться на таблицу типов проводок, *название* — полное название проводки, *шифр дебета* и *шифр кредита* — шифры счетов, участвующих в проводке.

Про каждую проводку важно знать: её тип, когда произошла и какое количество средств было переведено. Если 31 декабря 2002 г. проводкой номер 23450 начислили износ в размере 10 тыс. рублей, то запись в таблице проводок выглядит так:

Номер	Тип	Дата	Количество
23450	3	31.12.2002	10 000

Разумеется, объём хранимой в бухгалтерии информации куда больше. Существуют первичные документы, на основании которых сделаны проводки, отчётная информация за определённый период времени и многое другое. Но даже по предложенной модели легко подсчитать итоговый оборот средств по любому счёту за любой период времени. Независимо от количества проводок компьютер выполнит эту задачу без ошибок!

Программа бухгалтерского учёта способна облегчить и повседневный труд работников бухгалтерии. Например, при выдаче денег из кассы помимо бухгалтерской проводки необходимо заполнить расходный ордер. Данную работу также выполняет компьютер: бухгалтер указывает сумму выдачи, а компьютер делает проводку и печатает ордер. Это не только экономит время, но и гарантирует, что сумма, указанная в ордере, совпадает с суммой проводки.

СПОСОБЫ СЖАТИЯ ИНФОРМАЦИИ

Ресурсы надо экономить, и не только природные. Это в полной мере относится и к памяти ЭВМ, которую используют для хранения и передачи информации. Часто закодированную информацию можно преобразовать так, чтобы в результате она занимала гораздо меньше места. Такой процесс и называют упаковкой или сжатием информации.

При этом целью является именно экономия, а не желание спрятать информацию, зашифровать.

Технология сжатия информации делится на два больших класса — без потери информации, при этом исходное сообщение можно точно восстановить по упакованному, и сжатие с потерей, в этом случае распакованное сообщение будет отличаться от исходного. Как впоследствии окажется, такая потеря не всегда бывает фатальной, т. е. потерянной информацией можно было пренебречь. Однако во многих случаях внесение искажений нежелательно или вообще недопустимо, это характерно, например, для программ или текстов.

На сегодняшний день разработано много алгоритмов упаковки информации без потерь, но практиче-

АЛГОРИТМ ОБОБЩЁННОГО RLE-КОДИРОВАНИЯ

```
алг кодирование RLE
 дано текст
 надо напечатать коды RLE
нач цел счетчик, лит символ1, символ2
| счетчик:=0
| символ1:=прочитать из текста
| нц пока не конец текста
| | символ2:=прочитать из текста
| | если символ1=символ2 и счетчик<255
| | | то счетчик:=счетчик + 1
| | иначе
| | вывод счетчик, символ1
| | счетчик:=0
| | символ1:=символ2
| | все
| кц
кон
```

ски все они основаны на одной из двух простых идей.

Первая идея была предложена Дэвидом Хаффманом в 1952 г. и базировалась на том, что в обычном тексте частоты появления разных символов различны. При стандартном кодировании текста каждый символ кодируется одним байтом. Использование одного байта на один символ упрощает обработку текста. Но перед длительным хранением (или перед передачей по каналам связи) можно позволить себе более сложную кодировку текста. При упаковке по методу Хаффмана часто встречающиеся символы кодируются короткими последовательностями битов (короче 8), а более редкие — длинными (может быть, более 8). В результате в среднем получается менее 8 бит на символ.

Это легко проиллюстрировать. Пусть в тексте из 1000 байт 50 % пробелов. Тогда можно закодировать текст в виде последовательности двоичных 0 и 1.

Если в тексте встретился пробел, то в конец последовательности пишется 0, а если встретился непробел, то запишется 1, а за ней добавится двоичный код этого символа, т. е. 1 + 8 бит. Поскольку текст наполовину состоит из пробелов, то на их кодировку тратится 500 раз по одному биту, а на каждый из 500 непробелов расходуется по 9 бит.

Всего на весь текст будет потрачено 500 + 500 · 9 = 5000 бит, что значительно меньше, чем 8000 бит в исходном тексте.

Вторая основная идея упаковки состоит в том, что в сообщениях часто встречаются несколько подряд идущих одинаковых байтов, а некоторые последовательности байтов повторяются многократно. При упаковке графической информации чаще встречается первая ситуация, а при упаковке текстов — вторая, так как в русском языке редко встречаются две, а тем более большее число подряд идущих одинаковых букв. На рисунке, например, голубое небо использует один цвет, который многократно повторяется при кодировании этого изображения. Да и для хранения графической информации требуется значительно больше места, поэтому задача её упаковки наиболее важна. Графическая информация вообще очень редко хранится в компьютере в неупакованном виде.

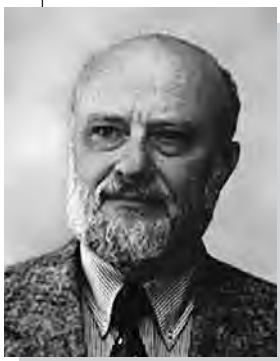
ГРУППОВОЕ КОДИРОВАНИЕ

RLE-кодирование (*англ.* Run-Length Encoding — «кодирование путём учёта числа повторений») — метод, применимый не только к изображениям,

Идеи Хаффмана используют сжатие CCITT (аббревиатура Международного комитета по телеграфии и телефонии) при передаче факсов.

Дэвид Альберт Хаффман.

Существует модификация алгоритма LZSS, который часто вместо символа в триаде хранит указатель на символ, если указатель занимает меньше места, чем байт.



Абрахам Лемпель.

Джекоб Зив.

но и к произвольным сообщениям, он позволяет компактно кодировать длинные последовательности одинаковых байтов, например:

АААААААААААААА

Такая группа обычно кодируется при помощи двух байтов. Первый байт содержит число символов в группе, а второй — повторяющийся символ:

15 А

Другой пример: ФФФАААААКУУУУК потребует уже не двух, а 10 байт:

3	Ф	4	А	1	К	4	У	1	К

Таким образом, можно сказать, что последовательность одиночных, неповторяющихся байтов при RLE-упаковке займёт даже больше места, чем изначально.

Можно предложить следующую модификацию метода RLE.

Пусть первый байт несёт не только информацию о числе повторений, но и о том, есть ли эти повторения. Если старший (самый левый) бит равен 1, то следующий байт данных надо при распаковке повторить столько, сколько записано в оставшихся 7 бит.

Например, байт 10000101_2 говорит, что следующий за ним байт нужно повторить 5 раз (так как $101_2 = 5$). Напротив, если старший бит первого байта равен 0, то просто надо взять несколько следующих байтов данных без всяких изменений. Сколько именно — опять записано в оставшихся 7 бит. Например, байт 00000011_2 говорит, что следующие за ним 3 байт нужно взять без изменений.

Например, на последовательность ФФФАААААКАЮК потребуется уже не 12 байт, а всего 9:

10000011_2	Ф	10000100_2	А	00000100_2	К	А	Ю	К

Алгоритмы RLE-кодирования очень просты и быстры, но эффективность сжатия, т. е. отношение числа байтов в исходной последовательности к числу байтов в полученной после сжатия, сильно зависит от того, что сжимается. Если попытаться упаковать чёрно-белое изображение, полученное сканированием с листа книги, то эффективность будет весьма высока, так как белого цвета значительно больше, чем чёрного, т. е. изображение включает большие непрерывные объёмы данных одного цвета. Однако сложные изображения с большим количеством цветов кодируются плохо, так как имеют мало групп одинакового цвета.

LZW-КОДИРОВАНИЕ

Схема сжатия Лемпеля — Зива — Уэлча (LZW) является одной из самых распространённых при упаковке изображений. В 1977 г. Абрахамом Лемпелем и Джекобом Зивом был создан первый алгоритм LZ77. Описанный ими алгоритм работал примерно так. Сжатие происходит за счёт замены уже встречавшихся ранее в последовательности одинаковой группы символов тройкой значений *(триадой):* указателя на группу, длину совпадающей группы и первого отличающегося символа, идущего за группой.

Максимальная длина фрагмента *F* определяется равной 10–20 символам. А «скользящее» окно, через которое просматривается текст, обозначалось *N*.*

При очередном шаге самая длинная последовательность в фрагменте «ло ко д» ищется в буфере «ышко весело пошло ко д». При этом найденная совпадающая группа символов может перекрывать фрагмент *F,* но, естественно, не совпадать с ним. В примере такая группа символов найдена — «ло_» (с пробелом в конце), она и заменится при сжатии на триаду (7, 3, к).

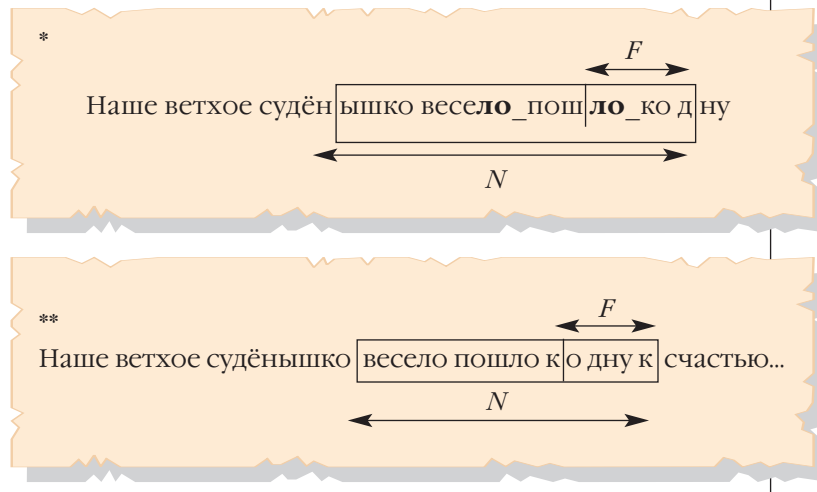
После этого окно сместится на длину группы + 1 (3 + 1 = 4) символов. **

Весьма вероятно, в буфере не будет найдено даже начало фрагмента, т. е. фрагмент начинается с символа,

просто отсутствующего в буфере. Тогда по правилам алгоритма этот символ будет заменён на триаду (0, 0, отсутствующий символ) и окно передвинется на 1.

В приведённом примере $F = 7$, $N = 22$, поэтому для кодировки в двоичном представлении триады потребуется всего два байта — один байт на символ, другой сразу на два числа. На длину группы не более 7, т. е. 3 бит, на указатель $N - F = 22 - 7 = 15$, т. е. 4 бит. Итого 3 + 4 = 7 бит, т. е. даже меньше байта. А для хранения всего окна просмотра требуется N байт. То есть метод LZ77 экономичен по использованию памяти ЭВМ.

Лемпель и Зив показали, что при достаточно большом N их алгоритм даёт очень хорошие результаты, а через год они представили новый алгоритм LZ78, который использовал отдельные словари. При этом повторяющиеся группы помещались в этот словарь, который добавлялся к сжатому файлу, а в кодируемом тексте устанавливались указатели на соответствующую сохранённую группу. Этот алгоритм и лёг в основу созданного в 1984 г. Терри Уэлчем, сотрудником фирмы Unisys, нового алгоритма, который стал носить название LZW.



АЛГОРИТМ LZW

Метод LZW обычно использует словарь, содержащий 4096 элементов. Элементы с номерами 0–255 содержат все 256 символов из 8-битной кодировки. Эти элементы словаря всегда постоянны, а остальные элементы с номерами 256–4095 изначально пусты, но в процессе работы заполняются словами по следующему алгоритму:

```
алг кодирование LZW
дано в словаре 256 символов
надо напечатать коды LZW
нач лит слово, символ
| слово:=""
| нц
| | символ:=Символ из текста
| | если Есть в словаре(слово + символ)
| | | то слово:=слово + символ  | дописывание
| | |                           символа к слову
| | | иначе вывод Код(слово)   | код из словаря
| | | Добавить к словарю (слово + символ)
| | | слово:=символ
| | всё
| кц
кон
```

Когда словарь переполняется, алгоритм Уэлча его очищает, как бы начиная работу сначала.

Существуют модификации этого алгоритма, сбрасывающие при переполнении не весь словарь, а только какую-то его часть, например первую половину.

Алгоритм распаковки работает по такому же принципу, но несколько сложнее:

```
алг декодирование LZW
дано в словаре 256 символов
надо напечатать раскодированные символы
нач лит слово, символ, цел к, к0
| к0:=Прочитать код
| вывод Словарь(к0)
| нц
| | к:=Прочитать код
| | слово:=Словарь(к0)
| | если Словарь(к)<>Пусто     | Слово есть в
| |                              словаре?
| | | то вывод Словарь(к)
| | | символ:=Первый символ в Словаре(к)
| | | иначе символ:=слово[1]   | первый символ
| | всё
| | Добавить к словарю (слово + символ)
| | к0:=к
| кц
кон
```

Бенуа Мандельброт.

В отличие от LZ77 этот алгоритм в процессе своей работы строит словарь, в который попадают специальным образом построенные слова. Если в тексте встречается группа, совпадающая со словом в словаре, то вместо неё записывается индекс этого слова в словаре.

Метод LZW имеет преимущество перед большинством словарных упаковщиков, так как словарь создаётся как на этапе упаковки, так и на этапе распаковки. Все алгоритмы данного семейства отличаются друг от друга способом образования словаря и своим поведением при переполнении словаря.

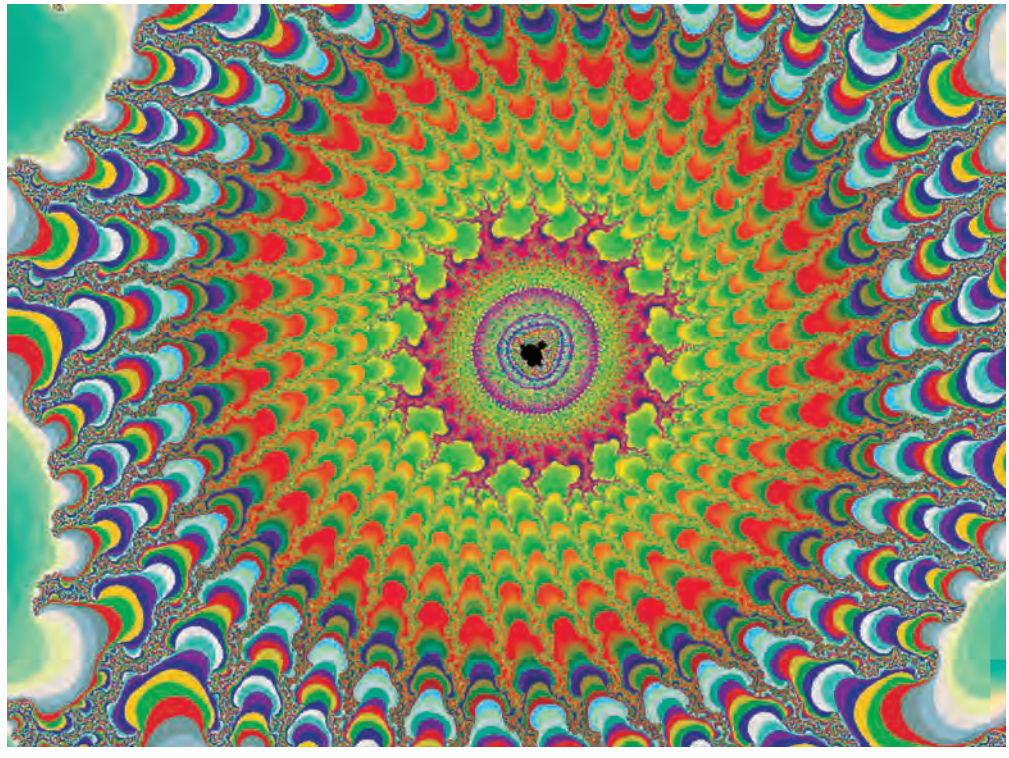
Метод LZW получил широкое распространение, однако, так как патент на него принадлежит Unisys Corporation, его использование в программах требует получения соответствующей лицензии.

ФРАКТАЛЬНОЕ СЖАТИЕ

Упрощённо *фрактал* — это структура, состоящая из подобных форм и рисунков разных масштабов и размеров. Этот термин впервые применил Бенуа Мандельброт для описания повторяющихся рисунков, которые он нашёл во многих различных структурах. Он обнаружил, что фракталы можно описать математически и создавать при помощи очень простых алгоритмов.

Если взглянуть на пол, то в его рисунке нетрудно обнаружить множество похожих повторяющихся фрагментов, при этом совершенно неважно, какая поверхность у пола. Он может быть паркетный, бетонный, покрытый ковром, всё равно рисунки будут повторяться размером от очень маленьких до очень больших. Если «скопировать» кусок поверхности, то можно найти на полу ещё несколько почти таких же. Если изменить размер копии, повернуть или произвести зеркальное отражение, то число похожих частей ещё возрастёт. Тогда можно перечислить все похожие области и все (математические) преобразования с копией, которые были проделаны. Полученные данные (системы уравнений) обычно назы-

Фрактальное изображение.

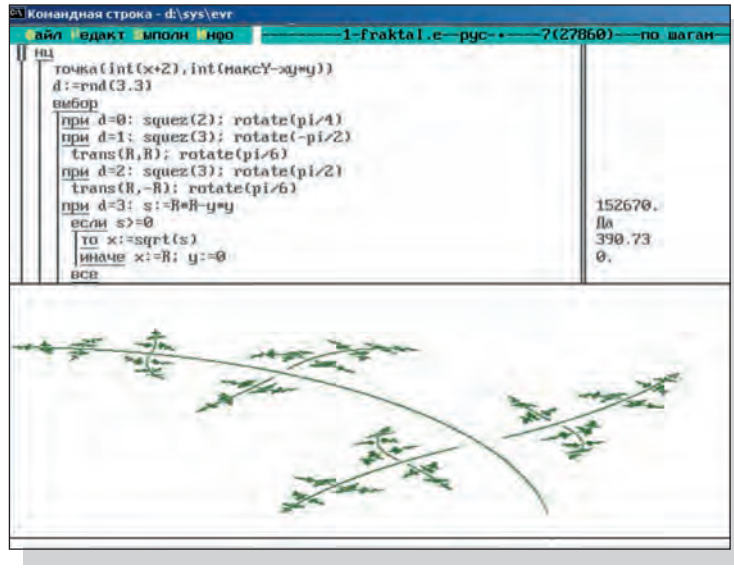


вают *фрактальными кодами*. При этом говорят, что они описывают поверхность через её фрактальные свойства. Используя их, можно воссоздать поверхность пола.

Поиск в изображении фрактальных рисунков требует чрезвычайно большой вычислительной работы. При этом фрактальное сжатие сопровождается потерями, так как процесс сравнения фракталов не предусматривает поиска точного их соответствия. Вместо этого ищется наилучший из возможных вариантов по времени кодирования, качеству соответствия оригиналу и размеру полученных данных. Как правило, требуется, чтобы результат был практически неотличим «на глаз» от первоисточника.

Но результаты дают огромный выигрыш. При распаковке фрактально-сжатого изображения требуется значительно меньше времени, чем при его запаковке. Иногда фрактальное изображение занимает памяти меньше своего оригинала в 100 и более раз. Кроме того, фракталы легко масштабировать без потери деталей и появления лишних деталей (артефактов).

Однако, как всегда, большинство алгоритмов фрактального сжатия запатентованы и закрыты для свободного использования в программах.



Программа на КуМире рисует фрактал.

СЖАТИЕ JPEG

Аббревиатура JPEG происходит от названия комитета по стандартам Joint Photographic Experts Group (Объединённая группа экспертов по фотографии). Алгоритмы сжатия LZW мало подходят для сжатия изображения, так как в цифровых фотографиях и отсканированных картинках присутствуют шумы, и их тем больше, чем больше глубина цвета (число бит на пиксель) при получении цифрового изображения.

Фактически JPEG не является одним алгоритмом сжатия, это целый набор методов сжатия. Как и при фрактальном сжатии, здесь также происходит потеря данных, тех, которые трудно заметить визуально. Небольшие изменения цвета не сильно заметны глазу человека, а вот малые изменения яркости, напротив, легко различимы. Основываясь на этом, схема JPEG сохраняет полутона изображения, но более «свободно» обращается с цветом. Сжатие эффективно, только если различия между соседними пикселями весьма незначительны. JPEG хорошо работает с изображениями глубиной более 5 бит на цветовой канал (RGB). При этом сам стандарт описывается для глубины в 8 бит.

Процесс сжатия происходит в несколько этапов. На первом этапе цветное изображение надо преобразовать из стандарта: например, RGB (красный, зелёный, синий) в YUV (яркость/цветность) или YC_bC_r (яркость и две цветоразностные компоненты).

При дальнейшей обработке яркость надо как можно тщательнее сохранить, при этом цветность менее существенна. Специфической чувствительностью глаза пользуются, уменьшая количество пикселей для каналов цветности, это называется *субдискретизацией*. Для обоих каналов цветности при субдискретизации может быть использована

Вы овладели теорией сжатия информации, если воспринимаете пакет апельсинового сока, как ZIP-файл кучки апельсинов.

Ещё в 1860 г. немецким учёным Густавом Фехнером был сформулирован закон (носящий его имя), подтвердивший, что связь между раздражителем и человеческим восприятием нелинейна: «Ощущения пропорциональны отношению логарифмов стимула».

схема 4 – 2 – 2, хорошо опробованная ещё на заре цветного телевидения в стандарте NTSC (Национальный комитет по телевизионным стандартам США). При этом яркость сохраняется для всех точек квадрата 2×2, а цветность одинакова в горизонтальных рядах.

Обратное косинус-преобразование соответственно совершает обратное: из частотного представления получается пространственное.

Далее картинка разбивается на блоки по 8×8 пикселей (при этом каждый цветовой и яркостной компонент обрабатывается независимо). JPEG основан на так называемом дискретном косинус-преобразовании (DCT), являющемся разновидностью преобразования Фурье. Оно позволяет переходить от пространственного представления изображения к его так называемому *спектральному* представлению, хорошо знакомому каждому, у кого есть музыкальный центр со спектроанализатором, где на маленьком экране красиво прыгают столбики в такт музыке.

DCT-преобразование применяется к каждому блоку 8×8. DCT-вычисления чрезвычайно сложны, и это наиболее трудоёмкий этап всего алгоритма сжатия JPEG. Выполнив его, получим разделение высокочастотной и низкочастотной информации, из которой состоит изображение. Сам по себе этот этап преобразования не предусматривает потерь, за исключением ошибок округления.

Элементы матрицы (нового блока) вычисляются по формуле, представленной ниже (для всех u и v от 0 до 7).

$$F(u,v)=\frac{1}{4}C_u C_v \sum_{i=0}^{7}\sum_{j=0}^{7} f(i,j)$$

$$\cos\frac{(2i+1)u\pi}{16}\cos\frac{(2j+1)v\pi}{16},$$

$$\text{где } C_u, C_v = \begin{cases} 1, & \text{при } u, v \neq 0, \\ \frac{1}{\sqrt{2}}, & \text{при } u, v = 0. \end{cases}$$

При более или менее равных коэффициентах фрагмента 8×8 перед DCT-преобразованием результатом будет матрица с тенденцией уменьшения абсолютных значений от левого верхнего угла к правому нижнему.

На следующем этапе, прежде чем отбросить больший объём информации, надо поделить все коэффициенты полученной матрицы на *коэффициенты квантования,* округляя результаты до целого. Матрица квантования представляет собой целые положительные коэффициенты, растущие от правого нижнего угла к левому верхнему.

Чем больше коэффициенты квантования, тем больше данных теряется. Каждый из 64 элементов матрицы имеет свой коэффициент квантования. Яркостные данные имеют меньшие коэффициенты квантования, чем такие же коэффициенты для цветности. То есть данные яркости сохраняются более тщательно. «Качество» JPEG-сжатия напрямую зависит от матриц квантования.

Таким образом, в результате квантования получится матрица с большим количеством нулевых элементов, имеющая в левом верхнем углу некоторое число ненулевых. На последнем этапе, на так называемом методе вторичного сжатия, при помощи хорошо знакомого метода Хаффмана избавляются от нулевых коэффициентов, обходя матрицу диагональными полосками, начиная с верхнего левого угла:

15	1	2	...
2	–1	0	...
1	0	0	...
...	...	...	...

Тогда в начале последовательности в основном будут ненулевые коэффициенты, в конце же — одни нули. Такие последовательности хорошо сжимает метод Хаффмана.

При восстановлении изображения весь процесс проходит в обратном направлении. Сначала декодирование по методу Хаффмана, далее домножение элементов матрицы на коэффициенты матрицы квантования. Потом применяются обратное косинус-пре-

образование и пересчёт в схемы RGB. По объёму вычислений восстановление изображения, JPEG-декодирование, сравнимо с JPEG-кодированием, поэтому для ускорения этих операций применялись аппаратные ускорители, а фирма Intel, выпустив на рынок процессор Pentium, тут же специально разработала его модификацию MMX (Multi-Media eXtension) — мультимедийное расширение, которое «умеет» выполнять некоторые операции с матрицами 8 × 8, ускоряющие прямое или обратное DCT-преобразование.

СЖАТИЕ ЗВУКА

Цифровой звук, если это не музыка, которую можно закодировать в виде MIDI, столь же неудобен для сжатия, как и картинка. Звуковой сигнал редко обладает избыточностью, т. е. имеет повторяющиеся участки (в основном из-за шумов). А значит, плохо сжимается с использованием алгоритмов компрессии без потерь, аналогичных LZW или методу Хаффмана.

В два раза звук можно упаковать с помощью метода со странным названием *компандирование* (от *англ.* compound — «соединение», «составной»).

Этот метод основан на законе, открытом психологами: если интенсивность раздражителя меняется в геометрической прогрессии, то интенсивность человеческого восприятия меняется в арифметической прогрессии.

В применении к звуку это означает, что если повысить громкость звука в 2, 4, 8 и более раз, то человеческое ухо будет воспринимать это как линейное увеличение интенсивности. То есть изменение громкости от 1 до 2 столь же заметно человеку, как и изменение громкости от 100 до 200, а изменение громкости от 100 до 101 человеком практически не ощущается.

Поэтому при компандировании значения амплитуды звука заменяются на логарифм этих значений.

При 8-битном кодировании звука $|a|<2^7$, следовательно, $\log_2|a|<7$. Значит, звук может быть закодирован тремя двоичными разрядами плюс ещё один разряд на знак амплитуды, итого получится 4 вместо 8.

То есть при компандировании 8-битного звука при некотором ухудшении качества звука происходит сжатие вдвое.

АДАПТИВНАЯ РАЗНОСТНАЯ КОМПРЕССИЯ

Один из способов уменьшения объёма аудиоданных — хранить не сами данные, а разность между двумя

Звук на CD, как правило принимаемый за эталон цифрового звучания, закодирован способом импульсно-кодовой модуляции, PCM (Pulse Code Modulation). Время звучания и диапазон амплитуды разбивались на соответствующие равные промежутки, являющиеся параметрами оцифровки, — частоту дискретизации и уровень дискретизации (число битов на канал). В результате на CD (при частоте около 44 кГц и 16 бит на канал) 1 с стереозвука требовала около 1,4 Мбит двоичных данных, или, как говорят, воспроизведения (запись) потоком 1400 кбит/с.

соседними значениями, если такая разность невелика. Как правило, это верно для человеческой речи. При этом разность кодируется меньшим числом битов, нежели исходные данные. Такой метод называется DPCM (Delta Pulse Code Modulation).

В другом, более экономном способе ADPCM (adaptive DPCM, или адаптивная разностная компрессия), значения разностей подвергаются ещё более грубому квантованию (число возможных значений), что позволяет уменьшить разрядность до 4 бит вместо 16, причём шаг квантования выбирается в зависимости от величины текущего изменения сигнала. Он увеличивается, если несколько раз подряд разность достигала 16 (4 бит выделено для квантования разности), и уменьшается, если разность несколько раз подряд была меньше 4. Именно в этом заключается «адаптивность», на которую указывает название метода. При этом принято говорить, что коэффициент сжатия 16-битных данных методом ADPCM равен 4 : 1.

Существуют методы сжатия особого вида звука — человеческой речи, основанные не на том, как слышит человек, а на том, как он говорит.

Качественно иные принципы лежат в основе алгоритмов сжатия по стандарту MPEG (Motion Pictures Experts Group).

MP 3

Естественно возникает вопрос: а нельзя ли воспользоваться каким-нибудь методом, аналогичным JPEG-преобразованию в графике? Ведь при сжатии картинки искажения, часто незаметные для глаза, позволяют сильно сократить размер изображения. Пусть в результате компрессии и декомпрессии звук не будет идентичным, но и не надо ставить перед собой цели абсолютно точного восстановления формы исходных звуковых колебаний. Главная задача — это максимальное сжатие звукового сигнала при минимальных слышимых (или вообще неслышимых) искажениях. Человеческий слух имеет ряд особенностей, позволяющих использовать достаточно эффективные алгоритмы компрессии звуковых данных при субъективно-минимальных потерях в качестве звука.

В 1940 г. Харви Флетчер, выдающийся американский физик, отец стереозвука, привлёк для исследований человеческого слуха большое число испытуемых. Он проанализировал зависимость абсолютного порога слышимости от частоты сигнала, т. е. при какой амплитуде звук определённой частоты не слышен для человека. В построенной на основе

опытов кривой максимальные значения порога находятся, как и ожидалось, на границах диапазона слышимости (около 20 Гц и ближе к 20 кГц), а минимум — приблизительно 5 кГц. Но главное, на что он обратил внимание, — это способность слуха адаптироваться к появлению новых звуков, что выражается в повышении порога слышимости. Иначе говоря, одни звуки способны делать неслышимыми другие, что и называют маскированием одного звука другим.

Маскирование имеет как частотное (или частотно-динамическое) свойство (громкие звуки маскируют близкие им по частоте более тихие, низкочастотные звуки маскируют высокочастотные), так и временно́е (эффект продолжается ещё некоторое время после того, как маскирующий звук прекратился).

Последнее свойство слуха при компрессии позволяет после громкого звукового сигнала некоторое непродолжительное время вообще не воспроизводить, а значит, и не сохранять никакого звука. Для грубой демонстрации этого эффекта можно привести такой пример: после выстрела из пушки в течение некоторого времени человек вообще ничего не слышит. То есть громкий щелчок продолжительностью 0,1 с может замаскировать последующие за ним тихие звуки на 0,5 с, которые не надо сохранять. Для точного воспроизведения такого звука для человека достаточно проиграть только малую часть от исходного сигнала — щелчок, так как оставшаяся часть всё равно не слышна. Говорят, что коэффициент компрессии в этом примере достигает

$$6 : 1 = \frac{0{,}5+0{,}1}{0,\,1},$$

а описанную процедуру сжатия обычно называют *маскированием во временной области*.

При *маскировании в частотной области* синусоидальный сигнал маскирует более тихие, близкие по частоте сигналы, в том числе и синусоидальные сигналы много меньшей амплитуды. Это даёт возможность выделить в исходном спектре сигнала

То же произойдёт, если проиграть маскирующий тон с частотой 1 кГц и уровнем 60 дБ и тестовый тон с частотой 1,1 кГц и уровнем 40 дБ, который при этом не будет слышен. Можно исключить маскирующий тон, оставив звучать тестовый на короткое время, пока он не станет вновь слышен. Эксперимент покажет, что это время равно примерно 5 мс.

Естественно, что те части спектра звука, которые лежат ниже абсолютного порога слышимости (минимальное значение звукового давления, которое способно воспринять человеческое ухо), ни кодировать, ни передавать не следует.

те части, которые практически не будут слышны. Удобно использовать разбиение спектра на полосы различной ширины, основываясь на особенностях слуха человека: в ушной улитке существуют области, каждая из которых отвечает за определённую частотную полосу (шириной от 100 Гц внизу спектра и до 3500 Гц в верхней части). Обычно выделяют 27 так называемых критических полос (*англ.* critical band): 0-я от 50 до 95 Гц, 1-я от 95 до 140 Гц, ... , 26-я от 20 250 Гц и выше.

Для выполнения алгоритма сжатия исходный сигнал разбивается на кадры, которые подвергаются частотному анализу. Алгоритм сжатия выглядит примерно так.

В 1987 г. при разработке алгоритмов аудиокодирования для цифрового вещания (Digital Audio Broadcasting DAB) при непосредственном участии профессора Дитера Сейтзера из Университета Эрлангена (Германия) был разработан известный стандарт ISO-MPEG Audio layer 3, который обычно и называют MP 3.

• При помощи специальных алгоритмов (ими могут быть быстрое преобразование Фурье или аналогичные) сигнал разделяется на 32 равные полосы спектра, при этом в одну получившуюся полосу могут попасть сразу несколько критических полос.

• Используя так называемую психоакустическую модель (в которую, как правило, и входит частотное маскирование), определяют уровень маскирования полосы соседними.

• Уровень в полосе, не превышающий вычисленный порог, считается равным нулю и не сохраняется. Наоборот, немаскированный уровень записывается в выходные данные.

Так, например:

Полоса №	...	7	8	9	...
Уровень (dB)	...	10	60	35	...

Уровень 8-й полосы равен 60 dB. На основании психоакустической модели делается вывод, что этот уровень может замаскировать сигнал 7-й полосы уровня 12 dB и 9-й полосы уровня 15 dB. Если посмотреть в таблицу, то можно заметить, что 7-ю полосу не надо сохранять, так как 10<12, а 9-ю, напротив, нужно записать, так как 35>15.

В дальнейшем на каждый ненулевой уровень выделяется некоторое число битов, достаточное для его примерного представления. Так, в той части спектра, где человеческое ухо имеет наименьший порог слышимости, информация кодируется шестнадцатью битами, а на краях, там, где ухо менее чувствительно к искажениям, шестью и менее битами. К полученному потоку битов можно, например, применить алгоритм сжатия Хаффмана.

▶ Жан Батист Жозеф Фурье.

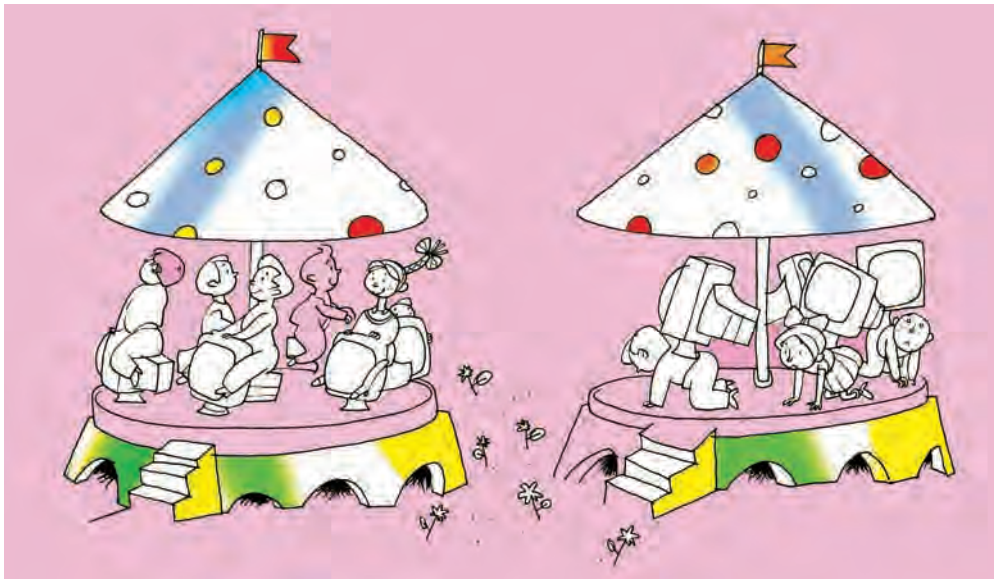
Различают три версии алгоритма, описанного MPEG-сжатием звука. В каждой версии данные разделяются на кадры, т. е. отдельный кадр состоит из 32 полос по 12 значений в каждой: 32 • 12 = 384.

В MPEG layer1 (дословно «слой 1») в частотном фильтре используются один кадр и алгоритмы, основанные на дискретном косинусе — преобразовании (DCT). Психоакустическая модель задействует только частотное маскирование. Алгоритм позволяет упаковывать при соотношении 1 : 4 с потоком 384 кбит/с.

MPEG layer 2 использует три кадра в частотном фильтре (предыдущий, текущий и последующий) общим объёмом 32 полосы по 12 значений в трёх кадрах: 32 • 12 • 3 = 1152. Модель использует и временное маскирование. Упаковывает с соотношением от 1 : 6 до 1 : 8.

MPEG layer 3, он же MP 3, использует частотный фильтр с разной величиной полос; психоакустическая модель включает временное маскирование стереосигнала. Имеет малые потери качества при высоком уровне компрессии: от 1 : 10 до 1 : 12.

MP 3 оказался одним из самых удачных стандартов сжатия звука, что позволило ему не только потеснить с рынка аудиопродукции мини-диски (MD) фирмы Sony, но и стать основным форматом записи для плееров, у которых вообще нет движущихся частей. В формате MP 3 на один CD можно записать от 6 до 12 ч музыки.

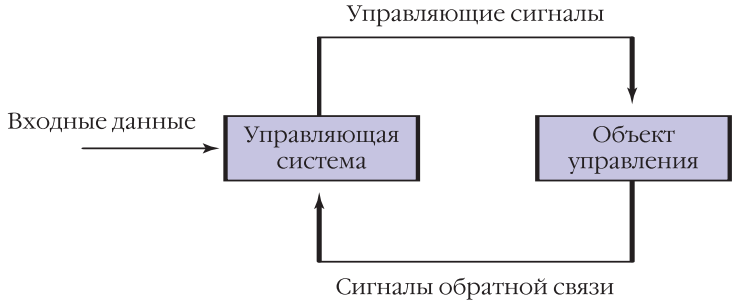


НАУКА ИНФОРМАТИКА

ИНФОРМАТИКА

В 1834 г. великий французский физик Андре Мари Ампер (1775—1836) предложил свою классификацию наук. Для обозначения науки об управлении обществом он использовал термин *кибернетика* (*греч.* «кибернетике» — «искусство управлять»).

Спустя 100 лет выдающийся американский учёный Норберт Винер (1894—1964), изучавший различные технические и биологические системы, обратил внимание, что работу любой *системы управления* можно представить в виде единой схемы:



В 1948 г. он выпустил в свет книгу «Кибернетика, или Управление и связь в животном и машине», в которой обобщил свои наблюдения, сформулировал общие принципы построения и работы управляющих систем, показал ключевую роль информации в них. Так началась вторая жизнь придуманного Ампером слова, ставшего названием новой науки.

Под *связью* кибернетика понимает процессы восприятия, хранения и передачи информации. *Управление* — это процесс воздействия *управляющей системы* на *объект управления*, обеспечивающий его работу по достижению поставленной цели. Оно осуществляется с помощью *управляющих сигналов*, вырабатываемых управляющей системой. Эти сигналы содержат информацию о требуемом поведении объекта управления. Она создаётся на основе анализа *входных данных* (информации о состоянии внешней среды) и сигналов *обратной связи* (несущих информацию о состоянии

объекта управления). Переработку информации, поступившей по каналам обратной связи, называют *контролем*.

Ощущая неполноту своего определения и словно идя вслед за Ампером, Винер в 1954 г. издал книгу «Кибернетика и общество», в которой распространил сферу влияния новой науки на социальные системы. Огромный вклад в пропаганду кибернетики внёс английский учёный Уильям Росс Эшби (1903—1972), систематизировавший идеи Винера в книге «Введение в кибернетику» (1956 г.).

Кибернетика изучает способность машин и живых организмов воспринимать определённую информацию, сохранять её в памяти, передавать по каналам связи и перерабатывать в управляющие сигналы. Задача кибернетики — выделение и изучение общих свойств процесса управления и систем управления.

Кибернетика — это самостоятельная наука со своим предметом исследования (управляющие системы). Как и в других науках, в ней можно выделить важнейший раздел — *теоретическую кибернетику*. Она разрабатывает аппарат и методы исследования, пригодные для изучения систем управления различной природы. Теоретическая кибернетика объединила несколько существовавших ранее разделов математики: математическую логику, теорию алгоритмов, теорию информации, теорию кодирования. Некоторые новые научные направления зародились уже в рамках самой кибернетики, среди них — теория автоматов, теория формальных языков, теория формальных грамматик, нечёткая математика.

Таким образом, кибернетика в значительной степени строится на математической основе. Но она имеет и собственный особый метод исследования. Это *компьютерное моделирование*, позволяющее изучать не объекты, а их описания *(модели)*. Моделирование здесь играет ту же роль, что и эксперимент в физике, химии, других естественных и технических науках. Крайне важно, что моделировать на компьютере можно даже объекты, которые нельзя описать с помощью уравнений или формул. Это ставит кибернетику (как и математику) в особое положение, ведь такой метод применим в самых разных науках. Соответствующие области знаний получили свои названия:

• *Техническая кибернетика,* используя результаты и выводы теоретической кибернетики, разрабатывает и исследует всевозможные технические управляющие системы — от простых систем автоматического регулирования до сложнейших автоматизированных систем управления, построенных на основе суперкомпьютеров.

• *В биологической кибернетике* выделяют несколько разделов: медицинская, психологическая, физиологическая кибернетика, бионика, нейрокибернетика. Все они занимаются моделированием биологических систем и математической обработкой результатов их исследования.

• *Экономическая кибернетика* изучает процессы управления экономикой, моделирует экономические системы.

• *Военная кибернетика* рассматривает общие вопросы управления войсками и методы повышения эффективности применения боевой техники.

• *Социальная кибернетика* исследует модели процессов, протекающих в человеческом обществе.

Информация (от *лат.* informatio — «разъяснение», «изложение»), как и управление, — центральное понятие кибернетики, наряду с материей и

Андре Мари Ампер.

Книга У. Р. Эшби «Введение в кибернетику.

ИНФОРМАТИКА В СССР

Гонения на кибернетику в послевоенные годы в СССР привели к тому, что наша страна по производству вычислительной техники оказалась далеко позади развитых капиталистических стран.

Все фундаментальные исследования, результаты которых в той или иной степени могли использоваться в военной области, автоматически получали гриф «секретно». Это произошло и с книгой американского математика Норберта Винера «Кибернетика, или Управление и связь в животном и машине» (1948 г.): сразу же после того как она попала в СССР, её засекретили.

Идеи, высказанные автором, противоречили официальной доктрине советского общества. Параллели, проведённые Винером между живым миром (человек и животные) и миром машин (моделирование в социальной сфере и экономике), являлись крамольными для граждан Советского Союза, которым внушалась мысль о несводимости «высших форм» к «низшим формам». Именно за это книга и попала в спецхран.

Сейчас трудно представить, сколько молодых и выдающихся людей могли бы стать настоящими учёными и принести пользу родной стране, если бы не те, кто, боясь за свою карьеру и тёплое местечко, развернул их травлю. Звания и должности получали не за талант или научные достижения, а за принадлежность к коммунистической партии и за рабоче-крестьянское происхождение.

Не случайно в философских словарях 50-х гг. кибернетика определялась как реакционная лженаука, возникшая в США после Второй мировой войны и получившая широкое распространение и в других капиталистических странах. В журнале «Вопросы философии» в 1953 г. появилась «заказная» статья под названием «Кому служит кибернетика». В ней одобрялись развитие и производство быстродействующих «арифмометров». Но наряду с этим отмечалось, что использование машин для моделирования и символьной обработки шло вразрез с марксистско-ленинским учением: «Теория кибернетики, пытающаяся распространить принципы действия вычислительных машин новейшей конструкции на самые разные природные и общественные явления без учёта их качественного своеобразия, является механицизмом, превращающимся в идеализм. Это пустоцвет на древе познания, возникший в результате одностороннего и чрезмерного раздувания одной из черт познания». Автор статьи подписался псевдонимом — Материалист. Вдруг установка ЦК изменится и будет дан зелёный свет развитию кибернетики? Именно поэтому в статье развитие компьютеростроения поддержано так аккуратно. К сожалению, подобное поведение являлось нормой.

А. И. Китов.

Однако в учёной среде всегда находились и те, кто во имя идеи, гражданской позиции и науки готов был идти до конца. Сейчас нет смысла перечислять всех, кто боролся за становление кибернетики. Отметим только, что статью С. Л. Соболева, А. И. Китова и А. А. Ляпунова «Основные черты кибернетики», в которой энергией является одной из фундаментальных основ мироздания. Хотя появление ЭВМ (устройств для накопления и переработки информации) и было одним из толчков к возникновению кибернетики, быстро выяснилось, что эта наука не даёт ответа на многие теоретические и практические вопросы, непосредственно связанные с развитием вычислительной техники. Поэтому уже в 40—50-х гг. XX в. начала формироваться новая наука, получившая в англоязычных странах название «Computer Science» (компьютерная наука). Во Франции её назвали «Informatique», объединив слова «Information» и «Automatique».

Судьба новой науки в СССР сначала складывалась весьма непросто. Ведь в первые годы своего развития кибернетика в значительной степени была наукой социальной, философской. Например, в Кратком философском словаре (1954 г.) говорилось: «Кибернетика — это реакционная лженаука, возникшая в США после Второй мировой войны... является... не только идеологическим оружием империалистической реакции, но и средством осуществления её агрессивных военных планов». Тем не менее скоро отношение к кибернетике стало меняться в лучшую сторону. Происходило это не само по себе — огромную роль в пропаганде кибернетики и в её развитии сыграли многие выдающиеся советские учёные: А. А. Ляпунов (1911—1973), А. И. Берг (1893—1979), И. А. Полетаев (1915—1983) и др.

В нашей стране кибернетику стали трактовать весьма широко, включая в неё и все вопросы, связанные с созданием и использованием ЭВМ. Но постепенно происходило обособление этих задач. В результате кибернетика как бы передала информатике задачи, связанные с понятием информации, а за собой оставила решение проблем управления.

Между тем в различных отраслях кибернетики широко используются практические достижения информатики, такие, как программные системы, высокопроизводительные компьютеры, сетевые технологии.

Теоретическую основу информатики составляет математика, и в пер-

вую очередь дискретная математика. Её содержание в значительной мере пересекается с содержанием теоретической кибернетики. Важную роль также играют вычислительная математика, ориентированная на создание методов решения задач с помощью компьютеров, и системный анализ. Информатика взяла из кибернетики и экспериментальный метод компьютерного моделирования, заключающийся в формировании гипотезы, создании модели, проведении экспериментов, сборе данных и анализе результатов.

Кроме изучения процессов манипулирования данными информатика занимается их реализацией и применением в различных сферах деятельности человека. Её важнейшей задачей является *автоматизация* (т. е. исключение участия человека) всех этих процессов. Поэтому инженерную часть информатики составляет создание устройств и систем, предназначенных для решения возникающих при этом проблем (конструирование). Говоря об устройствах и системах, надо иметь в виду их аппаратную составляющую и программное обеспечение.

Характерная особенность информатики — очень тесное взаимодействие её теоретических и инженерных аспектов. Этим она отличается от других наук. Например, в химии чётко выражена граница между теорией и химическим машиностроением. В информатике средства, обеспечивающие эксперимент (компьютеры, компиляторы, прикладные программы), одновременно являются и целью разработки, объектом приложения теории.

Ещё одна отличительная черта информатики — фундаментальный характер понятия эффективности. Она обязательно должна быть обеспечена на всех стадиях информационных процессов.

Как это было раньше с кибернетикой, иногда информатику делят, в зависимости от сферы её использования, на экономическую, правовую, медицинскую и т. д. При этом очень часто такое разъединение не подчёркивает специфики конкретной области применения, и речь идёт

рой они выступили в защиту науки, напечатали в тех же «Вопросах философии» в 1955 г. Это доказывает, что кибернетика не противоречила линии партии, а просто-напросто подрывала позиции функционеров из научной среды.

Отечественные учёные (в первую очередь военные) сумели доказать, хотя и с запозданием, важность развития кибернетики (информатики) и компьютеростроения.

Советские ЭВМ, которые иногда выпускались всего лишь чуть позже западных аналогов, по многим параметрам не уступали им. БЭСМ, «Сетунь» и другие вычислительные машины недаром оставили заметный след в науке и технике. Информатика в нашей стране стала так же сильна, как и отечественная математическая школа. Однако в производстве вычислительной техники, особенно в массовом промышленном компьютеростроении, СССР от Запада сильно отставал.

До тех пор пока вычислительную технику разрабатывали в НИИ и КБ, принималась масса интересных, неординарных научных решений. Но при воплощении их в жизнь или появлялись производственные трудности (особенно в области качества), или сталкивались ведомственные интересы. Ряд министерств осуществлял разработку и выпуск вычислительной техники часто почти идентичной по параметрам и показателям. Здесь, как и при капитализме, действовали жёсткие законы конкуренции. Результатом борьбы стали выпуск устаревшей техники или, что ещё хуже, слепое подражание Западу. Копирование зарубежной вычислительной техники приняло промышленные масштабы. Конечно, совсем неплохо изучать опыт конкурирующих держав. Однако идти у них на поводу, повторяя не только достижения, но и ошибки, — это, несомненно, проигрышная стратегия.

Конец 80-х гг. показал, что персональные компьютеры нужно выпускать не десятками, а тысячами, десятками и сотнями тысяч штук в год. Отечественная компьютерная промышленность была к этому не готова. Хочется верить, что уроки минувшего века не прошли для россиян даром. Но, к сожалению, научный мир слабо подвержен переменам, и то, что произошло с кибернетикой, в нём может повториться вновь.

КРИТИКА БУРЖУАЗНОЙ ИДЕОЛОГИИ

Кому служит кибернетика

Среди современных буржуазных социологических теорий, направленных на защиту капитализма, не последнее место занимают «теории», фетишизирующие технику, пытающиеся изобразить ее основным двигателем общественного развития. Некоторые буржуазные ученые склонны все общественные противоречия, существующие в капиталистическом обществе, все беды и несчастья отнести за счет техники. В «мистической силе» техники они видят причины войн, безработицы, кризисов. Эти «социологи» призывают к разрушению техники и возвращению к идиллическим временам первобытной жизни, когда не было ни машин, ни социальных конфликтов. Другие «социологи» из того же лагеря фетишизируют технику как силу положительную, способную якобы устранить все противоречия капиталистического строя.

Все эти измышления ученых лакеев империализма ничего общего не имеют с наукой и свидетельствуют лишь о вырождении современной буржуазной науки.

Статья в журнале «Вопросы философии». 1953 г.

ДОКУМЕНТАЛИСТИКА

Во многих учебниках можно прочитать, что слово «информатика» в середине 70-х гг. XX в. было заимствовано из французского языка. Однако это не совсем так.

К тому времени научная дисциплина именно с таким названием существовала и успешно развивалась уже более четверти века! Она исследовала структуру и общие свойства научной информации — данных, отображающих объективные закономерности природы, общества и мышления. Кроме того, изучала процессы обмена научной информацией — от устного общения на конференции до публикации в виде статьи или монографии. В практическом плане информатика вырабатывала способы сбора, хранения, поиска, распространения научной информации. Её непосредственной предшественницей была наука о сборе, обработке, хранении, поиске и распространении документов — документалистика, созданная 100 лет назад бельгийским учёным Полем Отле.

Нетрудно заметить связь информатики с делопроизводством, библиотечным и архивным делом, книговедением и т. д. Информатика (Information Science) возникла на их основе в 40—50-х гг., но особенно бурно стала развиваться после появления компьютеров.

Так что пришедшее к нам из Франции слово вовсе не было новым для русского языка, просто его более широкое значение вытеснило старое.

В целом информатику понимают сегодня как науку, которая занимается систематическим изучением процессов создания, описания, хранения, обработки и передачи данных.

просто об использовании компьютеров, баз данных, пакетов прикладных программ (это не относится к бурно развивающимся биоинформатике и нейроинформатике, в которых получено много важных результатов).

Интереснее рассмотреть внутреннюю структуру информатики. Сфера явлений, охватываемых информатикой, неоднородна и весьма сложна. Кроме того, информатика очень динамично развивающаяся наука, и хотя она достаточно молода, за время её существования те или иные разделы уже неоднократно появлялись и исчезали.

Несмотря на то что в разные периоды развития информатики её структура была различной, можно выделить несколько устойчивых областей. Каждая из них имеет свою чётко обозначенную сферу действия, свои сформировавшиеся подходы к решению задач и методы исследования. К ним относятся такие разделы, как алгоритмы и структуры данных; архитектура и организация компьютеров; языки программирования; операционные системы; технология программирования; взаимодействие человека и компьютера.

Технология программирования, в частности, разрабатывает и изучает методы создания больших программных систем, проверки их соответствия заданным требованиям и обеспечения надёжности работы. Сами методы при этом постоянно развиваются, совершенствуются, но цель данной отрасли информатики остаётся прежней. То же можно сказать и о других перечисленных разделах.

Остальные отрасли информатики более подвержены изменениям. Ещё в начале 90-х гг. XX в. в ней выделяли следующие предметные области: базы данных, поиск и восстановление информации; искусственный интеллект и робототехника.

Их важность в информатике не подлежит сомнению и сегодня, но за прошедшее время содержание этих областей так сильно изменилось, что названия целесообразно сформулировать по-иному: управление информацией; интеллектуальные системы.

Появились новые важные области: сетевые технологии и вычисления; компьютерная графика и визуализация.

Человечество вступает в следующий этап своего развития — информационное общество. Внедрение компьютерных технологий на глазах меняет нашу среду обитания и характер социальных отношений. Поэтому особое значение приобретают выделяемые в отдельную область социальные и профессиональные проблемы информатики. К их числу относятся ответственность за последствия принимаемых решений, спосо-

▶ Модель космического корабля «Шаттл».

бы обеспечения неприкосновенности частной жизни и гражданских свобод, противодействие компьютерной преступности, защита интеллектуальной собственности.

Подводя итог, можно сказать, что информатика, несмотря на огромные достижения, очень молодая наука и находится ещё в самом начале своего пути.

ДЖОН ФОН НЕЙМАН

Блестящий математик и физик Джон Льюис фон Нейман родился 28 декабря 1903 г. в Будапеште. При рождении родители дали ему имя Янош, в годы учёбы и работы в Швейцарии и Германии он называл себя Иоганном, а после переезда в США — Джоном. Джон фон Нейман был старшим ребёнком в семье состоятельного венгерского банкира. Свои необычайные способности мальчик проявил очень рано — в 6 лет он владел несколькими иностранными языками (свободно говорил на древнегреческом и часто шутил на нём с отцом) и мог разделить два восьмизначных числа в уме, в 8 — освоил основы высшей математики. В 12 лет легко запоминал несколько страниц текста.

Начальное образование он получил в Будапеште. Когда ему исполнилось 17 лет, отец решил, что сыну нужно заниматься чем-то более выгодным, чем математика, и уговорил его изучать химию. Для этого Джон отправился сначала в Германию, а затем в Швейцарию, параллельно он изучает математику в Будапеште. В 1926 г. Фон Нейман получает диплом химика и степень доктора математики. Некоторое время он преподаёт в Германии, а в 1930 г. его приглашают в Принстонский университет (США). Когда в 1933 г. в Принстоне открылся Институт перспективных исследований, Джон фон Нейман стал его сотрудником и окончательно обосновался в США. Джон был очень общительным человеком: обладая большими знаниями и будучи блестящим рассказчиком, он в любом обществе становился душой компании.

Джон фон Нейман.

Джон фон Нейман внёс большой вклад в создание и развитие целого ряда областей математики и физики. Он проводил фундаментальные исследования, связанные с математической логикой, теорией групп, алгеброй операторов, квантовой механикой, статистической физикой.

В 1944 г. фон Нейман и экономист Оскар Моргенштерн написали книгу «Теория игр и экономическое

Джон фон Нейман пользовался заслуженной любовью и уважением коллег. При доказательстве теорем он исписывал всю доску формулами и потом настолько быстро их стирал, что коллеги в шутку называли его манеру «доказательством методом стирания».

поведение», в которой развили не только математическую теорию игр, но и её применение к экономическим, военным и другим наукам. Также Джон фон Нейман является одним из создателей метода Монте-Карло — численного метода решения математических задач, основанного на моделировании случайных величин. А кроме того, Нейман участвовал в секретном проекте по созданию атомной бомбы.

Во время Второй мировой войны американским учёным пришлось выполнять военные заказы. Джон фон Нейман был направлен в группу разработчиков ЭНИАКа (электронно-вычислительная машина) консультантом по математическим вопросам. Внимательно изучив конструкцию, он пришёл к идее нового типа логической организации ЭВМ:

- устройства ввода/вывода информации;
- память компьютера;
- процессор, состоящий из устройства управления (УУ) и арифметико-логического устройства (АЛУ).

В 1946 г. вместе с Германом Гольдстейном и Артуром Берксом он написал и выпустил отчёт «Предварительное обсуждение логической конструкции электронной вычислительной машины», где высказано множество интересных идей, например были предложены принципы параллельного доступа к памяти. Архитектура первых двух поколений ЭВМ с последовательным выполнением команд в программе получила название «фон-неймановская архитектура ЭВМ».

Одна из ЭВМ, в проектировании и разработке которой фон Нейман принимал участие в 1954 г., сыграла основную роль в обработке информации при создании водородной бомбы. Джон фон Нейман, славившийся своим остроумием, в шутку называл эту машину «Маньяк» (MANIAC — Mathematical Analyzer Numerical Integrator and Computer).

Крупнейшим научным достижением фон Неймана в послевоенный период стало построение теории автоматов. В 1952 г. в работе «Вероятностная логика и синтез надёжных организмов из ненадёжных элементов» он показал путь создания надёжных ЭВМ и других автоматов. (По теории фон Неймана надёжность системы, построенной из ненадёжных элементов, может быть повышена за счёт увеличения числа элементов и соединений между ними.)

В 1954 г. Джон фон Нейман был назначен членом Комиссии по атомной энергии США, и весной 1955 г. он переезжает из Принстона в Вашингтон. Вскоре врачи обнаружили у него форму костного рака. Учёный страдал от невыносимых болей. До последних минут жизни служба безопасности военного ведомства, где он работал над секретными проектами, зорко следила, чтобы, находясь в бреду, фон Нейман не выдал какую-нибудь государственную тайну. Умер Джон фон Нейман 8 февраля 1957 г. в Вашингтоне.

Последняя книга, над которой работал фон Нейман, — «Вычислительная машина и мозг» — так и не была закончена.

Архитектурные принципы организации ЭВМ, указанные Джоном фон Нейманом, долгое время оставались почти неизменными, и лишь в конце 70-х гг. XX в. в архитектуре суперкомпьютеров и матричных процессоров появились отклонения от этих принципов.

НОРБЕРТ ВИНЕР

Имя этого человека, живо интересовавшегося проблемами естествознания, стало известно всему миру после выхода в 1948 г. его книги «Кибернетика». Норберт Винер — выдающийся американский математик, внёсший заметный вклад в теорию связи, участвовавший в создании первых вычислительных машин.

Ничто не предвещало такого сенсационного успеха книги. В 1946 г., находясь в Париже на конференции по математике, Винер встретился с издателем, который предложил ему написать книгу об обратной связи между автоматами (машинами) и нервной системой человека.

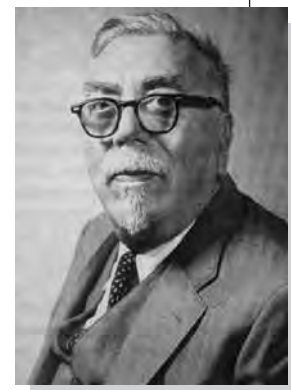
Начав трудиться над книгой, Винер долго размышлял над её заглавием. Он хотел найти слово, как-то объединявшее идеи управления, информации и связи... Свой выбор он остановил на греческом слове «кибернетика». (Первым употребил это слово древнегреческий мыслитель Платон; Ампер в XIX в. предложил называть так науку об управлении человеческим обществом.)

«Кибернетика» вышла из печати с большими погрешностями (в том числе и потому, что Винер не имел возможности сделать корректуру), и издатель невысоко оценивал коммерческое будущее книги. Но она стала истинным бестселлером, одной из самых популярных книг XX столетия. Слово «кибернетика» приобрело необычайную популярность, а идеи Винера совершили прорыв в сознании человечества.

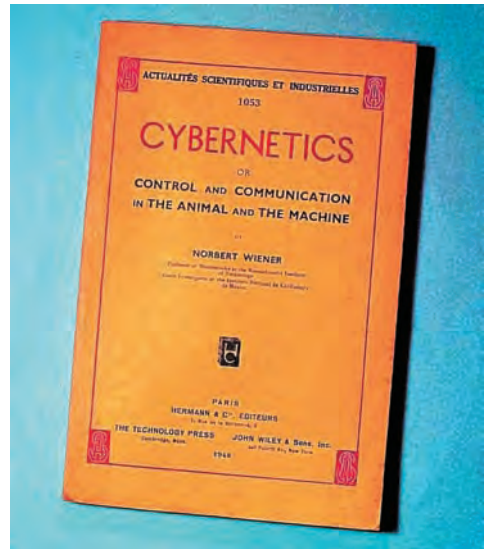
Норберт Винер родился 26 ноября 1894 г. в городке Колумбия (штат Миссури, США). Его отец был полиглотом, славистом, знатоком литературы, переводчиком. С раннего детства мальчик изучал древние языки, естественные науки, много читал. Он был вундеркиндом: среднюю школу окончил в 11 лет, высшее учебное заведение — в 14 (тогда же получил звание бакалавра наук), в 17 лет стал магистром искусств, а в 18 — доктором философии в области математической логики. Получив стипендию Гарвардского университета, Винер проходил стажировку в Кембриджском и Гёттингенском университетах. Он был знаком с Бертраном Расселом, Годфри Харди, Эдмундом Ландау, Давидом Гильбертом и другими известными математиками.

После Первой мировой войны Винер добился выдающихся результатов в математике. В 20-х гг. XX в. он определил так называемый *винеровский процесс,* характеризующий броуновское движение (см. статью «От атомов к молекулам» в томе «Физика» «Энциклопедии для детей»), и опубликовал ряд замечательных работ по гармоническому анализу, послуживших одной из отправных точек для работ Гельфанда по теории банаховых алгебр — крупного события в математике 30-х гг. С 1932 г. Винер — профессор Массачусетсского технологического института.

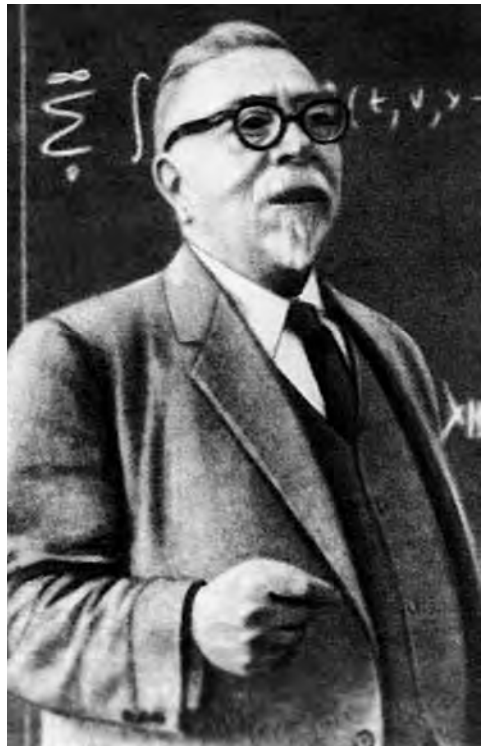
Во время Второй мировой войны Норберт Винер занимался прикладными проблемами, связанными, в частности, с нуждами авиации и зенитной артиллерии. Он дополнил теорию экстраполяции и фильтрации случайных процессов, созданную перед войной А. Н. Колмогоровым, в цикле работ, изданных для служебного пользования (ввиду их засекреченности). Винер применил математический аппарат,



Норберт Винер.



Фотография 1-го издания книги «Кибернетика».



Винер читает лекцию.

не входивший в традиционное инженерное образование. Поэтому книга, изданная в жёлтой обложке, получила среди инженеров, которым была адресована, название «жёлтая опасность». Американский математик Клод Шеннон «перевёл» её на более простой инженерный язык, и с той поры установились постоянные контакты между этими выдающимися научными деятелями XX столетия.

В военные годы Винер организовал в Принстоне семинар, в котором помимо математиков принимали участие нейрофизиологи, специалисты по теории связи и вычислительной технике. Этот семинар фактически стал началом научного направления, впоследствии получившего имя «кибернетика».

В СССР идеи кибернетики были восприняты резко отрицательно (см. статью «Информатика»).

Но уже в 1958 г. на русском языке вышли обе книги Норберта Винера: «Кибернетика, или Управление и связь в животном и машине» и «Кибернетика и общество». В том же году вышел 51-й, дополнительный том Большой советской энциклопедии со статьёй А. Н. Колмогорова «Кибернетика», где отмечалось, что это «новое научное направление, задачи которого были сформулированы в работах американского учёного Н. Винера, опубликованных в 1948 г.». Имя Винера стало неслыханно популярным. В последние годы жизни он много путешествовал, пропагандируя свои идеи. Побывал и в Москве, где встретился с математиками, биохимиками и психологами, сделал доклад в Политехническом музее. Винер умер 19 марта 1964 г.

АНДРЕЙ НИКОЛАЕВИЧ КОЛМОГОРОВ

Андрей Николаевич Колмогоров (1903—1987) был одним из крупнейших математиков XX столетия и одним из величайших российских учёных.

А. Н. Колмогоров родился 25 апреля в Тамбове. Мать умерла через несколько часов после рождения сына, и ребёнка воспитывала её сестра — Вера Яковлевна Колмогорова. Его детство прошло под Ярославлем, в имении деда.

Интерес к математике обнаружился у мальчика очень рано. В возрасте четырёх-пяти лет он вёл математический отдел в семейном детском журнале «Весенние ласточки», где публиковал придуманные задачи и свои математические «открытия» (в частности, что сумма первых n нечётных чисел равна n^2).

С 1910 г. Андрей жил в Москве. Он учился в одной из лучших частных московских гимназий. Круг его юношеских интересов был необычайно широк. Он всерьёз увлекался биологией, физикой, математикой. В возрасте 14 лет юноша самостоятельно изучил дифференциальное и интегральное исчисления по энциклопедическому словарю Брокгауза и Ефрона.

В 1920 г. Колмогоров поступил в Московский университет и стал учеником легендарного математика Н. Н. Лузина, объединившего под своим крылом лучших представителей московской математической школы. Тогда же формируется круг его гуманитарных интересов: он посещает семинар знаменитого русского историка профессора Московского университета С. В. Бахрушина и делает первый научный доклад о своей работе, посвящённой новгородскому землевладению.

Ещё будучи студентом, в 1922 г., в возрасте 19 лет, Колмогоров получает выдающийся результат в теории тригонометрических рядов, и его имя становится известным всему математическому миру. Тогда же он получает и фундаментальные результаты в математической логике.

Начиная с 1924 г. на протяжении последующих 40 лет А. Н. Колмогоров занимается исследованиями в области теории вероятностей. Он завоевал положение безусловного лидера в этой науке. Его книга «Основные понятия теории вероятностей» вошла в золотой фонд наряду с классическими мемуарами Якоба Бернулли и Пьера Симона Лапласа. В этой книге, в частности, была построена общепринятая аксиоматика теории вероятностей.

В 1929 г. Колмогоров оканчивает аспирантуру и с этого времени до конца своей жизни работает на механико-математическом факультете Московского университета.

Тридцатые годы — период необычайного взлёта его творческих достижений. Он создаёт теорию марковских случайных процессов, завершая усилия Альберта Эйнштейна, Макса (Карла Эрнста Людвига) Планка, Мариана Смолуховского и Норберта Винера; получает основополагающий результат по математической статистике; закладывает основы нескольких математических дисциплин (топологической алгебры, геометрии, теории топологических векторных пространств); вводит одно из важнейших понятий топологии (понятие верхних гомологий); строит теорию экстраполяции случайных процессов (пришедший чуть позже к тем же результатам Норберт Винер, создатель кибернетики, рассматривал их как одно из высших своих творческих достижений).

А. Н. Колмогоров.

В 1939 г., в возрасте 36 лет, А. Н. Колмогоров избирается действительным членом Академии наук.

В 40-х гг. XX в. Андрей Николаевич обращается к новой науке — локальной теории турбулентности и совершает в ней подлинный переворот. Сейчас невозможно даже обозреть размеры научного направления, выросшего из трёх его кратких заметок. Во время Великой Отечественной войны по заданию военных ведомств он создаёт статистическую теорию стрельбы, а в первые послевоенные годы разрабатывает основы теории статистического контроля продукции.

Расцвет его творчества — 50-е годы. Колмогоровым были получены выдающиеся результаты в небесной механике, которые сравнивают с достижениями Лапласа и Пуанкаре, и решена вместе с В. И. Арнольдом 13-я проблема Гильберта.

А. Н. Колмогоров принадлежал к тем учёным, которые стояли у истоков развития информатики и кибернетики

А. Н. Колмогоров.

Сложностью сообщения, по Колмогорову, называется минимальная плата по всевозможным методам шифровки сообщения. Это одно из наиболее известных определений внутреннего количества информации в сообщении.

в нашей стране и своим творчеством и авторитетом способствовали их становлению.

В начале 50-х гг. Колмогоров обдумывает и развивает основное понятие информатики — понятие алгоритма и даёт чёткое, ставшее общепринятым определение этого понятия. В начале 60-х гг. он размышляет над общей концепцией автомата, развивая идеи растущих и самоконструирующихся автоматов.

Он был одним из первых российских учёных, осознавших значение для математической науки трудов создателя *теории информации* Клода Шеннона. Основное понятие этой теории — понятие *энтропии* — Колмогоров обобщил и применил в весьма отдалённых областях математики: функциональном анализе, теории динамических систем и теории аппроксимации, что привело к созданию целых новых научных направлений. Но и самому шенноновскому понятию энтропии Колмогоров придал расширенный математический смысл, соединив с понятием *информации*.

Колмогоров много размышлял над философскими аспектами кибернетики, над понятиями «жизнь», «мышление», «управление», над проблемами будущего развития человечества. В одной из публикаций Андрей Николаевич назвал себя «отчаянным кибернетиком». В частности, на вопросы: «могут ли машины воспроизводить себе подобных и может ли в процессе самовоспроизведения происходить прогрессивная эволюция; могут ли машины испытывать эмоции: радоваться, грустить, быть недовольными чем-нибудь, чего-нибудь хотеть; могут ли, наконец, машины ставить перед собой задачи, не поставленные перед ними конструкторами?», он давал положительные ответы.

Колмогоров читает лекцию.

В последний период своей творческой деятельности Колмогоров начал развивать грандиозный замысел создания алгоритмических основ теории информации и теории вероятностей. Ему принадлежит знаменитое понятие колмогоровской сложности.

Колмогоров создал большую научную школу, многие его ученики стали выдающимися учёными, получили международное признание, 12 из них были избраны действительными членами и членами-корреспондентами Академии наук.

Последнюю четверть XX в. Колмогоров посвятил проблемам школьного математического образования.

А. Н. Колмогоров был в числе основателей и возглавлял многие математические журналы, такие, как «Успехи математических наук», «Теория вероятностей и её применение», «Квант» (журнал для юношества).

Характеризуя своего друга, академик Павел Сергеевич Александров писал, что «трудно найти математика в последних десятилетиях не только такой широты, а с таким воздействием на математические вкусы и на развитие математики».

Колмогоров — яркий пример универсального учёного. Он был и выдающимся логиком, и геометром, и аналитиком, и натурфилософом, и «математиком для математики», и человеком глубоких гуманитарных интересов.

А. Н. Колмогоров всегда искал и отстаивал истину. Он был исключительной, гениальной личностью. Трудно найти ещё примеры учёных такой поразительной силы, широты и глубины.

АЛАН МАТИСОН ТЬЮРИНГ

Сегодня Алан Тьюринг представляется нам как
Основатель Компьютерной Науки,
создатель доминирующей технологии конца XX века,
но этих слов никто не говорил в годы его жизни,
и в будущем его могут увидеть в совершенно ином свете.

Эндрю Ходжс. «Алан Тьюринг: Загадка»

Алан Тьюринг.

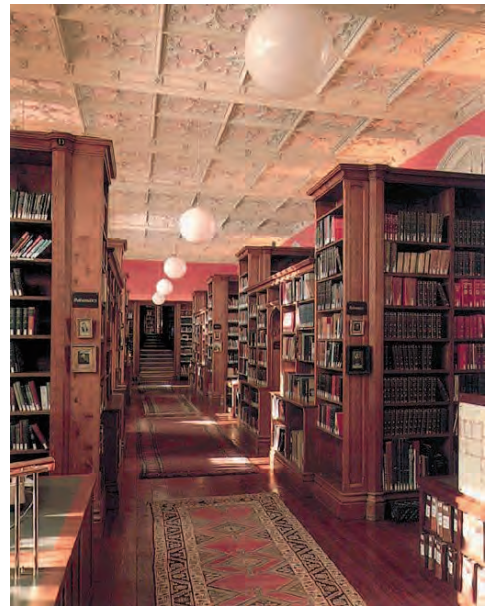
Алан Матисон Тьюринг появился на свет 23 июня 1912 г. в семье Джулиуса Матисона Тьюринга и Этель Сары Стоней. Он был вторым ребёнком. Родители Алана Тьюринга познакомились и обвенчались в Индии. Это была небогатая английская аристократическая семья, жившая в соответствии со строгими традициями империи. Среди их многочисленной родни был ирландский физик и математик Джордж Джонстон Стоней, который ввёл понятие об элементарном электрическом заряде и придумал термин «электрон».

В детстве Алан и его старший брат Джон довольно редко видели родителей. Пока их отец служил в Индии, дети оставались в Англии и жили на попечении друзей семьи. Традиционным английским воспитанием изучение основ естественных наук не предусматривалось. Но Алан обладал очень пытливым умом и с разрешения воспитателей читал научно-популярные книги (читать он научился самостоятельно в возрасте шести лет). Тогда кто-то из родных и подарил мальчику популярную книжку «Чудеса природы, которые должен знать каждый ребёнок» Эдвина Т. Брюстера. С этого всё и началось.

В 11 лет он ставил химические опыты, пытаясь извлечь йод из водорослей, что доставляло огромное беспокойство его матери. Она боялась, что увлечения сына, идущие вразрез с общепринятыми нормами, помешают ему поступить в одну из Public Schools (английское закрытое частное учебное заведение для мальчиков, учёба в котором считалась обязательной для детей аристократов). Но опасения оказались напрасны: в 1926 г. Алан был зачислен в престижную Шерборнскую школу (Sherborn Public School). Впрочем, вскоре пришлось опасаться уже того, сможет ли он окончить эту школу.

Об успеваемости Алана красноречиво свидетельствует классный журнал: последнее место в классе по английскому языку, предпоследнее — по латинскому, по остальным предметам чуть лучше. Вердикты учителей однообразны: «безнадёжное отставание», «безобразная успеваемость»... Директор школы писал: «Этот мальчик из тех, кто обречён стать большой проблемой для любой школы или сообщества...». Между тем в классном журнале имеются и другие записи: «Если он хочет быть только научным специалистом, он зря проводит время в Public School... Наверное, он будет математиком. Такие ученики, как он, рождаются один раз в 200 лет».

В 15 лет Алан самостоятельно освоил теорию относительности, но обстановка и стиль обучения в классической



Библиотека Королевского колледжа Кембриджского университета.

Алан Тьюринг в годы учёбы.

Тьюринг участвует в кроссе. 1946 г.

британской школе не располагали к развитию подобных интересов. Обязательные программные предметы оставляли подростка полностью равнодушным, успевал он еле-еле и в итоге оказался перед реальной перспективой вообще не получить аттестата.

В 1928 г. Тьюринг нашёл долгожданную родственную душу — нового одноклассника, весьма одарённого ученика по имени Кристофер Морком. Наконец-то Алан смог с кем-то поделиться своими размышлениями. Юноши стали неразлучными друзьями. После окончания школы оба собирались поступать в Кембриджский университет. Первая попытка сдать предварительные экзамены в Кембридж закончилась для Алана неудачей. Но он не слишком расстраивался, потому что искренне радовался за Кристофера, который успешно прошёл испытания и получил стипендию. Алан надеялся поступить со второй попытки, чтобы учиться вместе с другом.

Но 13 февраля 1930 г. Кристофер внезапно умер. Смерть лучшего друга потрясла 17-летнего Тьюринга — он задумался над тем, как материализован человеческий разум и высвобождается ли он после смерти. Ответы на эти вопросы Алан искал в книге «Природа физического мира» Артура Эддингтона (1882—1944), а объяснение свободы человеческого выбора — в квантовой механике.

В октябре 1931 г. Алан Тьюринг стал студентом Королевского колледжа Кембриджского университета. И сразу же стали заметны значительные успехи: прекрасная успеваемость, интенсивная научная работа, кембриджская степень бакалавра с отличием в 1934 г., степень магистра и аспирантская стипендия от Королевского колледжа в 1935 г., премия Смита за работу по теории вероятностей в 1936 г.

Размышления над природой человеческого разума в 1932 г. привели студента Тьюринга к изучению очень сложной работы Джона фон Неймана о логических основах квантовой механики и далее к исследованиям в области формальной логики. В 1935 г. он внезапно увлёкся знаменитой проблемой Гильберта о разрешимости и начал работать над ней. В апреле 1936 г., когда Алан уже завершил свою работу, появилась статья американского логика Алонзо Черча с тем же результатом, но полученным совершенно иным методом. Ответы обоих исследователей были отрицательными, однако Тьюринг пришёл к своему очень оригинально. Сначала Алан определил метод в виде механического процесса работы воображаемой машины, считывающей символы-инструкции с бумажной ленты. Затем он доказал, что его машине под силу выполнить любое вычисление, на которое способен человеческий ум в пределах формальной логики. Несмотря на то что Черча считают первым, кто решил знаменитую проблему, работа Алана Тьюринга «О вычислимых числах, с приложением к проблеме разрешимости» принесла ему славу первооткрывателя таких понятий, как «вычислительный алгоритм» и «универсальный компьютер».

Тьюринг блестяще закончил четырёхлетний курс обучения. Казалось, его ждёт успешная карьера слегка эксцентричного кембриджского профессора, занимающегося чистой математикой. Но в сентябре 1936 г. Алан получил двухгодичную аспирантскую стипендию от Принстонского университета (США) и приехал к Джону фон Нейману. Там он работал в области алгебры и теории чисел, доказав при этом соответствие между собственным определением вычислимости и определением Черча, а также продолжил развитие своих идей о логике вычислений, что и стало темой его докторской диссертации — самой сложной и глубокой математической работы Тьюринга. Он сконструировал шифровальную машину на электромагнитных реле и принялся за изучение науки шифрования.

В 1938 г., сразу после защиты диссертации, Алану предложили профессорскую должность в Принстонском университете, но он вернулся в Англию, в Кембридж. В сентябре 1939 г. Тьюринг начал работать на Правительственную школу кодов и шифров.

Давид Гильберт (1862—1943) ставил следующий вопрос: «Существует ли, хотя бы в принципе, определённый метод или процесс, посредством которого можно решить любую математическую проблему?».

Началась Вторая мировая война. Германия для кодировки радиосообщений применяла шифровальную электромеханическую машину «Энигма» («Загадка»), раскрыть её шифр никак не удавалось. Главным героем увлекательнейшей криптографической битвы стал Алан Тьюринг. Англичане «Энигме» противопоставили свою дешифрующую электромеханическую машину «Бомба», но она не справлялась. Тьюрингу удалось значительно усовершенствовать машину, благодаря чему был взломан шифр — тот, что использовали люфтваффе (Военно-воздушные силы Германии), после этого англичане оказались в курсе всех деталей операций, планируемых воздушным флотом Геринга. Но 1 февраля 1942 г., вскоре после вступления в войну США, германские субмарины перешли на усложнённый шифр («рыбий язык»), и «Бомба» потерпела неудачу. Алан принялся за создание новой машины «Колосс», в которой вместо электромеханических реле использовались 2 тыс. электронных ламп. «Колосс» легко победил «Энигму» и справился даже с «рыбьим языком».

С ноября 1942 по март 1943 г. Тьюринга командировали в США, где учёный представлял британскую сторону в обеспечении электронного шифрования переговоров между Рузвельтом и Черчиллем (позже по инициативе Черчилля математика наградили орденом Британской империи, но за какие заслуги, известно было только узкому кругу посвящённых).

Не забывал Тьюринг и о теоретических разработках. Одна из его ранних идей — аналогии в работе воображаемой «машины Тьюринга» и человеческого мозга — трансформировалась в концепцию «умного компьютера» (искусственный интеллект), чей IQ (коэффициент интеллекта) проверялся с помощью специально разработанной логической процедуры, сегодня называемой тестом Тьюринга.

В послевоенные годы Аланом Тьюрингом завладела идея создания универсального программируемого электронного компьютера, способного делать всё: считать и проводить сложные алгебраические вычисления, раскрывать шифры и играть в шахматы. Он мечтал о библиотеке программ, о едином национальном компьютерном центре с сетью терминалов.

В 1949 г. в университете Манчестера, куда перешёл Тьюринг и где он разрабатывал программы для электронного компьютера, появился компьютер Manchester Mark I, или, как его ещё называли, MADAM (от Manchester Automatic Digital Machine). В результате развития проекта MADAM был введён в действие один из первых в мире полностью работоспособных компьютеров — Ferranti Mark I.

В 1950 г. Тьюринг написал философскую статью «Вычислительные машины и интеллект» (в 60-х гг. вышла на русском языке в виде отдельной книги под названием «Может ли машина мыслить?») — классику будущей науки об искусственном интеллекте. В 1951 г. к нему пришло признание научных заслуг: учёного избрали членом Лондонского королевского учёного общества. Тьюринг начинает вычислительные эксперименты по нелинейному моделированию формообразования у растений и раковин. Он исследует то, как живые организмы растут и почему приобретают свою форму. Ответы на вопросы, волновавшие с детства, он обнаружил в нелинейности уравнений химических реакций и диффузии, моделируя их решения на компьютере. Его исследования показали, что электронный компьютер может использоваться в научных целях.

Хотя Алан Тьюринг — выходец из аристократической семьи, он никогда не был эстетом: кембриджские политические и литературные кружки были ему чужды. Он предпочитал заниматься любимой математикой, а в свободное время — проводить

Шифровальная машина «Энигма».

Во время своих химических занятий Алан Тьюринг играл в «Необитаемый остров» — игру, изобретённую им самим. Её цель заключалась в том, чтобы получать различные полезные химические вещества из подручных средств: стирального порошка, средства для мытья посуды, чернил и тому подобной бытовой химии...

химические опыты, решать шахматные головоломки, играть в го.

Друзей у него почти не было. Многих отталкивали его несколько беспорядочный стиль одежды, эксцентричная причёска и резкий скрипучий голос (к тому же иногда он сильно заикался).

Студенты шептались о том, что, подводя стрелки будильника, Алан не пользуется сигналами точного времени, а глядит на звёзды и производит ему одному известные вычисления.

Тьюринг всегда отличался завидным здоровьем, был прекрасным спортсменом, отдавая предпочтение гребле и марафонскому бегу. В свободное время он занимался в Валтонском спортивном клубе. По утрам частенько пробегал 10 км (от дома до лаборатории), чем удивлял коллег. А в период цветения растений, вызывающих у него аллергию, ездил на велосипеде, надев противогаз, за что его неоднократно останавливали полицейские. Свои усиленные тренировки Алан объяснял необходимостью снять стресс, вызванный работой. В 1948 г. в кроссе он показал результат лучше, чем спортсмен, вскоре ставший серебряным призёром Олимпийских игр. Тьюрингу же лишь досадная травма помешала участвовать в играх.

Жизнь этого разносторонне одарённого человека прервалась рано. Алан Тьюринг был арестован и 31 марта 1953 г. предстал перед судом по обвинению в гомосексуализме (то, что сегодня называют нетрадиционной сексуальной ориентацией, в послевоенной Англии считалось преступлением). Ему предоставили выбор: тюрьма или лечение. Он избрал лечение. Всё пережитое вызвало расстройство психического здоровья. В глубокой депрессии продолжал учёный научную работу — до того дня, когда поставил свой последний химический опыт...

Прислуга нашла его мёртвым 8 июня 1954 г. На недоеденном яблоке и на пальцах учёного обнаружили цианистый калий, тот же яд был рассыпан по столу, на котором он накануне ставил химические опыты. Заключение следователя прозвучало категорично: самоубийство. Но мать Тьюринга и многие его коллеги не поверили в это, ведь Алан недавно начал новые исследования в области теории программирования, биологии, квантовой физики, несмотря на то что после суда ему запрещалось работать в секретных компьютерных лабораториях... Тьюрингу был 41 год.

АНДРЕЙ ПЕТРОВИЧ ЕРШОВ

А. П. Ершов.

Академик Андрей Петрович Ершов (1931—1988) — один из основателей теоретического и системного программирования, создатель сибирской школы информатики. Его существенный вклад в становление информатики как новой отрасли науки и нового феномена общественной жизни широко признан в нашей стране и за рубежом.

Заканчивая школу, Андрей, как и многие юноши того времени, мечтал о ядерной физике. В 1949 г. он поступил в Московский университет, однако от карьеры физика-ядерщика пришлось отказаться уже при подаче документов: в годы Великой Отечественной войны 11—12-летним мальчишкой Андрей находился на оккупированной фашистами территории, и этого было достаточно, чтобы не допустить его к тем областям науки, которые содержали государственные тайны.

Ещё студентом МГУ под влиянием А. А. Ляпунова Ершов страстно и на всю жизнь увлёкся программированием, где каждый шаг являлся творческим и интересным поиском. Его наставники занимались изучением Программы автоматического присвоения адресов — ПАПА (она внедря-

ТРОПА В АКАДЕМГОРОДКЕ

А. П. Ершов

Двадцать лет хожу я на работу
по тропе, проложенной в лесу.
Если мне Господь послал заботу,
Я её здесь с лёгкостью несу.

Всем живым заполнено пространство —
Птицы, белки, травы, дерева...
Жизни ход и жизни постоянство —
Той тропы заветные слова.

Здесь недавно поселилась фея.
Смотрят в душу глаз её лучи.
Мне в лицо её дыханье веет,
Тихий голос строчками звучит.

Но всего родней, всего дороже
В непрерывном беге быстрых дней
Неслучайно встреченный прохожий,
Путь держащий по тропе моей.

Двадцать лет, не обронив ни слова,
Мы стремим друг другу быстрый взгляд.
Этот взгляд при каждой встрече новой
Мне приносит бодрости заряд.

Жить с людьми — заслон любой заботе.
Три семьи царят в моей судьбе:
Дома — первая, вторая — на работе,
Третья — на солнечной тропе.

Май 1983 г.

лась в первых вычислительных центрах и привела к развитию языков низкого уровня — автокодам). Андрей же исследует другое направление. Его дипломная работа называлась «Программирующая программа» — ПП. По существу, речь шла о первом трансляторе в высокоуровневой системе программирования. Этот труд во многом определил жизненный выбор молодого учёного.

После окончания университета Андрей Петрович поступил на работу в Институт точной механики и вычислительной техники Академии наук — организацию, в которой складывался один из первых советских коллективов программистов.

Школа № 162. Новосибирск. 1984 г.

А. П. Ершов с учащимися Всесоюзной летней школы.

В 1957 г. его назначают заведующим отделом автоматизации программирования во вновь созданном Вычислительном центре АН СССР. В связи с образованием Сибирского отделения АН СССР по просьбе директора Института математики СО АН СССР академика С. Л. Соболева Андрей Петрович берёт на себя обязанность организатора и фактического руководителя отдела программирования этого института, а затем переходит в Вычислительный центр СО РАН. Впрочем, «по просьбе директора» — это скорее официальная формулировка. В действительности же он в числе первопроходцев-энтузиастов добровольно отправился осваивать новую Сибирь — строить Академгородок, ставший ему по-настоящему родным. Ершов близко воспринимал научную и общественную жизнь Академгородка, с высочайшим чувством ответственности организовывал научные конференции, принимал гостей и друзей со всего мира, активно участвовал в деятельности Научного совета по проблемам образования при Президиуме СО АН СССР.

А.П. ЕРШОВ — УЧЁНЫЙ И ЧЕЛОВЕК

Вокруг Андрея Петровича Ершова всегда кипела жизнь — его окружали студенты и школьники, аспиранты и молодые исследователи, зарубежные учёные и сибирские журналисты. Андрей Петрович не был ни администратором по призванию, ни воспитателем по образованию, но тем не менее к нему постоянно обращались за советами, консультациями, с просьбами. Приходили и в радости: «Родился сын!», и в печали: «Что делать? Провалился на защите». Среди его аспирантов дружно и увлечённо работали узбеки и евреи, русские и татары, украинцы и литовцы. У него для каждого находилось нужное слово, правильный тон. Речь идёт о человеке предельно загруженном, ценящем и умеющем считать каждую секунду. Но никто и никогда не слышал от него, казалось бы, очень естественной реплики: «Я занят! Мне некогда!». Интеллигент высшей пробы, он навсегда остался в памяти коллег, друзей и родных настоящим человеком.

Не часто пользовался он служебным автомобилем: пешая прогулка по лесным тропинкам Академгородка ранним утром на работу и поздним вечером домой стали частью его повседневной жизни.

Академик А. П. Ершов уделял много внимания проблемам информационного обеспечения учёных. Он очень любил книги и собирал их всю жизнь. В его личной библиотеке хранилось более 30 тыс. книг, журналов, трудов конференций, препринтов и отдельных оттисков статей практически на всех европейских языках. После смерти Андрея Петровича его наследники передали библиотеку в Институт систем информатики. Теперь это Мемориальная библиотека имени А. П. Ершова.

Андрей Петрович Ершов был не только талантливым учёным, учителем и борцом за свои идеи, но и выдающейся, разносторонне одарённой личностью. Он любил музыку, поэзию, писал сам и переводил на русский язык стихи английских поэтов. Им сделан замечательный перевод киплинговского стихотворения «Когда» («К сыну»), который специалисты считают лучшим (а ведь существуют переводы и таких столпов поэзии, как С. Я. Маршак и М. Л. Лозинский). Любимым местом отдыха были музыкальные вечера в Доме учёных.

Андрей Петрович прекрасно играл на гитаре и пел. Приятно вспомнить вечер у костра в сосновом лесу на берегу Оби. Школьники, слушавшие днём лекции «О человеческом и эстетическом факторах в программировании», окружают плотным кольцом любимого учителя, взявшего в руки гитару, и тихонько подпевают. Все, кто имел счастье общаться с академиком Ершовым, будут всегда помнить его блестящие идеи, выдающиеся достижения и необыкновенную доброжелательность.

Работал он удивительно. Конечно, святыми были рабочий кабинет и университетская библиотека. Но выдающийся учёный не видел затруднений, если приходилось писать на откидном столике самолётного кресла или в междугородном автобусе Новосибирск — Барнаул. Младшие коллеги и ученики Ершова старались подражать стилю его работы даже в мелочах. Каждая страничка рукописей имела пометки даты, часа и минут. Редкая сосредоточенность и целеустремлённость позволяли не забывать о широком диапазоне поглощавших его дел.

Даже встречу в палате Онкологического центра незадолго до кончины Ершова нельзя было назвать посещением больного. Рядом с больничной койкой стоял столик с книгами и рукописями, Андрей Петрович постоянно работал, говорил о делах и заботах посетителя, а не о мучившей его болезни.

Фундаментальные исследования А. П. Ершова в области схем программ и теории компиляции оказали заметное влияние на его многочисленных учеников и последователей.

Книга А. П. Ершова «Программирующая программа для быстродействующей электронной счётной машины» — одна из первых в мире монографий по автоматизации программирования.

За существенный вклад в теорию смешанных вычислений в 1983 г. А. П. Ершов был удостоен премии имени академика А. Н. Крылова.

Язык программирования «Альфа» и оптимизирующий «Альфа»-транслятор, первая советская система разделения времени «Аист-0», система учебной информатики «Школьница», система подготовки печатных изданий «Рубин», многопроцессорная рабочая станция «Мрамор» — все эти значительные проекты инициировались А. П. Ершовым и выполнялись под его руководством.

Благодаря уникальным способностям научного предвидения Андрей Петрович одним из первых в нашей стране осознал ключевую роль вычислительной техники в прогрессе науки и общества. Его блестящие идеи заложили основу для развития в России таких научных направлений, как параллельное программирование и искусственный интеллект.

В 1980 г. академик начал эксперименты по преподаванию программирования в средней школе. Это привело к введению в школьную программу курса информатики и вычислительной техники. Широко используемый сейчас термин «школьная информатика» рождён в кабинете Ершова в Вычислительном центре СО АН. С его непосредственным авторским участием (предложил описание школьного алгоритмического языка) подготовлен первый советский школьный учебник по информатике. Кроме того, он создал ряд учебных телепередач по данному курсу. А. П. Ершов был постоянным и самым дорогим гостем всех Всесоюзных летних школ юных программистов, регулярно проводившихся в новосибирском Академгородке.

Велика роль А. П. Ершова как организатора науки: он принимал самое активное участие в подготовке множества международных научных конференций и конгрессов, был редактором или членом редколлегии как русских журналов «Микропроцессорные средства и системы», «Кибернетика», «Программирование», так и международных — «Acta Informatica», «Information Processing Letters», «Theoretical Computer Science».

Имя Ершова носят Институт систем информатики СО РАН и благотворительный Фонд имени А. П. Ершова, основной целью которого является развитие информатики как изобретательства, творчества и искусства.

А. П. Ершов в 1986 г. Физические эксперименты.

ЗАЧЕМ И КАК УЧИТЬ ДЕТЕЙ ПРОГРАММИРОВАНИЮ

ТРИ КАТЕГОРИИ ПРОГРАММИСТОВ

С каждым годом всё больше становится вычислительных машин. Ещё быстрее увеличивается количество задач, решаемых с помощью компьютеров. Казалось бы, перспектива ясна: скоро всем придётся писать программы на алгоритмических языках.

С первых шагов развития вычислительной техники среди программистов наметилось разделение труда, ставшее особенно отчётливым в наши дни. *Системные программисты* — это программисты-профессионалы. Они создают программы (операционные системы, трансляторы с языков высокого уровня, программы управления базами данных и т. д.), облегчающие общение с машиной для всех компьютерных пользователей. Системные программисты получают высшее, как правило университетское, образование.

Второй «слой» программистов — *прикладные программисты.* Это специалисты по экономике, геологии, полиграфии, метеорологии, медицине и т. д., которые, не зная ни теории программирования, ни особенностей программирования на конкретных компьютерах, умеют описать задачи из своих предметных областей на одном из наиболее подходящих для этой цели языков программирования. Прикладных программистов готовят обычно профильные высшие учебные заведения.

Усилиями системных и прикладных программистов решение задачи на машине может быть сведено до уровня подстановки параметров

Компьютерный класс.

в уже готовые процедуры. Именно таким путём открывается доступ к компьютерам для более широкого круга людей — *параметрических пользователей.* Они могут не знать ни устройства компьютера, ни алгоритма решаемой задачи, ни языка, на котором написана программа. Таким параметрическим пользователем является и кассир в железнодорожной кассе, и оператор банка, и почтовый служащий.

Число системных и прикладных программистов сравнительно невелико, да и увеличивается со временем не так уж значительно. А вот число «простых» параметрических пользователей растёт и будет расти стремительно. Сейчас совершенно ясно, что скоро ими станет бóльшая часть населения земного шара. Но для успешного общения с компьютером даже параметрическим пользователям необходимы определённые умения и навыки.

ЧТО МОЖЕТ ПРОГРАММИСТ

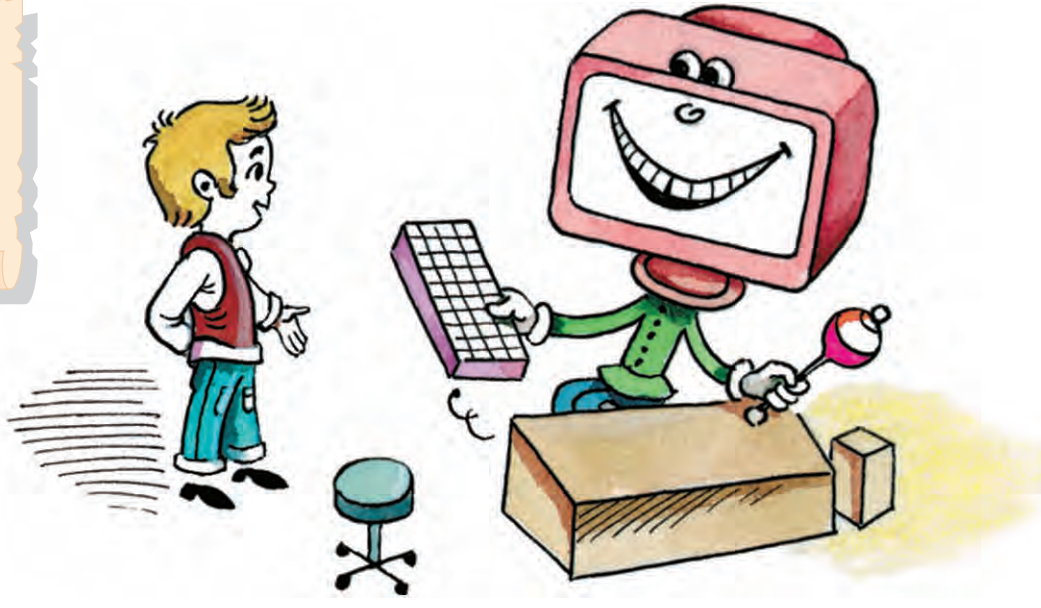
Вот только часть умений и навыков, которые понадобятся каждому параметрическому пользователю.

Во-первых, необходимо научиться планировать действия исходя из ограниченного набора средств.

Во-вторых, уметь формально и полно описать все объекты в качестве данных или результатов, указать их взаимосвязи, или, как говорят программисты, построить информационную модель задачи.

В-третьих, требуется знать, как быстро найти нужную информацию (ведь большую часть времени, затра-

Вы прирождённый программист, если в ваших наручных часах стоит процессор классом не ниже Pentium IV.



ченного на решение задач на машине, занимают операции поиска информации).

В-четвёртых, очень важно правильно строить свои команды машине: если компьютер имеет богатый набор процедур (или, как сказал бы системный программист, хорошее программное обеспечение), то с ним можно общаться с помощью крупных порций информации — процедур; если же такие процедуры заранее не написаны, то обращаться к машине можно, лишь описав требуемые действия с помощью более мелких порций информации — операторов или директив.

Действительно, приведённые выше навыки и умения связывались до сих пор исключительно с необходимостью эффективного использования вычислительной техники. Между тем каждому из них соответствует и более широкое значение.

Например, умение планировать необходимо не только компьютерному пользователю, но и инженеру, агроному, офицеру, учителю, экономисту и т. д.

Умение моделировать, т. е. предвидеть свойства создаваемой конструкции, наиболее важно для творческих работников — технологов, инженеров, физиков, астрономов...

Правильная организация поиска — залог успеха в работе всякого, кто обращается за нужными сведениями к разнообразным хранилищам человеческих знаний: библиотекам, архивам, информационным системам. Недаром говорят, что, не имея возможности передать ученику за его школьные годы все накопленные человечеством знания, учитель должен в первую очередь научить школьника учиться, т. е. уметь отыскивать нужную ему информацию.

Наконец, дисциплина общения в любом человеческом коллективе не менее существенна, чем при «разговоре» с компьютером. И дело здесь не в одной лишь точности высказываний или вежливости формы. Чтобы беседа получилась деловой, надо учитывать, например, уровень знаний собеседника.

ИДЁМ ПО МАГАЗИНАМ...

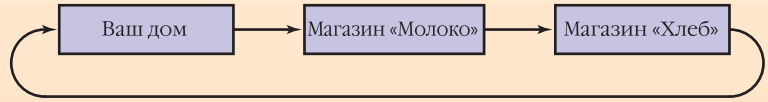
Информатика окружает нас со всех сторон. Иногда кажется, что знания информатики даются нам с рождения. Большинство педагогов волнует вопрос: происходит ли обучение алгоритмическому мышлению в естественных условиях? К сожалению, нет. Вот факт, поражающий преподавателей информатики из года в год уже много лет.

Через неделю после начала курса информатики на уроке, который назовём «прививка интереса», предлагается среди прочего простенькая задача: «Вам нужно, купив молоко и хлеб, вернуться домой. При этом не надо выполнять ничего сверхъестественного и выдумывать дополнительные условия. Просто купить и вернуться».

Ни разу ни в одном классе ни один ученик не купил сначала хлеб, потом молоко, а только наоборот, т. е. попусту тратил свои силы. Все тащили молоко до магазина «Хлеб» и обратно.

С прискорбием можно констатировать, что планирование (или назовём его программированием) не является обычным (т. е. естественным) и распространённым способом мышления. Алгоритмическое мышление не «автоматическое» умение, как, например, поддержание равновесия при ходьбе или чтение знакомых букв. То есть оказывается, что это умение (даже если ему научить) надо специально «включать» в каждом отдельном случае. Сигналом для включения обычно бывает неуспех при лобовом решении задачи.

Пока не возникнет препятствий, последовательность действий всегда примитивна (сначала — молоко, потом — хлеб) и очень часто неэффективна.



ПРОГРАММИРОВАНИЕ ПРИДЁТ НА ПОМОЩЬ

Умения планировать, моделировать, искать информацию, общаться настолько фундаментальны, что обучение следует начинать как можно раньше. Но ни физика, ни литература, ни даже математика не располагают (в полном объёме) необходимыми инструментами для этого. В программировании же такие инструменты есть! Действительно, тому, кто понял, что такое процедура,

условный оператор, цикл, умению планировать легко обучиться. Для информационного моделирования хорошо служат средства описания данных, имеющиеся в современных языках программирования: множества, таблицы, каталоги, файлы. Точно так же для организации поиска используются различные поисковые механизмы (каталоги, многоуровневые индексы, функции расстановки и т. д.), а для организации процессов общения — синтаксические описания и макросредства.

На вопрос, как обучать детей, ответ подсказал опыт учителей и программистов, занимающихся разработкой учебных программ. Они считают, что в первую очередь необходимо научиться управлять простыми исполнителями-роботами, каждый из которых предназначается для конкретной, в том числе учебной, задачи. Поэтому на уроках информатики у учащихся начальной школы на экранах можно увидеть совсем не строчки программ, написанных на языке ЛОГО или Pascal, а исполнителей — Перевозчика, Плюсика, Чертёжника, Кенгурёнка, Кукарачу... Малыши учатся играя. И только узнав в игре с исполнителями главные понятия, нужные для программирования (команда, система команд, имя информационного объекта, ветвления, повторения), можно приступать к освоению языков программирования, сначала учебных, а затем и профессиональных.

ПРОГРАММИРОВАНИЕ — ВТОРАЯ ГРАМОТНОСТЬ

Содержание курса информатики в советской (теперь российской) школе значительно изменилось с момента его введения — от обязательного обучения алгоритмизации, когда школьников учили программированию на одном из популярных языков типа Basic и Pascal, до редактирования текстов с помощью Microsoft Word, когда из учащихся готовят секретарей-машинисток. Умение предварительно планировать свои (или чужие) действия, составлять программы (план дальнейших работ) используется повсеместно. Например, чтобы записать в определённое время кинофильм с телевизора на видеомагнитофон, сначала придётся освоить инструкцию по эксплуатации, которая в данном случае описывает язык программирования видеомагнитофона:

установить дату записи = сегодня;
установить начало записи = 10.00;
установить окончание записи = 11.25;
установить номер телеканала = 4;
установить скорость записи = LP.

Действительно, программирование видеомагнитофона — несложный процесс. Но число современных программируемых домашних приборов растёт с каждым днём. Это и мобильные телефоны, где программируется не только запоминание номеров, установка различных параметров, это и микроволновые печи, кухонные кофеварки, хлебопечки и т. д. Компьютер — тот же домашний прибор, только более сложный и универсальный.

В голове или на листке бумаги можно составить программу, используя *универсальный язык программирования.* Выполнять программу может компьютер, человек или абстрактный *универсальный выполнитель,* который понимает язык и умеет заставлять *исполнителей* выполнять команды. Так мама учит ребёнка звонить бабушке по телефону:

«Нажми клавишу 2,
нажми клавишу 3,
нажми клавишу 2...».

Здесь ребёнок является выполнителем программы, а телефон — исполнителем.

На уроках математики тоже происходит выполнение программ, например, когда для решения квадратного уравнения $ax^2+bx+c=0$ нужно вычислить дискриминант $D=b^2-4ac$:

если $D<0$, *то* корней нет,
если $D=0$, *то* корни совпадают,
если...

В этом случае ученик по сути является исполнителем со своим набором команд.

Так же как компьютер часто соединяет в себе целый набор исполнителей и сам является выполнителем программы, так и школьнику не раз приходится выступать в роли выполнителя программ, написанных на различных языках программирования. Действительно, преподаватели на уроках математики и физики, на лабораторных занятиях по биологии и химии никогда не формализуют язык,

Вы прирождённый программист, если на ваших наручных часах кнопок не меньше, чем на клавиатуре вашего ноутбука.

Вы прирождённый программист, если ваш консервный нож имеет разъём USB и используется вами как MP3-плеер.

ЧТО ТАМ ВНУТРИ?

Ребёнок, разобрав электронные часы, принцип их работы понять не сможет (и взрослый подчас тоже). Так же происходит мифологизация объекта под названием «компьютер». Для многих людей он — чудо. И это абсолютно естественно для человека.

С другой стороны, подавляющее большинство детей к 9—10 годам представляют, как работают механические часы. Про пружину, шестерёнки и стрелки разными способами им рассказывали не один раз. А некоторые разбирали будильники и всё видели собственными глазами.

Что получается? Если мы видим поверхность, из которой торчат две стрелки, и знаем про шестерёнки (но не знаем ни об их количестве, ни о размере, ни о последовательности их сцепления), то этого оказывается достаточно, чтобы эффект чуда не возникал.

Сделаем очевидный вывод. О любом объекте (в частности, электронных часах или компьютере) нужно рассказать очень немного, чтобы он был демифологизирован. Обязательно надо разъяснить принципы устройства компьютера и составления программ для него (и лишь попутно учить включать компьютер). Иными словами, в курсе информатики следует говорить о шестерёнках, а не показывать, как заводить будильник, и слушать, как он тикает.

на котором излагаются программы (последовательность действий). Школьники с той или иной долей успеха учатся его понимать. Качество исполнения порой зависит не от умения и сообразительности школьника, а от таланта и опыта педагога, сумевшего формализовать язык изложения и научить ему школьников. Вот где пригодились бы навыки алгоритмизации! Учителя, конечно, считают идеальными учеников, понимающих их с полуслова и выполняющих все задания, т. е. понимающих язык программирования и выполняющих программы.

Примеры удачной формализации языков существуют. У военных есть уставы, где прописано всё — от языка управления (программирование военнослужащих) до программы (поведение солдат и офицеров). Почти формализован «канцелярский» язык. На этом языке ведётся переписка в учреждениях, издаются приказы и распоряжения.

Ребёнок в начальных классах должен получить объём знаний, позволяющий сознательно выполнять программы по всем школьным дисциплинам. Сила математики и других предметов, сила, в конце концов, инструкций по эксплуатации сложной бытовой аппаратуры именно в методических приёмах, позволяющих, не описывая языка программирования, использовать человека в качестве выполнителя.

В средних классах для ученика важно научиться составлять программу на языке универсального выполнителя. От предмета к предмету меняется набор исполнителей — от арифметических действий к химическим или физическим опытам. Ученик как выполнитель программ учителей в состоянии понимать и создавать программы на различных языках.

В старших классах возрастёт трудность задач, появятся сложные выполнители и языки с развитой структурой. Это потребует изучения определённой технологии программирования (составления программ). В итоге ученик обязан уметь проектировать язык программирования выполнителей и исполнителей, необходимых для решения конкретной задачи. Это уже не умение заполнять и использовать ежедневник, а проектирование такого ежедневника.

Можно было бы закончить на мажорной ноте, высоко подняв лозунг — название статьи, призвав молодое поколение к глубокому изучению программирования как основе поведения современного человека. Но программирование и проектирование (а с ним и аналитическое мышление) не свойственно не только большей части общества, но и принципы функционирования самого общества, закреплённые за-

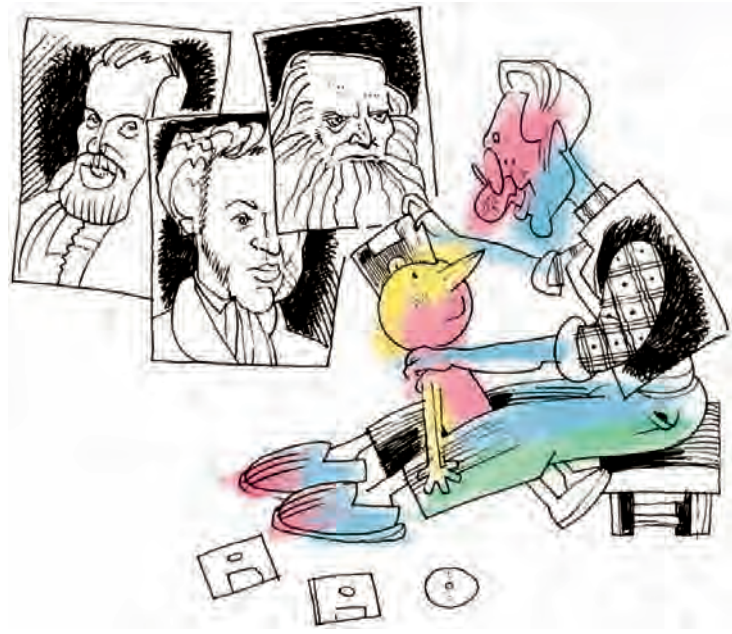
конодательно, часто далеки от простой логики.

Не секрет, что юриспруденция одна из сложнейших для изучения и понимания наук. Не случайно, что мало кто из граждан в состоянии самостоятельно разобраться в тонкостях законодательства, хотя каждый обязан это законодательство выполнять. Таким образом, имеется уникальная ситуация, когда нужно выполнить программу, не понимая языка программирования. Для разъяснения этого языка гражданам (исполнителям) и существуют юристы, которые также часто не могут однозначно истолковать, как надо выполнять тот или иной закон (программу). Спорные моменты разрешает суд, который в силу своей независимости (в том числе и от логики) также вправе неоднозначно истолковать и(или) применить тут или иную норму права (правила выполнения программы). То есть по закону суд может принять неверное решение (неправильно выполнить программу). Здесь правильно или неправильно принятое решение (исполнение программы) зависит только от судьи (выполнителя), что с точки зрения программирования достаточно странно.

Так в Гражданско-процессуальном кодексе Российской Федерации (основной программы) есть статья 67 про оценку доказательств судами. Начало этой статьи звучит так: «Суд оценивает доказательства по своему внутреннему убеждению, основанному на всестороннем, полном, объективном и непосредственном исследовании имеющихся в деле доказательств». Если внимательно прочитать начало предложения, то можно сделать ошеломляющий вывод о том, что алгоритм поведения судьи при принятии доказательств недетерминирован, потому что результат принятия судом доказательства (выполнение программы) — «внутренняя убеждённость» судьи (неизвестное свойство исполнителя). Таким образом, никто не застрахован от событий аналогичных тем, что произошли с двумя молодыми учёными в «Сказке и тройке» — классическом произведении А. и Б. Стругацких: «Эдик снова заговорил. Он взывал к теням Ломоносова и Эйнштейна, он цитировал передовые центральных газет, он воспевал науку и наших мудрых организаторов, но всё было вотще. Лавра Федотовича это затруднение наконец утомило, и, прервавши оратора, он произнес только одно слово:

— “Неубедительно”.

Раздался тяжёлый удар. Большая Круглая Печать впилась в мою заявку».



НА ОШИБКАХ УЧАТСЯ...

За ошибку наказывают. И потому ошибаться не просто не любят, а боятся. А за страхом следует либо бездействие, либо новая ошибка.

Однако мы знаем, что есть люди, которые начинают работу с мыслью: «Сейчас я всё напишу неправильно». И они этого не боятся. Эти люди — программисты. В процессе преподавания информатики можно научить не только не бояться делать ошибки, но и радоваться им, поскольку каждая найденная ошибка уменьшает количество ещё не найденных.

Продолжая тему ошибок, заметим, что попутно (если занятия сопровождаются работой на компьютерах) можно научить разумно и быстро отлаживать свою программу, говоря иначе, «ставить диагноз по симптомам» (а это необходимо следователю, врачу, политику, журналисту, геологоразведчику и др.), поскольку поиск ошибки в программе позволяет устанавливать и перепроверять причину сбоя. Но обучение этому — отдельная задача, которая никак не должна смыкаться с обучением думать алгоритмически.

АЛГОРИТМ

АЛГОРИТМЫ И ПРОГРАММЫ

ПОНЯТИЕ АЛГОРИТМА

Каждый человек в своей повседневной жизни, в учёбе или на работе сталкивается с необходимостью решения огромного количества задач самой разной сложности. Некоторые из них настолько трудны, что требуют длительных размышлений для поиска решения (а иногда найти его так и не удаётся). Другие же, напротив, столь просты и привычны, что решаются автоматически. Решение задачи «Поступить в институт» намного труднее и требует выполнения гораздо большего количества сложных действий, чем решение задачи «Купить мороженое». При этом выполнение даже самой простой задачи обычно осуществляется в несколько последовательных этапов, или *шагов*. Если проанализировать ход выполнения этих задач, то окажется, что практически любое действие, которое надо при этом произвести, может быть чётко сформулировано и записано.

Например, процесс приготовления обеда можно описать так:

1) сварить суп;
2) приготовить второе блюдо;
3) приготовить десерт.

Такая запись (назовём её *инструкцией*) малополезна на практике. Ясно, что шаги должны быть более простыми, поэтому необходимо несколько детализировать её (ограничимся первым пунктом):

1) вымыть, почистить и нарезать овощи;
2) зажечь плиту;
3) вскипятить в кастрюле воду;
4) положить в кипящую воду мясо;
5) когда мясо сварится, положить в кастрюлю овощи;
6) когда овощи сварятся, выключить плиту.

Здесь определена последовательность действий, следуя которой любой человек сумеет приготовить суп. Аналогично в виде последовательности шагов можно описать процесс решения многих задач, известных из школьного курса математики: приведение дробей к общему знаменателю, решение системы линейных уравнений путём последовательного исключения неизвестных, построение треугольника по трём сторонам с помощью циркуля и линейки и т. д. Такая последовательность шагов в решении задачи и называется *алгоритмом*.

Каждое отдельное действие — это *шаг алгоритма* («почистить овощи», «умножить *a* на *b*» и т. д.). Последовательность шагов алгоритма строго фиксирована, т. е. шаги должны быть *упорядоченными*. Правда, существуют так называемые *параллельные алгоритмы,* для которых это требование не соблюдается. Порядок записи шагов в них не обязательно должен совпадать с порядком выполнения соответствующих действий.

Понятие алгоритма близко к другим понятиям, таким, как *метод* (метод Гаусса решения систем линейных уравнений), *способ* (способ построения треугольника по трём сторонам с помощью циркуля и линейки) или даже *рецепт*. Другие понятия, такие, как *процедура* (судебная процедура), *программа* (программа действий правительства по выводу экономики из кризиса), *процесс* (мирный процесс на Ближнем Востоке), также имеют немало общего с алгоритмом. Поэтому необходимо чётко сформулировать отличительные особенности именно алгоритмов.

Прежде всего, алгоритм подразумевает наличие исходных данных и некоторого результата. Следовательно, алгоритм — это точно определённая инструкция, последовательно применяя которую к исходным данным, можно получить решение задачи.

Многие алгоритмы обладают тем свойством, что остаются правильными для разных наборов исходных данных (приготовление супа одинаково для разных используемых овощей; вычисление площади прямоугольника одинаково для любых значений *a* и *b*). Такое свойство алгоритмов называют *массовостью*. Оно позволяет многократно применять один и тот же алгоритм для решения задачи. При этом необходимо иметь в виду, что для каждого алгоритма есть некоторое множество объектов, допустимых в качестве исходных данных. Например, *a* и *b* могут быть только вещественными положительными числами; в алгоритме деления вещественных чисел делимое может быть любым, а делитель не должен быть равен нулю.

Всегда ли удаётся получить результат, применяя алгоритм к некоторому набору из множеств допустимых исходных данных? Не всегда: так, при делении 7 на 3 оба целых числа допустимы, но результат 2,333... является бесконечной десятичной дробью. Об алгоритмах такого рода говорят, что они *потенциально выполнимы,* но состоят из бесконечного числа шагов. В связи с этим надо ввести понятие *конечности* алгоритма. Смысл его состоит в том, что выполнение

алгоритма должно обязательно приводить к его завершению. В то же время можно привести примеры формально *бесконечных* алгоритмов, широко применяемых на практике. Например, алгоритм работы системы сбора метеорологических данных состоит в непрерывном повторении последовательности действий «Измерить температуру воздуха» и «Определить атмосферное давление», выполняемых с некоторой частотой (ежеминутно, ежечасно или ежедневно) всё время существования данной системы.

На практике обычно интересует результат, выдаваемый алгоритмом через конечное число шагов и за конечное время (в самом деле, какова польза от алгоритма расчёта прогноза погоды на завтра, если результат вычисляется неделю). Свойство конечности имеет особо важный смысл, так как часто требуется не потенциальная, а именно *реальная* выполнимость алгоритма.

Вернёмся к примеру с делением. Если оборвать процесс на каком-либо шаге, то полученное число, вообще говоря, не будет частным от деления, но его тем не менее можно принять за приближённый результат, вычисленный с устраивающей в конкретной задаче точностью.

С понятием алгоритма тесно связано представление об *исполнителе алгоритма*. Неважно, кто выступает в этой роли: человек (повар, абитуриент или математик в приведённых выше примерах) или же некоторое техническое устройство (в том числе компьютер). Главное требование — исполнитель алгоритма должен уметь выполнять каждый его шаг.

Отсюда вытекает следующее важное требование к алгоритму — *определённость*. На каждом шаге у исполнителя алгоритма должно быть достаточно информации, чтобы его выполнить. Кроме того, исполнителю алгоритма нужно чётко знать, каким образом он выполняется. Легко заметить, что некоторые шаги приведённого выше алгоритма приготовления супа, в свою очередь, нуждаются в уточнении (так, пункт «Зажечь плиту» потребует выполнения последовательности действий: «Взять коробок спичек», «Зажечь спичку», «Повернуть вентиль газовой плиты», «Поджечь газ»). Следовательно, чтобы выполнялось требование определённости, шаги инструкции должны быть достаточно простыми, *элементарными,* а исполнитель алгоритма должен *однозначно* понимать смысл каждого шага последовательности действий, составляющих алгоритм (при вычислении площади прямоугольника любому исполнителю алгоритма, осуществляющему алгоритм, необходимо уметь умножать, а также трактовать знак операции «x» именно как умножение, а не как, например, сложение). Здесь возникает важная проблема: *в каком виде, в какой форме* будет представлен алгоритм, чтобы выполнялось это условие. Фактически речь идёт о том, на каком *языке* записан алгоритм.

ПЕРВЫЕ АЛГОРИТМЫ

Считается, что самым древним нетривиальным алгоритмом является способ нахождения наибольшего общего делителя двух целых чисел. Он был изложен в трудах древнегреческого математика Эвклида — седьмой книге «Начал» около 2300 лет тому назад. (Правда, есть предположение, что алгоритм Евклида лишь интерпретация алгоритма, предложенного Эвдоксом за 75 лет до этого.)

Алгоритм Евклида нахождения НОД двух целых положительных чисел.

«Пусть *A, C* — два данных положительных целых числа. Надо найти их наибольший общий делитель. Если *C* делит *A*, то *C* является общим делителем чисел *C* и *A*, так как делит и само себя. Очевидно, что *C* будет и наибольшим делителем, поскольку нет числа, большего *C*, которое делило бы *C*.

Если же *C* не делит *A*, то начинаем непрерывно вычитать меньшее из чисел *A, C* из большего до тех пор, пока не получим некоторое число, которое нацело делит предыдущее вычитаемое. Рано или поздно такое число получится, потому что если разность будет равна единице, то единица будет делить предыдущее вычитаемое.

Итак, пусть *E* — положительный остаток от деления *A* на *C*, *F* — положительный остаток от деления *C* на *E*, *F* делит *E*. Так как *F* делит *E*, а *E* делит *C* – *F*, то *F* также делит *C* – *F*, но *F* делит само себя, следовательно, *F* делит *C*. Но *C* делит *A* – *E;* поэтому *F* также делит *A* – *E*. Но оно также делит *E,* поэтому оно делит и *A*. Следовательно, *F* — общий делитель чисел *A* и *C*.

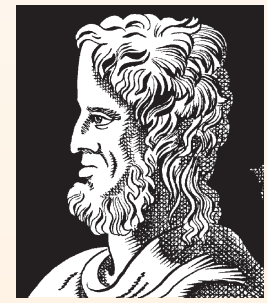
Теперь я утверждаю, что он является и наибольшим таким делителем. Действительно, если *F* — не наибольший общий делитель чисел *A* и *C*, то некоторое большее число будет делить их оба. Пусть таким числом будет *G*.

Поскольку *G* делит *C*, которое, в свою очередь, делит *A* – *E*, то *G* делит *A* – *E; G* делит также все *A*, следовательно, оно делит и остаток *E*. Но *E* делит *C* – *F*, поэтому *G* делит *C* – *F*. Но *G* делит также все *C*, следовательно, оно делит и остаток *F;* таким образом, большее число делит меньшее, а это невозможно.

Поэтому нет такого числа, большего, чем *F*, которое делило бы числа *A* и *C*, и, значит, *F* — их наибольший общий делитель».

В древнеегипетских папирусах можно найти ещё более древние примеры вычислительных процедур. Так, египтянам был известен метод умножения, основанный на удвоении и сложении (кстати, на этой же идее базируются некоторые широко применяемые современные алгоритмы).

В отличие от Древнего Египта в Вавилоне было принято кроме решения примера давать пояснения процесса вычислений, пригодные в общем случае. Например, если один из сомножителей оказывался равным единице, умножение всё-таки выполнялось. И хотя вавилоняне использовали алгоритмы за полторы тысячи лет до греков, историки науки единодушно признают приоритет Евклида. Дело в том, что алгоритм Евклида имеет достаточно сложную структуру: он является *итерационным* (так называют алгоритмы, содержащие неоднократное повторение некоторых действий). Вавилонские алгоритмы представляют исключительно исторический интерес, в то время как алгоритм Евклида не потерял своего практического значения до сих пор.



Эвдокс.

Ещё один старинный алгоритм, который широко использовался на практике.

Алгоритм аль-Каши вычисления значения x^n, где n — положительное число.

Шаг 1. Вводим три величины — $N := n$; $y := 1$; $z := x$.

В этот момент справедливо соотношение $x^n = y \times z^n$.

Шаг 2. Делим N нацело на 2, $N := N$ *div* 2; одновременно определяем, было ли до того N чётным.

Если N было чётным, то переходим к шагу 5.

Шаг 3. Умножаем y на z, $y := y \times z$.

Шаг 4. Если N равно нулю, то ответ равен y.

Шаг 5. Умножаем z на себя, $z := z \times z$.

Возвращаемся к шагу 2.

Этот алгоритм был предложен в самом начале XV в. (впрочем, около 200 г. до н. э. в индийском математическом трактате «Чанда-сутра» приводится описание аналогичного метода). Именно такую процедуру ещё в XVIII в. до н. э. египетские математики использовали для умножения.

Шаг 1. Вводим три величины — $N := n$; $y := 0$; $z := x$.

Шаг 2. Делим N нацело на 2, $N := N$ div 2; одновременно определяем, было ли N до того чётным.

Если N было чётным, то переходим к шагу 5.

Шаг 3. Увеличиваем y на z, $y := y + z$.

Шаг 4. Если N равно нулю, то ответ равен y.

Шаг 5. Складываем z с собой, $z := z + z$.

Возвращаемся к шагу 2.

В шагах 3 и 5 умножение заменено сложением, и в шаге 1 y приравнивается не единице, а нулю. В результате выполнения такого модифицированного алгоритма получаем произведение двух чисел n и x: $y = n \times x$. Это весьма удобный для практических вычислений «вручную» способ умножения, который сводится к более простым операциям удвоения, деления пополам и сложения. Именно такой приём применяется при вычислениях на счётах — в Европе он традиционно носит название «русский крестьянский метод».

Урбен Жан Жозеф Леверье.

Неоднозначность восприятия алгоритма часто может быть связана не с его сутью, а с особенностями представления. Даже в детализированной инструкции не определено точно, что означает «мясо сварилось» или «овощи сварились». В зависимости от особенностей того или иного исполнителя алгоритма (предположим, от опытности повара) эти шаги могут быть выполнены по-разному: мясо, например, окажется недоваренным, а овощи, напротив, разварятся. Итак, каждый шаг алгоритма должен быть таким, чтобы не возникало разночтений.

С определённостью перекликается ещё одно важное свойство алгоритма — *детерминированность*. Его можно понимать следующим образом: если алгоритм многократно применять к одному и тому же набору исходных данных, то всегда должен получаться один и тот же результат.

Итак, алгоритм — это система правил, которая сформулирована на языке, понятном исполнителю алгоритма, определяет процесс перехода от допустимых исходных данных к некоторому результату и обладает свойствами массовости, конечности, определённости, детерминированности.

Это определение алгоритма не является строгим (хотя бы потому, что в нём используются не определённые точно термины, например «правило»).

На протяжении многих веков понятие алгоритма связывалось с числами и относительно простыми действиями над ними. Да и сама математика была по большей части наукой о вычислениях, наукой прикладной. Чаще всего алгоритмы представлялись в виде математических формул. Порядок элементарных шагов алгоритма задавался расстановкой скобок, а сами шаги заключались в выполнении арифметических операций и операций отношения (проверки равенства, неравенства и т. д.). И хотя эти формулы могли быть достаточно громоздкими, а вычисления вручную — крайне трудоёмкими (французский астроном Жан Жозеф Леверье провёл десятки лет в расчётах орбит планет Солнечной системы, в результате чего ему удалось обнаружить неизвестную ранее планету Нептун), суть самого вычислительного процесса оставалась вполне очевидной. У математиков не возникала потребность в осознании и строгом определении понятия алгоритма, в его обобщении.

Но с развитием математики появлялись новые объекты, которыми приходилось оперировать учёным: векторы, матрицы, графы, множества и др. Как определить для них однозначность или как установить конечность алгоритма, какие шаги могут считаться элементарными (например, является ли таковым обращение матрицы или нахождение пересечения двух множеств)?

Возникла идея о существовании алгоритмически неразрешимых проб-

ЕЩЁ ОДНА ФОРМАЛИЗАЦИЯ АЛГОРИТМА

В. А. Успенский.

Алгоритм есть описание действий; изучая лишь запись алгоритма, нельзя понять его динамику. Да и результат выполнения при таком изучении далеко не всегда очевиден.

Схема Колмогорова — Успенского была одной из попыток построить некоторый формальный механизм выполнения алгоритмов, в котором алгоритм и его исполнитель — человек или некоторый прибор — рассматриваются в неразрывном единстве. Здесь действия исполнителя алгоритма чётко определены, а их количество ограниченно.

Предполагается, что обрабатываемая человеком информация записывается на листах бумаги, при этом имеется неограниченный запас чистых листов и неограниченный резерв места для хранения исписанных листов. Преобразование информации, реализуемое человеком, разбивается на отдельные дискретные шаги. На каждом шаге человек обозревает некоторое число исписанных листов и в зависимости от содержания записей на них по строго заданным правилам производит изменения.

Изменения могут быть трёх видов:

— стирание всей обозреваемой информации или некоторой её части;

— запись на обозреваемых листах новой информации;

— изменение совокупности обозреваемых листов.

Однако данная модель оказывается недостаточно формализованной, и её применение для реализации конкретных алгоритмов затруднено. По своему удобству она значительно уступает машинам Поста и Тьюринга, хотя имеет с ними значительное сходство.

лем, таких, для которых нельзя найти процедуру решения. Следовательно, надо было научиться *математически строго доказывать* факт отсутствия соответствующего алгоритма. Это можно сделать только в том случае, если существует строгое определение алгоритма. Попытки выработать такое определение привели на рубеже 20—30-х гг. XX в. к возникновению *теории алгоритмов*.

Почти в это же время возникла и практическая потребность в осуществлении сложных вычислений. На повестку дня встала необходимость решения задач из разных научных и инженерных дисциплин: радиолокации, самолётостроения, моделирования физических процессов (особенно из области ядерной физики, баллистики и многих других). Вычислительные возможности человека были существенно ограничены, средства механизации, а тем более автоматизации вычислений практически отсутствовали. Появление первых проектов вычислительных машин стимулировало исследование возможностей *практического* применения алгоритмов, использование которых ввиду их трудоёмкости было ранее недоступно. Дальнейший бурный прогресс вычислительной техники определил развитие как теоретических, так и прикладных аспектов изучения алгоритмов.

КАК ЗАПИСЫВАТЬ АЛГОРИТМЫ

Описание алгоритма Евклида нахождения НОД двух целых положительных чисел состоит из трёх шагов и занимает всего несколько строчек.

Шаг 1. Разделить m на n, где $m>n$. Пусть p — остаток от деления.

Шаг 2. Если p равно нулю, то n и есть искомый НОД.

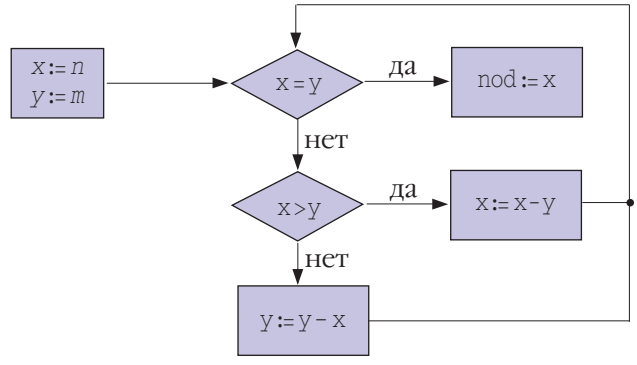
Шаг 3. Если p не равно нулю, то сделаем m равным n, а n — равным p. Вернуться к шагу 1.

В то же время даже значительно «упрощённое» описание, данное Евклидом, требует около страницы текста(!), и понять из его описания, что именно и для чего делается, далеко не просто.

Для записи алгоритмов необходим некоторый язык. При этом крайне важно, какой именно язык выбран. Неудобство русского (равно как и любого другого естественного) языка для этих целей очевидно. Фраза «мальчик взял с собой тетрадь в клетку» может сообщать нам о школьнике, подготовившемся к уроку арифметики. Но может иметься в виду юный любитель животных, который собирается у клетки с тигром записывать свои наблюдения и ощущения... Слова могут иметь самые разные смысловые оттенки, и понять их без детального анализа всего текста подчас просто невозможно.

Весьма распространённым является графический способ записи алгоритмов на языке *блок-схемы*. Она представляет собой набор элементов (блоков), соединённых стрелками. Каждый элемент — это некоторая часть алгоритма. Элементы блок-схемы делятся на два вида: содержащие какие-либо действия обозначают прямоугольниками, а содержащие проверку условия — ромбами. Из прямоугольников всегда выходит только одна стрелка (входить может несколько), а из ромбов — две. Первая помечается словом «да», а вторая — «нет» (они показывают, выполнено условие или нет).

На рисунке показана блок-схема алгоритма нахождения НОД:



Такой способ удобен для анализа разрабатываемого алгоритма, для проверки правильности логики его работы.

Евклид.

Каждый блок на рисунке соответствует элементарному шагу алгоритма. Если же записываемый алгоритм более сложен, то неизбежно встаёт вопрос: как избежать его излишней детализации?

Сделать это можно, укрупняя шаги алгоритма. Тогда каждый блок будет содержать достаточно сложные действия и даже представлять целый алгоритм (точнее, подалгоритм). Например, блок может заменять написанный ранее и уже проверенный алгоритм. Если при этом есть уверенность в правильности работы блока, то в детали его работы лучше не вдаваться. Такое «крупноблочное» представление бывает весьма полезно и на начальной стадии разработки алгоритма, когда известно, *что* надо сделать в этом блоке, но ещё не известно *как*. Уточнение алгоритма работы блока может быть сделано позднее.

Построение блок-схем из элементов всего лишь нескольких типов даёт возможность достаточно просто преобразовывать их в компьютерные программы, позволяет формализовать этот процесс.

Известны и другие способы наглядной записи алгоритмов — в первую очередь операторные схемы, предложенные А. А. Ляпуновым. Приведём схему Ляпунова, описывающую алгоритм НОД:

$$A_1 \overset{4}{\downarrow} P_1(x=y) \overset{1}{\uparrow} A_2 \overset{21}{\uparrow\downarrow} P_2(x>y) \overset{3}{\uparrow} A_3 \overset{43}{\uparrow\downarrow} A_4 \overset{42}{\uparrow\downarrow} S$$

Здесь через A_i обозначены операторы, производящие вычисления (например, A_4 соответствует присваиванию $y := y - x$), а через P_i — логические операторы (P_1 соответствует проверке $x = y$). Стрелки ↑ и ↓ означают переход, им соответствуют начало и конец стрелки на блок-схеме. Если проверка даёт отрицательный результат, то происходит переход по стрелке, иначе выполняется следующий оператор. S — это оператор остановки. Операторам-вычислителям соответствуют прямоугольники, а логическим операторам — ромбы блок-схемы. Операторные схемы легко переписать в виде блок-схем, но, в отличие от последних, они не столь наглядны, и это затрудняет их практическое использование.

Запишем теперь алгоритм Евклида на языке программирования Pascal:

```
var
   n, m, x, y, nod: integer;
      begin
         x:=n; y:=m;
         while x<>y do
            if x>y then x:=x-y
                   else y:=y-x;
         nod:=x;
      end;
```

Очевидно явное сходство с записью как на языке блок-схем, так и на языке операторных схем Ляпунова. Конечно, компьютер по-прежнему не способен воспринимать этот текст непосредственно, необходима его предварительная обработка специальной программой — транслятором. Однако по сравнению с другими формами записи алгоритма такую обработку реализовать значительно проще. Кроме того, она достаточно удобна и для человека.

ПРОГРАММА

Слово «программа» имеет несколько значений: план деятельности или работ, краткое изложение содержания учебного предмета, содержание концерта или циркового представления и некоторые другие. Среди них отдельно выделено одно, специальное значение: *программа* — это описание алгоритма решения задачи на языке компьютера.

Именно данное значение слова подчёркивает тесную связь алгоритмов и программ. Одним из первых обратил внимание на эту связь в 1945 г. английский математик Алан Тьюринг, который считал, что программы (он называл их «таблицами инструкций») должны создаваться математиками, имеющими опыт вычислительной работы и склонность к творчеству («решению головоломок»). Тьюринг особо подчёркивал важность перевода известного процесса (т. е. алгоритма) в форму таблиц инструкций. При этом он не сомневался, что такой перевод — «очаровательное занятие».

Вообще же можно сказать, что программа — это запись алгоритма на языке, понятном исполнителю алгоритма. Особый тип исполнителя алгоритмов — компьютер, поэтому необходимо создавать специальные средства, позволяющие, с одной стороны, разработчику в удобном виде записывать алгоритмы, а с другой — дающие компьютеру возможность понимать написанное, — *языки программирования*, или *алгоритмические языки*.

Один из первых языков программирования — Algol-60 изначально

разрабатывался не только как средство написания программ, но и как средство записи алгоритмов. Это подчёркивается самим названием языка: английские слова algorithmic language можно понимать как «алгоритмический язык» (синоним слов «язык программирования») и как «язык для записи алгоритмов». На протяжении многих лет Algol-60 фактически являлся стандартом для записи алгоритмов, общепринятым при их публикации.

О ПРОИСХОЖДЕНИИ СЛОВА «АЛГОРИТМ»

Известно, что в раннем Средневековье слово algorism использовали для обозначения способа арифметических вычислений на бумаге без применения счётных досок (абаков). Именно в таком значении оно вошло в некоторые европейские языки. Например, в авторитетном словаре английского языка «Webster's New World Dictionary», изданном в 1957 г., оно снабжено пометкой «устаревшее» и объясняется как выполнение арифметических действий с помощью арабских цифр.

Несмотря на то что известно, когда появился термин «алгоритм», лингвисты по-разному пытались толковать его происхождение. Одни выводили algorism из греческих «алгирос» — «больной» и «арифмос» — «число». Правда, не понятно, почему числа «больные»? Другие склонялись к ещё более экстравагантному объяснению, связывая слово с неким мифическим древним испанским правителем King Algor of Castil. Свой вариант предлагает и Энциклопедический словарь Брокгауза и Ефрона (1890 г.). В нём «алгорифм» (кстати, до революции использовалось и написание «алгориѳм», через «фиту») производится от арабского слова «аль-горетм», т. е. «корень».

Но истину удалось установить не лингвистам, а историкам математики. Они доказали, что слово происходит от имени великого среднеазиатского учёного, автора популярнейшего на протяжении многих веков учебника по математике аль-Хорезми, жившего в первой половине IX в. В латинской транскрипции его имя записывается

Аль-Хорезми.

как Abu ‘Abd Allah Muhammad ibn Musa al-Khwarismi и означает «Мухаммад, сын Мусы, отец Абдуллы, родом из Хорезма». Хорезм — это историческая область на территории современного Узбекистана, центром которой является древний город Хива.

Почти все словари сходятся в том, что первоначально слово имело форму algorismi и лишь спустя какое-то время потеряло последнюю букву, приобретя более удобный для европейского произношения вид algorism. Позднее и оно, в свою очередь, неоднократно подвергалось искажениям, последнее из которых, скорее всего, связано со словом arithmetic, имеющим греческое происхождение. Уже в новом написании слово встречается в XVIII в. в одном из германских математических словарей «Vollstandiges mathematisches Lexicon», изданном в Лейпциге в 1747 г. Термин algorithmus объясняется в нём как понятие о четырёх арифметических операциях. Но такое значение не было единственным, ведь терминология математической науки в те времена ещё только формировалась. В частности, латинское выражение «algorithmus infinitesimalis» применялось к способам выполнения действий с бесконечно малыми величинами.

Постепенно все старинные значения вышли из употребления. К началу XX в. слово «алгоритм» уже означало «всякий арифметический или алгебраический процесс, который выполняется по строго определённым правилам», именно так оно объясняется в Большой советской энциклопедии (1926 г.).

Можно проследить процесс проникновения слова «алгоритм» в русский язык. Его не было ни в «Толковом словаре живого великорусского языка» В. И. Даля (1863 г.; а вот слово «алгебра» там присутствует), ни в «Толковом словаре русского языка» под редакцией Д. Н. Ушакова (1935 г.). Зато слово «алгорифм» есть и в популярном дореволюционном Энциклопедическом словаре братьев Гранат, и в уже упоминавшемся первом издании БСЭ, где говорится, что в Средние века так называли «правило, по которому выполняется то или другое из четырёх арифметических действий по десятичной системе счисления». А вот «Математическая энциклопедия» даёт его расширенное толкование: не только как искусства счёта в десятичной позиционной системе счисления, но и как самой десятичной системы.

То, что алгоритм воспринимался как термин сугубо специальный и малозначительный, подтверждается отсутствием соответствующих статей в менее объёмных изданиях. В частности, его нет даже в десятитомной Малой советской энциклопедии (1957 г.), не говоря об однотомных энциклопедических словарях. Но зато спустя десять лет в третьем издании БСЭ (1969 г.) «алгоритм» уже характеризуется как «одно из основных

СПОР АЛГОРИСМИКОВ И АБАЦИСТОВ

Многие века абак фактически был единственным средством для практических вычислений. В историю математики вошло упорное противостояние лагерей *абацистов* и *алгорисмиков* (первых иногда называли *гербекистами)*. Началось оно с появления в 1202 г. знаменитой «Книги абака» Леонардо Пизанского (Фибоначчи), где впервые было сказано, что кроме счёта на пальцах и на абаке возможен также счёт на бумаге. Алгорисмики предложили использовать для вычислений арабские цифры, только-только начавшие проникать в Европу благодаря переводам сочинений восточных математиков. Но прошло целых три столетия, прежде чем новый способ счёта окончательно утвердился, — столько времени потребовалось, чтобы выработать общепризнанные обозначения, усовершенствовать и приспособить к записи на бумаге методы вычислений. И хотя трактат Луки Пачоли «Сумма знаний по арифметике, геометрии, отношениям и пропорциональности» (1494 г.) подвёл черту под этим спором, счётная доска ещё долго не выходила из употребления.

Лука Пачоли.

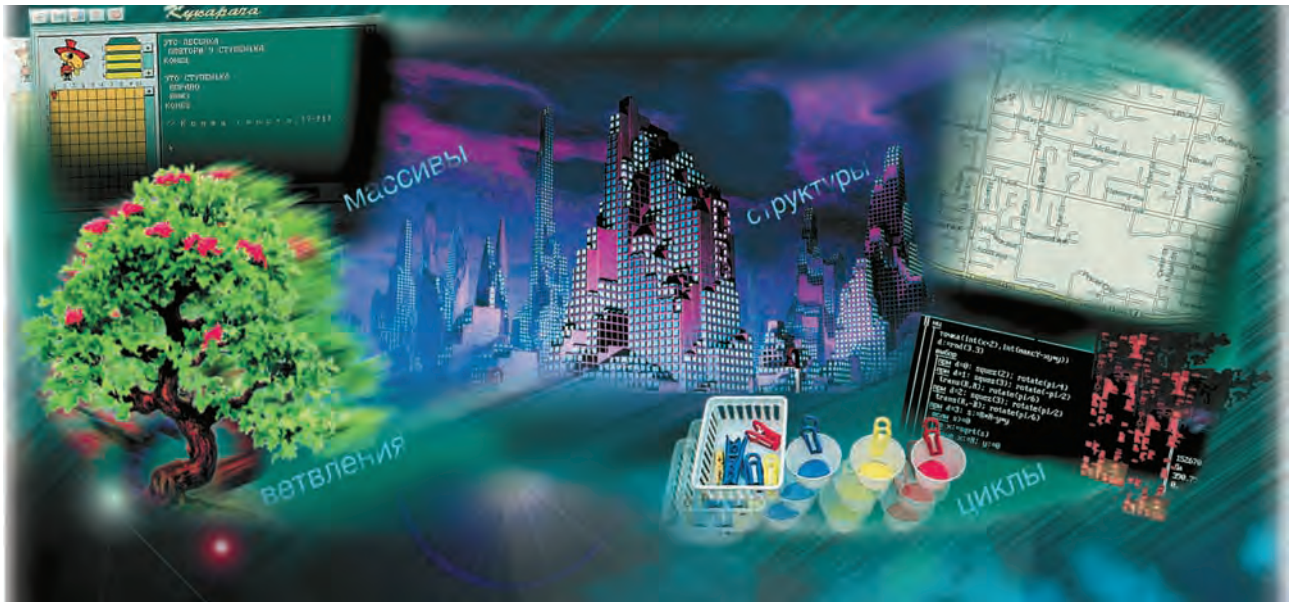
понятий (категорий) математики, не обладающих формальным определением в терминах более простых понятий, а абстрагируемых непосредственно из опыта». За 40 лет алгоритм из понятия, знакомого лишь узкому кругу специалистов, превратился в одно из ключевых понятий математики — и признанием этого стало включение слова не в энциклопедии, а в словари. Например, оно присутствует в академическом «Словаре русского языка» (1981 г.) — именно как термин из области математики.

Одновременно с развитием понятия алгоритма постепенно происходила и его экспансия из чистой математики в другие сферы. И начало ей положило появление компьютеров, благодаря которому слово «алгоритм» обрело новую жизнь. Например, в третьем томе «Детской энциклопедии» (1959 г.) о вычислительных машинах говорится немало, но они всё ещё воспринимаются скорее как некий атрибут светлого будущего, поэтому и алгоритмы ни разу не упоминаются на её страницах. Но в начале 70-х гг., когда компьютеры перестали быть экзотической диковинкой, слово «алгоритм» стремительно входит в обиход. В «Энциклопедии кибернетики» (1974 г.) в статье «Алгоритм» оно теперь связывается с реализацией на вычислительных машинах, а в «Военной энциклопедии» (1976 г.) даже появляется отдельная статья «Алгоритм решения задачи на ЭВМ».

За последние полтора-два десятилетия XX в. компьютер сделался неотъемлемым атрибутом нашей жизни, общеупотребительной становится и компьютерная лексика. Слово «алгоритм» в наши дни известно каждому. Оно настолько уверенно шагнуло в разговорную речь, что сейчас мы нередко встречаем на страницах газет, слышим в выступлениях политиков выражения вроде «алгоритм поведения», «алгоритм успеха» или даже «алгоритм предательства». Слово живёт, обогащается всё новыми значениями и смысловыми оттенками, так что, скорее всего, в словарях будущего его уже никогда не придётся снабжать пометкой «устаревшее».

Фибоначчи.

Дама Арифметика решает спор абацистов и алгоризмиков. Гравюра. 1504 г.



ОСНОВЫ ПРОГРАММИРОВАНИЯ

ПРОСТЕЙШИЕ ПРОГРАММЫ

Любой компьютерной программе — и простой, и сложной — совершенно необходима связь с внешним миром: с пользователем, в чьих интересах выполняется программа, или с другой программой, использующей нашу программу как вспомогательное средство.

Эта связь с внешним по отношению к программе окружением чаще всего выражается в передаче ей каких-либо значений — величин, которые программа использует в качестве исходных данных.

Даже в тех случаях, когда передача исходных данных формально отсутствует (например, в демонстрационных программах, которые показывают заранее запрограммированные картинки и сообщения), программа начинает работать лишь после того, как из внешнего мира поступит запускающий её сигнал в виде команды вызова программы.

Итак, необходимым элементом обычной программы является команда ввода, которая служит для передачи вводимого извне (пользователем или другой программой) значения и сохранения его в указанном месте памяти. Например, команда

```
ввод a
```

приостанавливает работу программы: она ждёт неограниченно долго,

когда пользователь введёт (например, наберёт на клавиатуре) полный набор символов вводимого значения. При работе с клавиатуры окончание ввода отмечается нажатием клавиши выполнения Enter. После нажатия на Enter программа продолжает свою работу. При этом самым первым её действием становится размещение введённых данных в области памяти, обозначенной в команде ввод буквой *a*, т. е. *a* — это имя величины, которая будет использована далее в программе.

Вот пример простейшей программы, которая вычисляет среднее арифметическое двух величин *a* и *b*:

```
ввод a
ввод b
c := (a+b) /2
```

После того как введены исходные данные *a* и *b*, программа вычисляет их среднее арифметическое и размещает результат этого вычисления в ячейке, названной *c*. Команду, содержащую символ :=, называют *присваиванием*.

Программа может считаться завершённой, если её результат нужен другим программам (которые «знают», что нужный им результат находится в ячейке *c*). Но если результат нужен пользователю, который хочет увидеть его выведенным на экран или распечатанным на принтере, то программу нужно завершить командой вывода величины *c*:

```
ввод  a
ввод  b
c := (a+b) /2
вывод  c
```

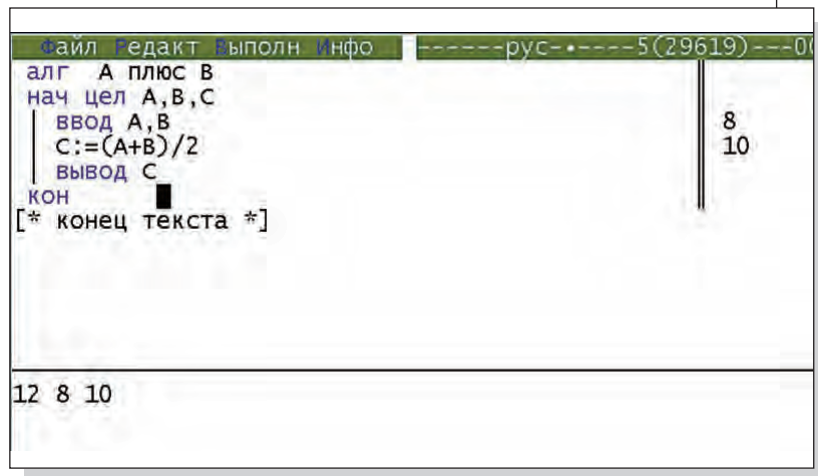
Далеко не всегда результатом выполнения программы является числовое значение или текстовое сообщение, которое программа выдаёт пользователю. Нередко выполнение программы состоит в осуществлении действий, изменяющих среду, в которой она работает. Такая программа не вырабатывает и не выдаёт никаких значений.

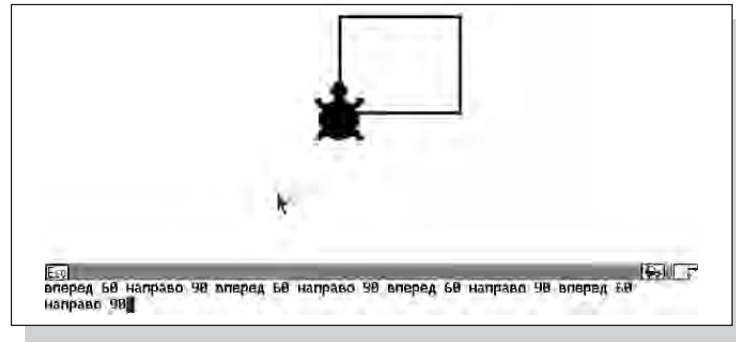
Например, написанная на языке ЛОГО программа переместит Черепашку по квадратному маршруту и даже вычертит на экране этот квадрат, если ввести угол 90°, но никакое значение или сообщение в результате работы этой программы не будет выдано:

```
вводчисла : сторона
вводчисла : угол
опустить перо
вперёд : сторона
вправо : УГОЛ
вперёд : сторона
вправо : УГОЛ
вперёд : сторона
вправо : УГОЛ
вперёд : сторона
вправо : УГОЛ
```

Несколько пояснений:

- непосредственно перед именами величин (без пробела) в ЛОГО принято ставить двоеточие;
- слово «числа», написанное вслед за словом «ввод», необходимо для того, чтобы программа смогла распознать,





Каким будет значение величины один после выполнения следующей программы:

```
ввод два
ввод три
один := два + три - 4
```

при условии, что при первой команде ввода пользователь набирает на клавиатуре число 5, а при второй — число 6?
Ответ: один = 7.

что вводимая величина является числом (существует также команда «ввод текста» для ввода текстовых значений);

• наконец, вместо предложенного здесь способа ввода в ЛОГО обычно указывают необходимые значения непосредственно в команде (в качестве её параметров). Таким образом, более привычной будет следующая запись той же программы (если перемещение выполняется на 20 шагов, а поворот — на 90°):

```
вперёд 20
вправо 90
вперёд 20
вправо 90
вперёд 20
вправо 90
вперёд 20
вправо 90
```

Программы, в которых предусматривается выработка результирующего значения для последующего использования его в других программах, называют *функциями*.

Программы, которые не вырабатывают результирующее значение, называют *процедурами*.

Приведённые здесь в качестве примеров программы относят к простейшим, потому что они *линейны*: последовательность команд фиксирована, при каждом осуществлении программы (независимо от значений исходных данных) вторая команда выполняется вслед за первой, третья после второй и т. д.

Большинство программ устроено сложнее. В них встречаются ветвления, которые заставляют менять порядок следования команд, и циклы, которые организуют многократные повторения отдельных групп команд.

УПРАВЛЯЮЩИЕ СТРУКТУРЫ — ВЕТВЛЕНИЯ

Простейшие программы линейны: команды данных программ выполняются в порядке, в котором они записаны, т. е. последовательно друг за другом. Иногда такую группу последовательно выполняемых команд называют *серией*.

Программирование было бы очень скучным и трудоёмким занятием, если бы существовали только линейные программы. Во всех языках программирования предусматриваются управляющие команды, которые позволяют менять порядок выполнения команд в программе. Команды, управляющие порядком исполнения программы, называют *управляющими структурами*.

Существует три основных типа управляющих структур:

- *серии*;
- *ветвления* (условные команды);
- *циклы* (команды повторения).

Общий вид условной команды:

```
если <условие>
| то <команда-1>
| иначе <команда-2>
всё
```

Примеры условий: $a = 15$, $a + 1 < c$, $15 = 5$. Первое из них верно (т. е. даёт ответ «да») лишь в случае, когда величина имеет значение 15. Второе условие истинно для всех a и c, при которых может выполняться неравенство $a + 1 < c$. А вот третье условие не выполняется никогда. Говорят ещё, что такое условие тождественно-ложно.

Подчёркиванием выделяется ключевое слово языка, а в угловых скобках указывается языковая категория, требующая конкретизации — определения или толкования.

Условие обычно имеет вид утверждения, проверка которого даёт однозначный ответ — «да» или «нет» (0 или 1, истина или ложь).

Первая операция, которая выполняется любой условной командой, — это проверка утверждения-условия. Его значение используется для выбора одной из двух *ветвей* условной команды. Ветвь, на которой стоит <*команда-1*>, называют иногда ветвью то, а другую ветвь — <*команда-2*> — ветвью иначе.

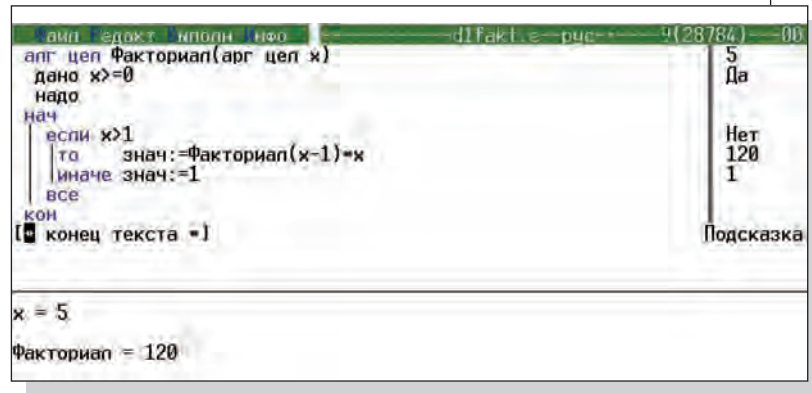
Например,

```
a := 10
  если a > 5
  | то вывод "всё в порядке"
  | иначе вывод "аварийная ситуация"
  всё
c := a + 1
```

программа выведет на экран «спокойное» сообщение, поскольку истинность условия `a > 5` обеспечена действием предшествующей команды присваивание `a := 10`. Сообщение `"аварийная ситуация"` при этом не выводится, так как следующая исполняемая команда — присваивание — `c := a + 1`.

Обычно как на ветви то, так и на ветви иначе можно задавать не только одну команду. Однако если группа команд, которые желательно выполнить на ветви условной команды, может понадобиться в программе неоднократно, то рекомендуется создать процедуру из этой группы команд («запроцедурить» её), и тогда на соответствующей ветви выражение <*команда*> записывается как вызов процедуры, т. е. как простое упоминание имени процедуры.

В языке программирования КуМир, используемом, в частности, для управления учебным исполнителем Робот, предусматриваются условия, проверяющие значения двух физических величин, характеризующих среду, в которой работает этот исполнитель, — радиацию и температуру. Текущее значение этих величин находится в областях памяти (ячейках), названных ключевыми словами — «радиация» и «температура». Поэтому в этом языке допустимы, например, такие условия: радиация <0,7, температура >100.



Ветви то и иначе рекомендуется записывать с небольшими смещениями вправо от соответствующего ключевого слова если и размещать знак окончания условной команды — слово всё — строго под её началом (соответствующим слову если). Эта рекомендация (но не требование языка!) связана с тем, что на месте выражений <*команда-1*> и <*команда-2*> могут записываться, в свою очередь, ветвления, образующие цепочки *вложенных* условных команд. В этом случае аккуратные смещения ветвей условных команд помогают человеку, читающему программу, разобраться в самых сложных ветвлениях без использования громоздких и избыточных блок-схем. Вот как выглядит описание программы, которая находит максимум среди трёх чисел *a*, *b* и *c*:

```
если a > b
| то
| если a > c
| | то max := a
| | иначе max := c
| всё
| иначе
| если b > c
| | то max := b
| | иначе max := c
| всё
всё
```

Условные команды записываются иногда и в сокращённой форме. Если, например, команду на ветви то следует выполнять только при положительном результате проверки условия, а в случае отрицательного результата никаких дополнительных действий предпринимать не нужно, то условную команду можно записать так:

```
a := 1
если a < 3
| то b := "подходящий результат"
всё
c := 10
```

Здесь ветвь иначе в условной команде отсутствует, а величина *b* получает текстовое значение "подходящий результат" только в том случае, если условие истинно.

Вот программа, которая даёт совет, как выбрать одежду по погоде:

```
вывод "надень рубашку"
если t < 0
| то вывод "надень шубу"
всё
```

При любой температуре она порекомендует надеть рубашку, но в холодный день добавит ещё один совет — надеть шубу.

Уже приводился пример, когда одна условная команда выполнялась внутри другой. Случаи подобного рода нередки, если приходится выбирать одну из нескольких возможностей.

В следующем примере проектирования программы задаётся (например, с клавиатуры) код дня недели по таблице:

понедельник	1
вторник	2
среда	3
четверг	4
пятница	5
суббота	6
воскресенье	7

Необходимо этот код преобразовать в текстовое значение величины день.

Такую программу можно записать, конечно, с помощью классической команды если. Эта программа представляет собой сложную вложенную условную команду.

```
ввод код
если код = 1
| то день := "понедельник"
| иначе
| если код = 2
| | то день := "вторник"
| | иначе
| | если код = 3
| | | то день := "среда"
| | | иначе
| | | если код = 4
| | | | то день := "четверг"
| | | | иначе
| | | | если код = 5
| | | | | то день := "пятница"
| | | | | иначе
| | | | | если код = 6
| | | | | | то день := "суббота"
| | | | | | иначе
| | | | | | если код = 7
| | | | | | | то день := "воскресенье"
| | | | | | | иначе вывод "ошибка кода"
| | | | | | всё
| | | | | всё
| | | | всё
| | | всё
| | всё
| всё
всё
```

Однако подобные ситуации столь часты в программировании, что ряд языков предусматривает специальные условные команды, позволяющие выбирать одну ветвь из нескольких возможных.

С помощью команды выбор наша программа записывается короче и нагляднее:

```
выбор
| при код = 1: день := "понедельник"
| при код = 2: день := "вторник"
| при код = 3: день := "среда"
| при код = 4: день := "четверг"
| при код = 5: день := "пятница"
| при код = 6: день := "суббота"
| при код = 7: день := "воскресенье"
| иначе вывод "ошибка кода"
всё
```

При проектировании программы со сложными условиями следует проявлять внимательность: быть может, всё решается гораздо проще, чем кажется вначале. Представим себе тест, где компьютер должен распознать один из четырёх предложенных ответов. Ответы закодированы числами 1, 2, 3 и 4. Именно одно из этих чисел программа ожидает от пользователя, вводящего с клавиатуры значение после команды ввод. Предположим, что правильным в рассматриваемом тесте оказывается ответ, закодированный числом 3. При всей внешней схожести этой задачи с кодированием дней недели она решается существенно проще:

```
ввод код
если код = 3
| то вывод "ответ верный"
| иначе вывод "вы не правы"
всё
```

УПРАВЛЯЮЩИЕ СТРУКТУРЫ — ЦИКЛЫ

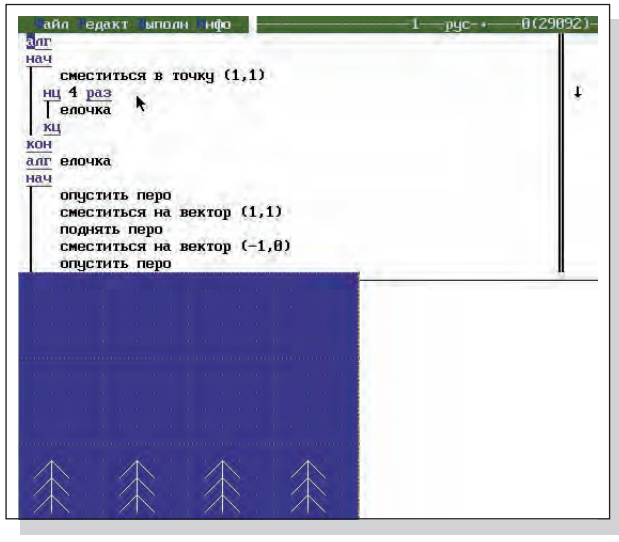
Для выполнения многократно повторяющихся операций в языках программирования существуют *команды повторения,* или *циклы.*

В языках программирования встречается четыре типа циклов (записываемых иногда в разной форме), но в каждом из них присутствуют обязательные элементы:

- *заголовок цикла*;
- *условие завершения цикла*;
- *тело цикла*;
- *счётчик числа повторений.*

Заголовком служит ключевое слово, отмечающее начало команды повторения. Для каждого типа циклов в качестве заголовка используется своё ключевое слово.

Условие завершения цикла записывается обычно в такой же форме, как и условие команды ветвления. Условие — это утверждение, требующее проверки, в результате которой может быть получен единственный из двух возможных ответов — «да» или «нет» (0 или 1, истина или ложь). Например, утверждение-условие $a < 10$ истинно при значениях a, равных 0, 3, 7 и ложно при a, равных 12, 100, 105.



Группа (серия) команд, выполняемых при каждом повторении цикла, образует тело цикла. В нём могут присутствовать и другие управляющие структуры — команды ветвления и повторения. Если в теле одного цикла работает другой цикл, такие команды повторения называют *вложенными.* Границы тела цикла отмечаются скобками. Например, в языке программирования КуМир открывающая скобка, или начало цикла, — это ключевое слово нц, закрывающая скобка, или конец цикла, — слово кц.

Наконец, элемент, который может отсутствовать в некоторых типах циклов, — это счётчик числа повторений, часто оказывающий хорошую службу программисту или пользователю, когда важно зафиксировать число требуемых повторений.

Итак, четыре типа циклов, используемые в языках программирования:

- циклы с фиксированным числом повторений;
- циклы с предусловием;
- циклы с постусловием;
- циклы с задаваемым шагом изменения переменной цикла.

Наиболее простыми считаются циклы с фиксированным числом повторений. Их общий вид:

```
нц <выражение> раз
| <серия>
кц
```

По такой команде сначала определяется числовое значение арифметического выражения <*выражение*> (оно должно быть целым), а затем найденное значение n служит для фиксации числа повторений. Таким образом, хотя в этом типе циклов условие завер-

шения формально отсутствует, его участие в выполнении команды очевидно: цикл завершается, как только неявно формируемое и увеличивающееся на единицу при каждом повторении тела цикла значение счётчика сравняется с *n*. Так, цикл

```
a := 10
b := 12
i := 1
s := 0
нц a + b раз
| s := s + m [i]
| i := i+1
кц
```

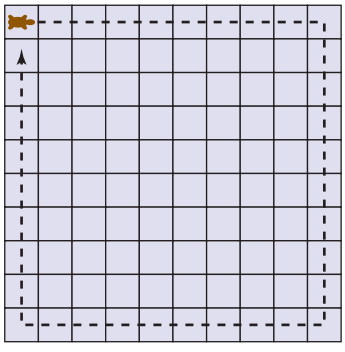
суммирует первые 22 значения массива *m*, от *m*[1] до *m*[22].

В языке управления исполнителем Кукарача, где в качестве тела цикла <*серия*> допускается единственная команда, цикл с фиксированным числом повторений выглядит совсем просто:

```
ПОВТОРИ n <команда>
```

Так, заставить погулять Кукарачу, который понимает и умеет выполнять четыре команды — ВВЕРХ, ВНИЗ, ВЛЕВО и ВПРАВО, перемещаясь при их выполнении на одну ячейку в заданном направлении по квадратному полю размером 10×10, можно с помощью всего четырёх команд:

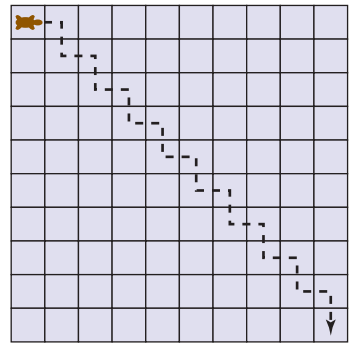
```
ПОВТОРИ 9 ВПРАВО
ПОВТОРИ 9 ВНИЗ
ПОВТОРИ 9 ВЛЕВО
ПОВТОРИ 9 ВВЕРХ
```



Как же быть, если потребуется выполнение нескольких команд в теле такого цикла? Для этого есть две возможности:

• если группа команд, которые желательно выполнить в теле команды повторения, может понадобиться в программе неоднократно, то рекомендуется создать процедуру из этой группы команд, и тогда на соответствующей ветви выражение <*команда*> записать как вызов процедуры. Вот как Кукарача решает задачу спуска вниз по лесенке:

```
ЭТО СПУСК
  ПОВТОРИ 9 ЛЕСЕНКА
КОНЕЦ
ЭТО ЛЕСЕНКА
  ВПРАВО
  ВНИЗ
КОНЕЦ
```



• если такая группа используется однократно, в теле команды цикла, то тогда можно и не прибегать к запроцедуриванию, а заключить группу в фигурные скобки {} (по существу, те же нц и кц); все команды, окаймлённые открывающей и закрывающей скобками, воспринимаются компьютером как единая команда:

```
ЭТО СПУСК
   ПОВТОРИ 9
   {
      ВПРАВО
      ВНИЗ
   }
КОНЕЦ
```

В цикле с предусловием число повторений тела не всегда легко определить заранее. Да в этом и нет необходимости — завершение цикла отслеживается условием, которое записывают в заголовке цикла:

```
нц пока <условие>
| <тело цикла>
кц
```

Такая команда начинается с проверки условия. Если оно истинно, то выполняется тело цикла и вновь проверяется условие — до тех пор пока оно не перестанет быть истинным. Это последнее условие и является признаком завершения команды повторения. Такая последовательность операций предполагает, что в теле цикла происходит изменение значения той переменной, которая фигурирует в условии. Если условие постоянно истинно, то цикл становится бесконечным, если, наоборот, условие изначально ложно, то тело не выполнится ни разу, а управление перейдёт к команде, записанной вслед за циклом.

Задачу, поставленную в программе, суммирующей первые 22 значения списка *m*, от *m*[1] до *m*[22], можно решить и с помощью цикла с предусловием:

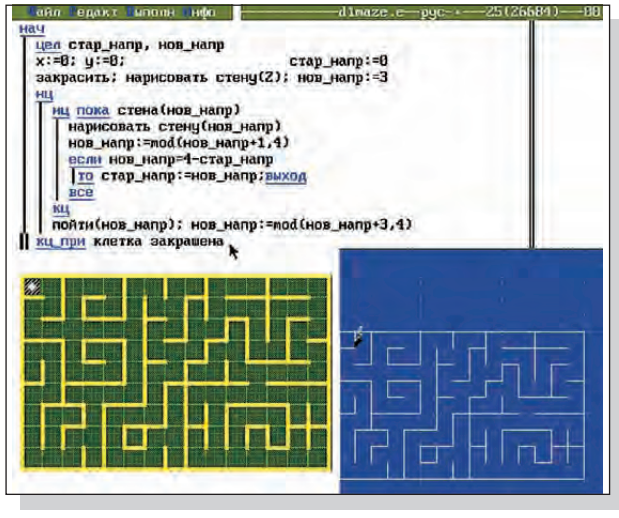
```
a := 10
b := 12
i := 1
s := 0
нц пока i < a + b
| s := s + m [i]
| i := i+1
кц
```

Цикл с постусловием имеет следующий вид:

```
нц
| <тело цикла>
кц при <условие>
```

Он отличается от предыдущего тем, что условие проверяется не перед началом тела, а в конце повторяемой серии. Это означает, что в любом случае (независимо от условия) тело цикла будет выполнено, по крайней мере, один раз. Вариант программы с циклом с предусловием, записанный в виде цикла с постусловием, имеет вид:

```
a:= 10
b:= 12
i:= 1
s:= 0
нц
| s:= s + m [i]
| i:= i + 1
кц при i < a + b
```



В циклах с задаваемым шагом изменения переменной цикла появляется ещё два понятия — *переменная* цикла и *шаг* её изменения при каждом повторении тела. В таком цикле нет необходимости устанавливать начальное значение счётчика (или, как говорят, *инициализировать* значение этой переменной), поскольку и на-

чальное, и конечное значения этой переменной и шаг её изменения могут быть заданы в команде повторения, имеющей такой вид:

```
нц для <переменная> от <выражение-1>
 | до <выражение-2>
 | [шаг <выражение-3>]
 | <тело цикла>
кц
```

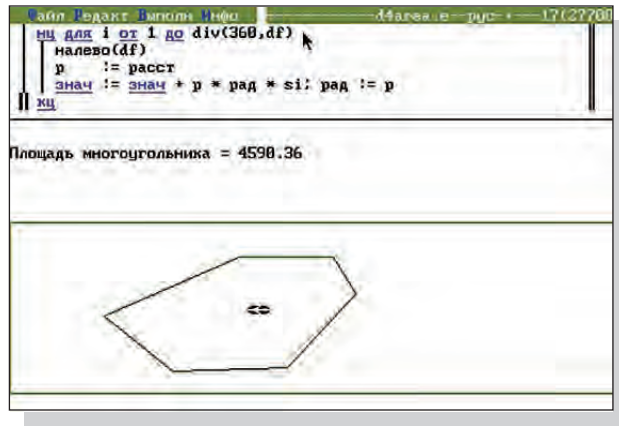
Вообще говоря, переменная цикла совсем не обязательно должна совпадать со счётчиком. Однако в таком виде переменную цикла используют чаще всего. В общем представлении этого цикла элемент заголовка окаймлён квадратными скобками. Это означает, что данных элементов в конкретном представлении команды может и не быть. Так, отсутствие элемента шага <*выражение-3*> заставляет команду повторения менять значение цикла при каждом выполнении тела на 1. Тем самым значение выражения <*выражение-3*> по умолчанию предполагается равным 1. Цикл с задаваемым шагом в нашем простом примере записывается так:

```
a := 10
b := 12
s := 0
нц для i от 1 до a + b
 | s := s + m [i]
кц
```

Здесь переменной цикла объявлена величина *i*. Необходимости в изменении счётчика в теле цикла нет, так как команда *i* := *i* + 1 предполагается определением цикла (единица в качестве шага изменения в этом примере принята по умолчанию).

Последний тип цикла наиболее компактен. Если же учесть, что на месте элементов <*выражение-1*>, <*выражение-2*> и <*выражение-3*> можно записывать сложные формулы, то становятся очевидными широкие возможности этого типа команды повторения.

Заметим, что во всех записях примеров программ с командами повторения тело цикла смещено вправо относительно команд нц и кц, расположенных на одной вертикали. Эта рекомендация (но не требование языка!)



связана с тем, что в теле цикла могут встречаться, в свою очередь, циклы и ветвления, образуя цепочки вложенных управляющих структур. В этом случае аккуратные смещения команд повторения и условных команд помогают человеку, читающему программу, разобраться в самых сложных ветвлениях без использования громоздких и избыточных блок-схем.

Если вооружить математика карандашом и бумагой и попросить его решить две задачи, одна из которых — вычисление сложного интеграла, а вторая — сложение трёх сотен чисел, то результат получится удивительный: первую задачу математик, конечно, решит, а во второй почти наверняка допустит ошибку. Хотя каждое отдельное сложение без труда выполнит даже младший школьник. Поэтому если математику придётся делать выбор, какую из этих двух задач он предпочёл бы решать, то, без сомнения, он остановится на первой: со сложной задачей иметь дело интереснее, она соответствует его интеллекту. А вторая

задача — это скучная, но тем не менее требующая высокой сосредоточенности рутинная работа, и потому в ней неизбежны ошибки.

Именно здесь — между оригинальной задачей, требующей творческого к себе отношения, и однообразной, трудоёмкой работой — пролегает граница человеческих и машинных возможностей.

Машина без усталости, напряжения и негативных эмоций справится не только с тремя сотнями выполняемых подряд однотипных арифметических операций, но и с тремя тысячами, и с тремя миллионами действий.

МАССИВЫ

Одна из фундаментальных структур данных, применяемых в информатике, — *массивы.* Это понятие часто встречается и в иных, синонимичных терминах — «векторы», «матрицы» и др. Например, язык программирования КуМир (используемый здесь для иллюстрирования различных механизмов программирования) оперирует термином «таблица» или «табличные величины», т. е. конечные упорядоченные совокупности однотипных элементов.

Вот примеры массивов-таблиц:

1) `вещ таб зарплата [1:20]`
(табличная величина названа *зарплата,* её элементы — 20 вещественных чисел);

2) `лит таб забег [1:7]`
(в таблицу *забег* входит семь элементов — текстовых, или литерных, величин; в ней собраны фамилии участников легкоатлетического забега);

3) `лит таб кросс [1:4, 1:7]`
(таблица *кросс* состоит из четырёх строк и семи столбцов; в ней перечислены фамилии всех участников легкоатлетического кросса, распределённых по четырём забегам, в каждом из которых по семь бегунов).

Описания данных таблиц (в соответствии с правилами языка КуМир) построены следующим образом:

1) сначала указывается тип элементов, образующих таблицу; первая таблица составлена вещественными числами (вещ), вторая — литерными величинами (лит);

2) затем следует тип табличной величины — ключевое слово таб;

3) далее задаётся имя описываемой табличной величины *(зарплата, забег, кросс);*

4) завершает описание конструкция, определяющая размеры таблицы.

Поскольку таблица относится к *составным* (а не *простым*) структурам данных, то именно конструкция, задающая размеры таблицы, является специфической в её описании. В квадратных скобках дана пара целых чисел, разделённых двоеточием. Первое число в этой паре указывает номер первого элемента таблицы, второе — номер её последнего элемента (предполагается последовательная порядковая нумерация). Это означает, что если первый элемент в паре является единицей, то её второй элемент выражает число элементов в таблице. Так обстоит дело в первых двух приведённых примерах: в таблице *зарплата* – 20 элементов, в таблице *забег* — 7.

Указанные таблицы — *линейные,* или *одномерные.* Этот факт отражён тем, что в квадратных скобках указана только одна пара целых чисел. Номер элемента в линейной таблице называется его *индексом.*

Например, таблица *забег* имеет вид:

```
лит таб забег [1:7]
```

Значения элементов	Индексы элементов
Кошкин	1
Мышкин	2
Блошкин	3
Плошкин	4
Мошкин	5
Ложкин	6
Рожкин	7

Фамилия Мошкин — это значение 5-го элемента таблицы *забег*. С помощью индекса этот факт записывается так:

```
забег [5] = «Мошкин»
```

А вот таблица *кросс* — *двухмерная:* в ней четыре строки, и в каждой — семь элементов. Эти параметры легко прочитываются в двух парах целых чисел, стоящих в квадратных скобках после имени таблицы *кросс*. Аналогичным образом описываются трёхмерные таблицы и таблицы более высоких размерностей. Чтобы указать элемент таблицы, имеющей размерность *n*, надо использовать не один индекс, а *n* — по значению размерности. Так, в двухмерной таблице *кросс*

```
кросс [3, 1] = «Иванов»,
кросс [2, 4] = «Залкинд»,
кросс [4, 5] = «Лосев».
```

Поскольку каждый элемент однозначно в ней распознаётся с помощью индексов, то над элементами таблицы допустимы все операции, применяемые к величинам соответствующего типа. Например, по команде

```
зарплата [17] := зарплата [1] + зарплата [3]
```

компьютер выберет из таблицы *зарплата* два числа — *зарплата* [1] и *зарплата* [3], сложит их и результат присвоит элементу этой же таблицы с индексом 17, а по команде

```
вывод кросс [4, 7]
```

напечатает слово — Зайцев.

Упорядоченность элементов таблицы позволяет эффективно осуществлять групповые операции. Вот как

```
лит таб кросс [1:4, 1:7]
```

Кошкин	Мышкин	Блошкин	Плошкин	Мошкин	Ложкин	Рожкин
Галкин	Палкин	Ёлкин	Залкинд	Чалкин	Салкин	Малкин
Иванов	Петров	Сидоров	Семёнов	Кузьмин	Фомин	Егоров
Волков	Бобров	Лисицын	Енотов	Лосев	Сурков	Зайцев
1	2	3	4	5	6	7

можно, воспользовавшись циклом, выполнить увеличение сразу для всех элементов таблицы *зарплата* на 9 %:

```
алг увелич
(арг рез вещ зарплата [1:20])
нач цел i
| нц для i от 1 до 20
| | зарплата[i]:= зарплата[i]*1.09
| кц
кон
```

Здесь *i* — имя величины целого типа (управляющая переменная цикла), которая в этом примере увеличивается на 1 при каждом выполнении тела цикла от 1 до 20. Таблица *зарплата* является в этом алгоритме одновременно и аргументом, и результатом.

В описанных выше трёх примерах таблицы имели фиксированное число элементов. Вместе с тем лучше использовать в качестве этих количественных значений переменные величины. Если, например, в описании таблицы *зарплата* указать не фиксированные ограничения, а поставить после двоеточия целую переменную *n*, то такую таблицу можно применить в алгоритме при любом числе элементов:

```
алг повышение
(рез вещ зарплата[1:n])
нач цел i
| нц для i от 1 до n
| | зарплата[i]:= зарплата[i] * 1.09
| кц
кон
```

Таблицы удобны при поиске среди множества элементов по заданным признакам, при определении наибольших и наименьших значений, в разнообразных сортировках.

Пример: найти в упорядоченном множестве числовых положительных элементов, т. е. в таблице, наибольший элемент и его индекс — номер в этом множестве.

```
алг максимум
(арг вещ зарплата[1:n],
рез вещ мак, цел инд)
нач цел i; мак:= 0; инд:= 1
| нц для i от 1 до n
| | если зарплата[i] > мак
| | | то мак:= зарплата[i]; инд:= i
| | всё
| кц
кон
```

СТРУКТУРЫ ДАННЫХ

«Алгоритмы + структуры данных = программы» — так называется книга Никлауса Вирта, известного швейцарского специалиста по программированию, блестящего педагога, автора языков программирования Pascal, Modula-2, Oberon. Обе составляющие программы, обозначенные Виртом, в равной степени важны. Не только несовершенный алгоритм, но и неудачная организация работы с данными может привести к замедлению работы программы в десятки, а иногда и в миллионы раз. С другой стороны, владение теорией программирования и умение систематически применять её на практике позволяют разрабатывать эффективные и в то же время красивые программы. Искусство работы с данными является важнейшей составляющей искусства программирования.

Никлаус Вирт.

МАССИВ КАК БАЗОВАЯ СТРУКТУРА

Как известно, две главные составные части компьютера — это процессор и оперативная память. Оперативная память компьютера состоит из элементов, каждый из которых имеет свой адрес. В современных компьютерах минимальным адресуемым элементом является байт. Важно также понятие машинного слова — это несколько подряд идущих байтов, обозначающих одно целое число. Разрядность машинного слова (количество битов) в современных процессорах составляет 32 или 64 бит, соответственно машинное слово включает 4 или 8 байт. Процессор все действия производит со словами (машинное

слово — наиболее естественный элемент данных для процессора). Адресом машинного слова является наименьший из адресов байтов, составляющих его. Таким образом, адреса машинных слов кратны четырём или восьми в зависимости от разрядности процессора.

Оперативная память, с точки зрения программиста, — это массив элементов. Каждый элемент массива имеет свой адрес, который обычно называют индексом. Любой элемент можно прочитать или записать мгновенно, за одно элементарное действие. Структуры данных, в которых непосредственный доступ возможен к любым их элементам, называют структурами данных с прямым, или произвольным, доступом (*англ.* random access). Примером такой структуры является массив (или, как его ещё называют, таблица). Структуры данных, в которых непосредственный доступ возможен лишь к одному или нескольким элементам (для доступа к остальным элементам надо произвести дополнительные действия), называются структурами последовательного доступа. Бытовым примером такой структуры является магнитофонная кассета, на которой записаны песни. В любой момент времени можно слушать лишь одну песню. Чтобы добраться до иных музыкальных фрагментов, надо перемотать ленту в ту или другую сторону.

Третья важнейшая составляющая компьютера — магнитный диск с логической точки зрения также является массивом. Элементарной единицей чтения и записи для него является блок. Размер блока зависит от конструкции конкретного диска, обычно он кратен числу 512. За одну элементарную операцию можно прочесть или записать один блок с заданным адресом.

Итак, наиболее важные запоминающие устройства компьютера — оперативная память и магнитный диск представляют собой массивы. Массив для программиста — это некая данность, так же как, например, для математика данность — целые числа. Работа с элементами массива

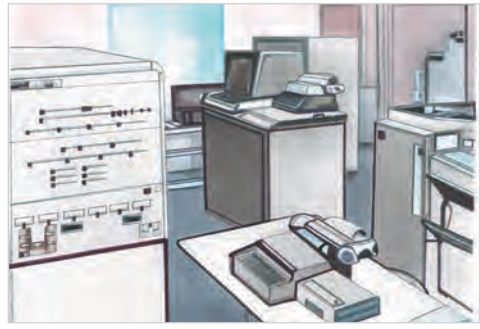
осуществляется исключительно быстро, они доступны без каких-либо предварительных действий.

Тем не менее одних массивов недостаточно для написания эффективных программ. Большинство реальных задач требует более сложной организации работы с данными. Массивы выступают лишь в роли базовых элементов, из которых строятся более сложные структуры.

В качестве примера можно рассмотреть библиотеку. С точки зрения программирования библиотека — это структура данных для работы с книгами. Стеллажи в библиотеке с пронумерованными полками являются аналогом массива. Роли элементов этого массива играют книги на полках, роли индексов — номера полок. Но библиотека — это не только набор стеллажей. Её структура включает каталоги (алфавитный и систематический). Существуют также системы доставки запросов в книгохранилище (например, пневматическую почту) и доставки книг читателям, наконец, персонал библиотеки, выполняющий запросы читателей, а также осуществляющий хранение и пополнение библиотечного фонда.

ПРОСТЕЙШИЕ СТРУКТУРЫ ДАННЫХ

Наиболее важными из простейших структур данных являются *стек* и *очередь*. Эти структуры встречаются в программировании буквально на каждом шагу, в самых разнообразных случаях. Особенно интересен стек, который может иметь весьма неожиданные применения. При разработке серии компьютеров IBM 360 в начале 70-х гг. XX в. фирма IBM совершила драматическую ошибку, не предусмотрев аппаратную реализацию стека. Эта серия содержала много других неудачных решений, но, к сожалению, была скопирована в Советском Союзе под названием ЕС ЭВМ («Единая серия»).



Машина серии ЕС ЭВМ 1022.

ОЧЕРЕДЬ

Очередь в программировании аналогична очереди в повседневной жизни. Она содержит элементы, как бы выстроенные друг за другом в цепочку. У очереди есть начало и конец. В отличие от обычной очереди, которую всегда можно при желании покинуть, в программе из середины очереди удалять элементы нельзя.

Очередь можно представить в виде трубки. В один конец трубки (конец очереди) добавляются шарики — элементы очереди, из другого конца (начало очереди) они извлекаются. Элементы в середине очереди, т. е. шарики внутри трубки, недоступны. Таким образом, концы трубки несимметричны, шарики внутри неё движутся только в одном направлении.

Если продолжить аналогию с трубкой, то ясно, что в принципе можно добавлять элементы в оба её конца и забирать их также из обоих концов. Подобная структура данных в программировании тоже существует, её название — «дек» (от *англ.* Double Ended Queue — «очередь с двумя концами»), но используется она значительно реже, чем очередь.

Роль очереди в программировании почти соответствует её роли в обыденной жизни. Очередь практически всегда связана с обслуживанием запросов в тех случаях, когда они не могут быть выполнены сразу. Например, когда человек, набирая текст, нажимает клавишу на клавиатуре (т. е. просит машину выполнить какое-то действие), то действие компьютера состоит в добавлении к тексту одного символа или же перерисовке области экрана, прокрутке окна, переформатировании абзаца и т. п.

Любая, даже самая простая, операционная система всегда в той или иной степени многозадачна. Это значит, что в момент нажатия пользователем клавиши операционная система может быть занята какой-либо другой работой. Тем не менее операционная система ни в коем случае не имеет права проигнорировать команду пользователя. Поэтому при нажатии клавиши на какое-то время прекращается выполнение предыдущей задачи. Машина запоминает текущий момент в работе и переключается на обслуживание запроса. Эта операция должна быть очень короткой, чтобы не нарушилось выполнение других задач. Поэтому команда, отдаваемая нажатием на клавишу, просто добавляется в конец очереди за-

просов, ждущих выполнения. После этого компьютер возвращается к той работе, которая была прервана нажатием на клавишу. Запрос, поставленный в очередь, будет выполнен не сразу, а только когда наступит его черёд. Каждая программа имеет свою очередь запросов. По мере продвижения в работе она считывает очередной запрос из начала очереди и выполняет его. Очередь поддерживается операционной системой.

Подход, состоящий не в прямом запуске нужных процедур, а в посылке сообщений, которые добавляются в очередь запросов, имеет много преимуществ и является одной из характеристик объектно-ориентированного программирования, широко распространившегося на рубеже XX—XXI вв. Например, если программе нужно завершить свою работу по какой-либо причине, она не вызывает сразу команду завершения, которая опасна, потому что нарушает логику работы и может привести к потере данных. Вместо этого программа посылает самой себе сообщение о необходимости завершения работы, которое ставится в очередь запросов и выполняется после запросов, поступивших ранее.

РЕАЛИЗАЦИЯ ОЧЕРЕДИ НА БАЗЕ МАССИВА

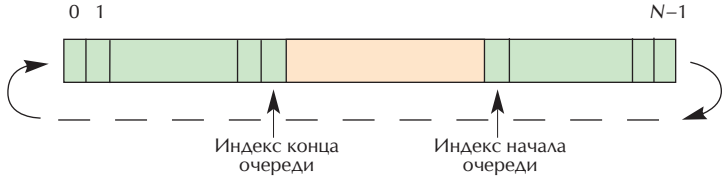
Если принять массив как изначальную данность, все остальные структуры данных должны реализовываться на его базе. Реализация одной структуры данных на основе другой — это описание её работы в терминах. Реализация обязательно включает в себя описание «идеи реализации» (т. е. того, как хранить элементы реализуемой структуры в базовой структуре, какие дополнительные переменные нужны) и набор подпрограмм, каждая из которых моделирует определённое предписание реализуемой структуры при помощи предписаний базовой структуры.

Реализация может быть многоэтапной, а массив не всегда выступает в качестве непосредственной базы реализации. Вообще, все реализации структур данных делятся на два класса — непрерывные и ссылочные. В непрерывных реализациях элементы структуры данных хранятся в массиве подряд, без промежутков, занимая таким образом непрерывную часть массива. В ссылочных, или списковых, реализациях элементы структуры данных хранятся в хаотическом порядке, в массиве или непосредственно в динамической оперативной памяти, которую программисты называют «кучей». При этом вместе с каждым элементом хранятся ссылки на соседние элементы. В случае списка таких ссылок может быть две (на следующий и предыдущий) или одна (ссылка на следующий элемент). Элементы списка как бы образуют связную цепочку, и чтобы добраться до элементов в середине списка, нужно пройти по этой цепочке, начиная с её первого элемента.

В случае очереди наиболее популярны две реализации — непрерывная на базе массива, которую называют также реализацией на базе кольцевого буфера, и ссылочная реализация, или реализация на базе списка. При непрерывной реализации очереди в качестве базы выступает массив фиксированной длины N, таким образом очередь ограничена и не может содержать более N элементов. Индексы элементов массива изменяются в пределах от 0 до $N-1$. Кроме массива реализация очереди содержит три простые переменные: индекс начала очереди, индекс конца очереди, число элементов очереди. Элементы очереди хранятся в отрезке массива от индекса начала до индекса конца.

При добавлении нового элемента в конец очереди индекс конца сначала увеличивается на единицу, а затем новый элемент записывается в ячейку массива с этим индексом. Аналогично при извлечении элемента из начала очереди ячейка массива с индексом начала очереди запоминается в качестве результата операции, а затем индекс начала очереди увеличивается на единицу.

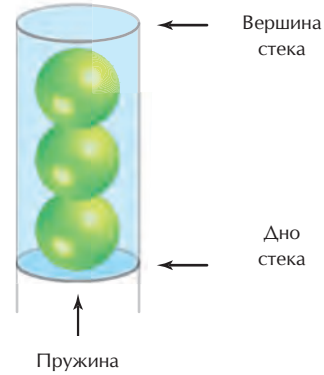
Ключевая идея реализации очереди состоит в том, что массив как бы замыкается в кольцо. За последним элементом массива следует его первый элемент (напомним, что последний элемент имеет индекс $N-1$, а первый — индекс 0). При сдвиге индекса конца очереди вправо в случае, когда он указывает на последний элемент массива, он переходит на первый элемент массива.



СТЕК

Стек — самая популярная и, пожалуй, самая важная структура данных в программировании. Стек представляет собой запоминающее устройство, из которого элементы извлекаются в порядке, обратном их добавлению. Это как бы неправильная очередь, в которой первым обслуживают того, кто встал в неё последним. В литературе по программированию общепринятыми являются аббревиатуры, обозначающие порядок работы очереди и стека. Порядок работы очереди — FIFO, что означает «первым пришёл — первым уйдёшь» (*англ.* First In First Out). Порядок работы стека — LIFO, «первым пришёл — последним уйдёшь» (*англ.* Last In First Out).

Примером стека может служить стопка бумаг на столе или стопка тарелок. Само слово «стек» так и переводится с английского — «стопка». Тарелки берутся в порядке, обратном их добавлению. Доступна только верхняя тарелка, т. е. тарелка «наверху» стека.



Стек легко представить в виде расположенной вертикально трубки с пружинным дном. Верхний конец трубки открыт, и в него можно добавлять элементы. Каждый новый элемент проталкивает на одну позицию вниз элементы, помещённые в стек ранее. При извлечении элемента из стека все остальные как бы выталкиваются вверх. (Общепринятые английские термины здесь очень выразительны: операция добавления элемента в стек обозначается словом push — «затолкнуть», «запихнуть», а операция удаления словом pop — «выстреливать».) Пример стека — магазин карабина.

ИСПОЛЬЗОВАНИЕ СТЕКА В ПРОГРАММИРОВАНИИ

Стек используется чрезвычайно часто, причём в самых разных ситуациях. Объединяет их общая цель: требование запомнить текущий момент в работе, которая ещё не выполнена до конца, при необходимости переключения на другую задачу. Почему именно стек используется для «замораживания» прерванного задания? Предположим, компьютер выполняет задачу *A*. В процессе её выполнения возникает необходимость сделать задачу *B*. Текущий момент решения задачи *A* запоминается, и компьютер переходит к выполнению задачи *B*. Но и здесь компьютер может переключиться на задачу *C*, и нужно будет запомнить состояние задачи *B*, прежде чем перейти к *C*. Позже, когда задача *C* окажется выполненной, сначала будет восстановлен текущий момент задачи *B*, затем, по окончании *B*, — текущий момент задачи *A*. Таким образом, восстановление происходит в последовательности, обратной запоминанию, что соответствует порядку работы стека.

Стек позволяет организовать рекурсию, т. е. обращение подпрограммы к самой себе либо непосредственно, либо через цепочку других вызовов. Пусть, например, подпрограмма *A* выполняет алгоритм, зависящий от входного параметра *X* и, возможно, от состояния общих данных. Для самых простых значений *X* алгоритм реализуется непосредственно. В случае более сложных значений *X* алгоритм осуществляется через применение того же алгоритма для более простых значений *X*. При этом подпрограмма *A* обращается сама к себе, передавая в качестве нужного параметра более простое значение *X*. В этом случае подпрограмма *A* переключается на новый набор переменных, не уничтожая предыдущего набора, который сохраняется в стеке. По окончании этих промежуточных действий начальное значение входного параметра *X* восстанавливается из стека и подпрограмма продолжает работу с того места, где она была прервана обращением к самой себе.

При этом не приходится специально сохранять в стеке значения локальных переменных подпрограммы, так как это происходит автоматически: в самом начале работы подпрограмма захватывает место в аппаратном стеке под свои локальные переменные — этот блок памяти называют обычно «блок локальных переменных подпрограммы» (*англ.* frame — «кадр»). В момент окончания работы подпрограмма освобождает память, удаляя из стека блок своих локальных переменных.

Кроме локальных переменных в аппаратном стеке сохраняются адреса возврата при вызовах подпрограмм. Предположим, в какой-то точке программы *A* вызывается подпрограмма *B*. Перед её вызовом адрес инструкции, следующей за инструкцией вызова *B*, сохраняется в стеке. Это так называемый адрес возврата в программу *A*. По окончании работы

Современный стандарт FORTRAN 90 уже вводит стековую память, устраняя недостатки ранних версий языка.

подпрограмма *B* извлекает из стека адрес возврата в программу *A* и передаёт управление по этому адресу. Таким образом, компьютер продолжает выполнение программы *A*, начиная с инструкции, следующей за инструкцией вызова. В большинстве процессоров имеются специальные команды, поддерживающие вызов подпрограммы с предварительным помещением адреса возврата в стек, а также возврат из подпрограммы по адресу, извлекаемому из стека. Обычно команда вызова обозначается словом call, команда возврата — словом return.

В стек помещаются также параметры подпрограммы или функции перед её вызовом. Порядок их помещения в стек зависит от соглашений, принятых в компиляторах языков высокого уровня. Так, в языке C или C++ наверху стека лежит первый аргумент функции, под ним второй и т. д. В языке Pascal всё наоборот — наверху стека лежит последний аргумент функции.

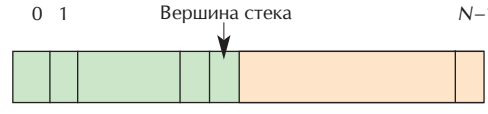
В FORTRAN IV (одном из самых старых и удачных языков программирования) аргументы передавались через специальную область памяти, которая располагалась не в стеке, поскольку в 60—70-х гг. XX в. ещё существовали компьютеры (например, IBM 360 или ЕС ЭВМ) без аппаратной реализации стека. Точно так же адреса возврата сохранялись не в стеке, а в фиксированных для каждой подпрограммы ячейках памяти. Программисты называют такую память статической, оттого что статические переменные занимают всегда одно и то же место в любой момент работы программы. При использовании только этой памяти рекурсия невозможна, поскольку при каждом новом вызове разрушаются предыдущие значения локальных переменных.

РЕАЛИЗАЦИЯ СТЕКА НА БАЗЕ МАССИВА

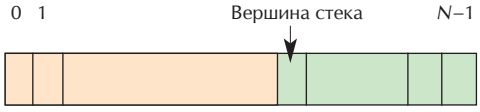
Реализация стека на базе массива является «классикой» программирования. Иногда даже само понятие стека неправильно отождествляется с этой реализацией.

Базой реализации является массив размера *N*, т. е. реализован может быть стек ограниченного размера, максимальная глубина которого не превышает параметра *N*. Индексы ячеек массива изменяются от 0 до $N-1$. Элементы стека хранятся в массиве следующим образом: элемент на дне стека располагается в начале массива, т. е. в ячейке с индексом 0. Элемент, расположенный над самым нижним элементом стека, хранится в ячейке с индексом 1 и т. д. Вершина стека расположена где-то в середине массива. Индекс элемента на вершине стека обозначается специальной переменной, которую обычно называют указателем стека (*англ.* Stack Pointer, или *SP*).

Когда стек пуст, его указатель имеет значение -1. При добавлении элемента в стек указатель сначала увеличивается на единицу, затем в ячейку массива с индексом, содержащемся в указателе стека, записывается добавляемый элемент. При извлечении элемента из стека содержимое ячейки массива с индексом, обозначенным в указателе стека, запоминается во временной переменной в качестве результата операции, затем указатель стека уменьшается на единицу. В описанной реализации стек растёт в сторону увеличения индексов ячеек массива.



Часто используется и другой вариант реализации стека на базе вектора, когда дно стека помещается в последнюю ячейку массива, т. е. в ячейку с индексом $N-1$. Элементы стека занимают непрерывный отрезок массива, начиная с ячейки, индекс которой хранится в *SP*, и заканчивая последней ячейкой массива. В этом варианте стек растёт в сторону уменьшения индексов. Если стек пуст, то его указатель имеет значение N.

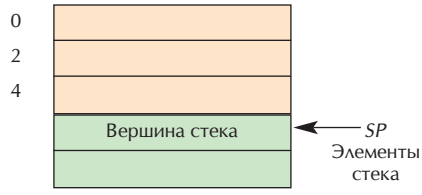


АППАРАТНАЯ РЕАЛИЗАЦИЯ СТЕКА

Строго говоря, в компьютере нет никакого аппаратного стека. Есть только оперативная память, которую можно рассматривать либо как массив байтов, либо как массив машинных слов. Аппаратный стек реализуется на базе оперативной памяти.

Машинное слово вмещает в себя адрес ячейки памяти, например, в 32-разрядной архитектуре адреса байтов могут изменяться от 0 до $2^{32}-1$. Адреса машинных слов кратны четырём и принимают значения $0, 4, 8, \ldots, 2^{32}-4$.

Кроме оперативной памяти процессор имеет свою внутреннюю, очень быструю память, которую называют *регистрами* процессора. Среди них есть регистры общего назначения, хранящие целые числа или адреса ячеек памяти, и так называемые плавающие регистры, хранящие вещественные числа (на программистском сленге вещественные числа называются плавающими). В дешёвых процессорах плавающие регистры могут отсутствовать, но регистры общего назначения имеются обязательно. Два из них играют особую роль — это счётчик команд *PC* (*англ.* Program Counter, иногда употребляется также обозначение *IP*, Instruction Pointer) и указатель стека *SP* (Stack Pointer). Регистр *PC* содержит адрес машинной команды (инструкции), которая будет выполняться на следующем шаге. Регистр *SP* содержит текущий адрес вершины стека. Аппаратная реализация стека совпадает с вариантом реализации стека на базе массива, в котором стек растёт в сторону уменьшения индексов ячеек массива.



АППАРАТНЫЙ СТЕК И ЛОКАЛЬНЫЕ ПЕРЕМЕННЫЕ ПОДПРОГРАММЫ

Аппаратный стек — это просто область оперативной памяти плюс регистр *SP*, который хранит адрес вершины стека. В аппаратном стеке можно размещать обычные переменные программы. Размещение локальных переменных в стеке обладает рядом серьёзнейших преимуществ по сравнению со статическим размещением переменных в фиксированных ячейках оперативной памяти — например, это позволяет организовывать рекурсию. В современных архитектурах весьма важное значение имеет поддержка параллельных процессов, работающих над общими статическими переменными. Это так называемые легковесные процессы, или нити (Thread), работающие параллельно в рамках одной программы. Разные нити работают параллельно над общими статическими данными, совершая таким образом некоторую совместную работу. При этом одна и та же подпрограмма может вызываться из разных нитей. В отличие от статических переменных, которые являются общими для всех нитей, локальные переменные подпрограммы должны располагаться в стеке, выделенном

На использовании нитей основана работа графических приложений в системе Microsoft Windows.

для каждой нити. Тогда наборы локальных данных одной и той же подпрограммы, вызываемой из разных нитей, будут различны, поскольку они располагаются в разных стеках. Если бы это было не так, то вызов подпрограммы, уже работающей в рамках одной нити, из другой нити разрушил бы набор локальных переменных этой подпрограммы.

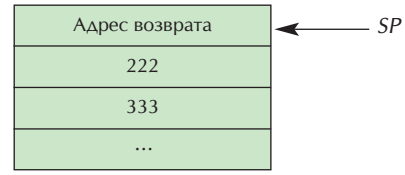
Как же размещаются локальные переменные подпрограммы в стеке? В языке С подпрограммы называются функциями. Функция может иметь аргументы и локальные переменные, т. е. переменные, существующие только в процессе выполнения функции. Предположим, функция зависит от двух входных аргументов *x* и *y* целого типа и использует три локальные переменные *a*, *b* и *c* также целого типа. Функция возвращает целое значение.

```
int f(int x, int y) {
int a, b, c;
...
}
```

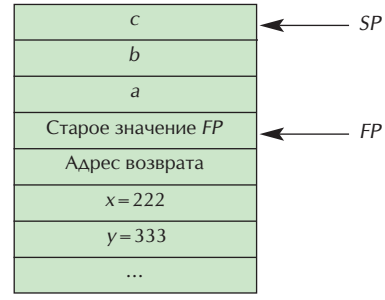
Пусть, в определённом месте программы вызывается функция *f* с аргументами *x* = 222, *y* = 333.

```
z = f(222, 333);
```

Вызывающая программа помещает фактические значения аргументов *x* и *y* функции *f* в стек, при этом на вершине стека лежит первый аргумент функции, под ним — второй аргумент. При выполнении инструкции вызова функции в стек помещается также адрес возврата. В начале работы функции *f* стек имеет следующий вид:



Перед выполнением функция *f* должна захватить в стеке область памяти под свои локальные переменные *a*, *b*, *c*. В языке С принято следующее соглашение: адрес блока локальных переменных функции в момент её работы помещается в специальный регистр процессора, который называется *FP* (*англ.* Frame Pointer — «указатель кадра»). В первую очередь функция *f* сохраняет в стеке предыдущее значение регистра *FP*. Затем значение указателя стека копируется в регистр *FP*. Потом функция *f* захватывает в стеке область памяти размером в три машинных слова под свои локальные переменные *a*, *b*, *c*. Для этого функция *f* просто уменьшает значение регистра *SP* на 12 (три машинных слова равны 12 байтам). После захвата кадра локальных переменных стек выглядит следующим образом:



Аргументы и локальные переменные функции *f* адресуются относительно регистра *FP*. Так, аргумент *x* имеет адрес *FP*+8, аргумент *y* — адрес *FP*+12. Переменная *a* имеет адрес *FP*–4, переменная *b* — адрес *FP*–8, переменная *c* — адрес *FP*–12.

По окончании работы функция *f* сначала увеличивает указатель стека на 12, удаляя таким образом из него свои локальные переменные *a*, *b*, *c*. За-

тем старое значение *FP* извлекается из стека и помещается в *FP* (т. е. регистр *FP* восстанавливает своё значение, существовавшее до вызова функции `f`). Затем происходит возврат в вызывающую программу: адрес возврата снимается со стека, и управление передаётся по адресу возврата. Результат функции `f` передаётся через нулевой регистр. Вызывающая программа сама удаляет из стека фактические значения аргументов *x* и *y*, помещённые в стек перед вызовом функции `f`.

СТЕКОВЫЙ КАЛЬКУЛЯТОР И ОБРАТНАЯ ПОЛЬСКАЯ ЗАПИСЬ ФОРМУЛЫ

В 1920 г. польский математик Ян Лукасевич предложил способ записи арифметических формул без использования скобок. В привычной нам записи знак операции записывается между аргументами: например, сумма чисел 2 и 3 записывается как 2+3. Ян Лукасевич предложил две другие формы записи: префиксную форму, в которой знак операции записывается перед аргументами, и постфиксную форму, в которой знак операции записывается после аргументов. В честь родины математика эти формы записи называют прямой и обратной польской записью. Обычная же форма записи называется инфиксной.

В польской записи скобки не нужны. Например, выражение

$$(2 + 3) \times (15 - 7)$$

записывается в прямой польской записи как

$$\times + 2\ 3 - 15\ 7,$$

а в обратной польской записи как

$$2\ 3 + 15\ 7 - \times.$$

Если прямая польская запись не получила большого распространения, то обратная польская запись оказалась чрезвычайно полезной. Её преимущество можно объяснить тем, что гораздо удобнее выполнять действие, если объекты, над которыми оно должно быть совершено, уже даны. Предположим, что кого-то просят налить стакан чаю. В обратной польской записи сначала дают стакан, затем чайник и только потом уже просят налить чай. В инфиксной форме прежде дают стакан, затем просят налить чай, кто-то должен сходить на кухню за чайником и только потом налить чай. В прямой польской записи сначала просят налить чай, кто-то должен раздобыть стакан, затем чайник и только потом выполнить требуемое действие.

Ян Лукасевич.

Обратная польская запись формулы позволяет проводить вычисления любой сложности, используя стек как запоминающее устройство для хранения промежуточных результатов. Такой стековый калькулятор был впервые выпущен фирмой Hewlett Packard (США). Обычные модели калькуляторов не позволяют вычислять по сложным формулам без использования бумаги и ручки для записи промежуточных результатов. В некоторых моделях предусмотрены скобки с одним или двумя уровнями вложенности, но более сложные вычисления провести на них невозможно. Кроме того, в обычных калькуляторах трудно понять, как результат и аргументы перемещаются в процессе ввода и вычисления между регистрами калькулятора. Напротив, устройство стекового калькулятора вполне понятно и легко запоминается. Калькулятор имеет память в виде стека. При вводе числа оно просто добавляется в стек. При нажатии на клавишу операции, например на клавишу «+», аргументы операции сначала извлекаются из

Калькулятор фирмы Hewlett Packard.

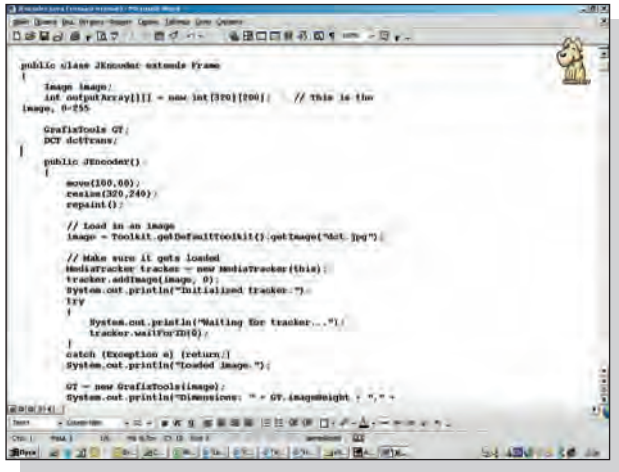
стека, затем над ними выполняется действие, наконец, результат операции помещается обратно в стек. Вершина стека всегда содержит результат последней операции и высвечивается на дисплее калькулятора.

Для вычисления надо сначала преобразовать выражение в обратную польскую запись (при некотором навыке это легко сделать в уме). Затем запись просматривается последовательно слева направо. Если видим число, то вводим его в калькулятор, добавляя его таким образом в стек. Если видим знак операции, то нажимаем соответствующую клавишу калькулятора, выполняя должную операцию с числами наверху стека.

Обратная польская запись формул оказалась исключительно удобной для работы с компьютером. Для вычислений используется стек, что позволяет работать с выражениями любой степени сложности. Реализация стекового вычислителя не представляет никакого труда. Кроме того, имеется простой алгоритм преобразования выражения из обычной записи в обратную польскую запись. Всё это привело к тому, что многие компиляторы языков высокого уровня используют обратную польскую запись в качестве внутренней формы представления программы.

Например, язык программирования Java является интерпретируемым, а не компилируемым языком. Это значит, что компилятор Java преобразует исходную Java-программу не в машинные коды, а в некоторый промежуточный язык, предназначенный для выполнения (интерпретации) на специальной Java-машине. В случае Java этот промежуточный язык называют байт-кодом. Байт-код представляет собой, условно говоря, обратную польскую запись Java-программы, а Java-машина — стековый вычислитель.

Фрагмент программы на языке Java.



```
public class JEncoder extends Frame
{
    Image image;
    int outputArray[][][] = new int[320][200][];    // this is the
image, 0-255

    GrafixTools GT;
    DCT dctTrans;

    public JEncoder()
    {
        move(100,80);
        resize(320,240);
        repaint();

        // Load in an image
        image = Toolkit.getDefaultToolkit().getImage("dct.jpg");

        // Make sure it gets loaded
        MediaTracker tracker = new MediaTracker(this);
        tracker.addImage(image, 0);
        System.out.println("Initialized tracker.");
        try
        {
            System.out.println("Waiting for tracker...");
            tracker.waitForID(0);
        }
        catch (Exception e) {return;}
        System.out.println("Loaded image.");

        GT = new GrafixTools(image);
        System.out.println("Dimensions: " + GT.imgWidth + "," +
```

ЯЗЫК POSTSCRIPT

Другой яркий пример использования обратной польской записи — графический язык PostScript, предназначенный для высококачественной печати текстов на лазерных принтерах. Он является стандартом типографского качества представления текстов, не зависящим от конкретной модели принтера.

То, что PostScript — это язык программирования, для многих людей, знакомых с типографским делом, но далёких от программирования, звучит непривычно. Общепринятое мнение, что компьютер работает с числами и в основном что-то вычисляет, не вполне верно. Не менее часто компьютерная программа работает с текстами и изображениями. Текст, содержащийся в текстовом файле, можно рассматривать с двух точек зрения. Например, трактовать его просто как текст статьи или книги. Но рассмотрим процесс печати текста на обычном (не графическом) принтере. Принтер соединён

с компьютером кабелем, и компьютер посылает через этот кабель один за другим коды символов, составляющих текст. В этом случае букву «А», имеющуюся в тексте, следует рассматривать как команду, предписывающую принтеру напечатать символ «А» в текущей точке страницы, используя заданный заранее шрифт. Затем координату *X* текущей точки надо увеличить на ширину буквы «А».

С этих позиций весь текст можно трактовать как программу его собственной печати. В случае любого текстового файла эта программа весьма примитивна, в ней, например, нет команд смены шрифтов, изменения текущей позиции, рисования линий и т. д. Понятно, что текст типографского качества не может быть представлен текстовым файлом.

В случае языка PostScript файл, пересылаемый на PostScript-принтер, представляет собой программу печати текста. Язык PostScript весьма богат, и вряд ли найдётся много людей, владеющих им. Чаще всего PostScript-программа создаётся другой программой обработки текста. Например, PostScript-файл создаётся для печати на принтере при помощи языка TeX. (TeX — это язык записи текстов, содержащих математические формулы, созданный замечательным математиком и теоретиком программирования Доналдом Кнутом.) Некоторые текстовые редакторы, например Adobe Acrobat или MS Word, в случае печати на профессиональном принтере также преобразуют текст в PostScript-программу. PostScript-файлы очень удобны для распространения, поскольку записываются в обычном текстовом формате и будут одинаково напечатаны в любой стране, независимо от национальных кодировок, версий операционных систем, наличия нужных шрифтов и т. п.

PostScript-программа представляет собой обратную польскую запись, т. е. любая команда записывается после своих аргументов. При выполнении PostScript-программы используется стек. Рассмотрим для примера простейшую программу, рисующую крест на странице. Символ % используется в языке PostScript в качестве комментария.

```
% Программа Крест
2.83 2.83 scale % установить миллиметровую шкалу
0.2 setlinewidth % установить толщину линии 0,2 мм
100 0 moveto % переместиться в точку (100, 0)
100 350 lineto % провести линию к точке (100, 350)
stroke % и начертить её
0 150 moveto % переместиться в точку (0, 150)
250 150 lineto % провести линию к точке (250, 150)
stroke % и начертить её
showpage % выдать страницу
```

Доналд Кнут.

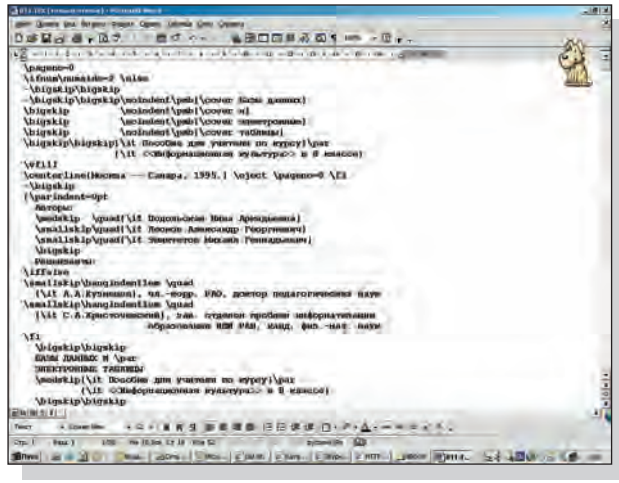
В первой строке программы устанавливается миллиметровая шкала. Дело в том, что по умолчанию для принтера единицей измерения является один пункт, или 1/72 дюйма. Любой программист знает, что 1 дюйм равен 25,4 мм. Надо вычислить соотношение одного миллиметра и одного пункта:

```
1 мм / 1 pt = 1 мм / (1/72) =
= 1 мм / (25,4/72) мм = 2,834645.
```

Если увеличить масштаб в 2,83 раза, происходит переход от пунктов к миллиметрам. Для изменения масштаба в стек сначала помещается число 2,83, соответствующее увеличению масштаба по *X*, затем в стек добавляется второе число 2,83, соответствующее увеличению масштаба по *Y*, и после этого выполняется команда `scale` (изменить масштаб). Из стека при этом вынимаются два числа, и масштаб по *X* и *Y* изменяется.

Разные команды могут иметь разное число аргументов. Например, вторая строка программы устанавливает толщину линии. Команда `set linewidth` имеет один аргумент, который помещается в стек перед её вызовом. При выполнении команды он снимается со стека, и толщина линии устанавливается равной этому снятому числу.

Третья строка перемещает текущую позицию курсора в точку с координатами $X = 100$, $Y = 0$. (В качестве единиц используются миллиметры, о чём мы позаботились в первой стро-



Фрагмент программы на языке TeX.

ке; начало координат находится в левом нижнем углу страницы.) В стек сначала добавляются два числа, соответствующие координатам *X* и *Y*, и затем выполняется команда `moveto`, которая снимает эти числа со стека и перемещает курсор в точку, чьи координаты были получены из стека. Язык PostScript является достаточно мощным языком программирования, в нём предусмотрены функции, циклы, ветвления и т. п. Используемая в нём обратная польская запись очень удобна для выполнения на компьютере. Поэтому интерпретатор языка PostScript легко встраивается в конструкцию принтера (который, конечно же, всегда оснащён более или менее сложным процессором), совсем ненамного увеличивая его стоимость.

Язык Java, байт-код которого представляет собой обратную польскую запись, тоже был первоначально разработан для программирования недорогих бытовых приборов, таких, как телевизор, стиральная машина и т. п. Стековый вычислитель устроен просто и может быть применён там, где быстродействие не играет особой роли, зато важны скорость и дешевизна разработки.

ДИНАМИЧЕСКИЕ СТРУКТУРЫ ДАННЫХ И ССЫЛОЧНЫЕ РЕАЛИЗАЦИИ

Все структуры данных можно условно разделить на два класса — структуры последовательного и произвольного доступа. В первых структурах доступ возможен к одному или нескольким элементам. Обычно это элементы в начале или конце структуры, а также иногда вблизи некоторого «выделенного» места, отмеченного специальным «указателем». Так, в случае очереди доступны лишь элементы в её начале и конце, в случае стека — элемент в его вершине. В структурах произвольного, или «прямого», доступа можно производить действия с любыми элементами структуры, при этом не требуется никаких переборов всех или части элементов.

Большинство структур данных реализуется на базе массива. Все реализации можно также разделить на два класса — непрерывные и ссылочные. В непрерывных реализациях элементы структуры данных располагаются последовательно друг за другом в некотором непрерывном отрезке массива, причём порядок их расположения в массиве соответствует их порядку в реализуемой структуре. В ссылочных реализациях элементы структуры данных хранятся в совершенно произвольном, хаотичном порядке. При этом вместе с каждым элементом хранятся ссылки на один или несколько «соседних» элементов. В качестве ссылок могут выступать либо индексы ячеек массива, либо адреса памяти. Можно представить себе шпионскую сеть, в которой каждый участник знает лишь координаты одного или двух своих коллег. Контрразведчикам, чтобы обезвредить сеть, нужно пройти последовательно по всей цепочке, начиная с выявленного шпиона.

Ссылочные реализации обладают двумя ярко выраженными недостатками:

1) для хранения ссылок требуется дополнительная память;

2) для доступа к некоторому элементу структуры необходимо сначала добраться до него, проходя последовательно по цепочке других элементов.

Казалось бы, зачем нужны такие реализаци? Все недостатки ссылочных реализаций компенсируются одним чрезвычайно важным достоинством: в них можно добавлять и удалять элементы в середине структуры данных, не перемещая остальные элементы.

МАССОВЫЕ ОПЕРАЦИИ

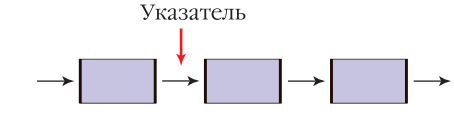
Массовые операции — это операции, затрагивающие значительную часть всех элементов структуры данных. Допустим, нужно добавить или удалить один элемент. Если при этом приходится переписывать значительную часть остальных элементов с одного места на другое, то говорят, что добавление или удаление приводит к массовым операциям. Хорошая реализация структуры данных — та, в которой таких операций либо нет совсем, либо они происходят очень редко. Массовые операции — это проклятие программиста, то, чего он всегда стремится избежать.

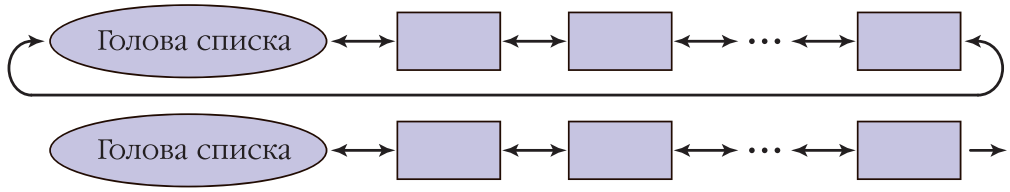
В непрерывных реализациях добавление или удаление элементов в середине структуры неизбежно приводит к массовым операциям. Поэтому структуры, в которых можно удалять или добавлять элементы в середине, обязательно должны быть реализованы ссылочным образом.

СПИСОК

Классический пример структуры данных последовательного доступа, в которой можно удалять и добавлять элементы в середине структуры, — это линейный список. Различают однонаправленный и двунаправленный списки (иногда говорят «односвязный» и «двусвязный»).

Элементы списка как бы выстроены в цепочку друг за другом. У списка есть начало и конец. Имеется также указатель списка, который располагается между элементами. Если мысленно вообразить, что соседние элементы списка связаны между собой верёвкой, то указатель — это красная ленточка, которая вешается на верёвку. В любой момент времени в списке доступны лишь два элемента — элементы до указателя и за указателем.





В однонаправленном списке указатель можно передвигать лишь вперёд, в направлении от начала к концу. Указатель также можно установить в начало списка, перед первым элементом. В отличие от однонаправленного списка двунаправленный абсолютно симметричен, указатель в нём передвигается как вперёд, так и назад и устанавливается как перед первым, так и за последним элементом списка.

В двунаправленном списке можно добавлять и удалять элементы до и за указателем. В однонаправленном списке элементы добавляются также с обеих сторон от указателя, а удаляются элементы только за указателем.

Удобно считать, что перед первым элементом списка располагается специальный «пустой» элемент, который называется «головой списка». Голова присутствует всегда, даже в пустом списке. Благодаря этому всегда можно считать, что перед указателем есть какой-то элемент, что упрощает процедуры добавления и удаления элементов.

В двунаправленном списке считается, что вслед за последним элементом вновь следует голова списка, т. е. список замкнут в кольцо.

Можно было бы точно так же замкнуть и однонаправленный список. Но гораздо чаще считают, что за последним элементом однонаправленного списка ничего не следует. Этот список, таким образом, представляет собой цепочку, заканчивающуюся ссылкой «в никуда».

ССЫЛОЧНАЯ РЕАЛИЗАЦИЯ СПИСКА

В программировании нередко отождествляют понятие списка с его ссылочной реализацией на базе массива или непосредственно на базе оперативной памяти.

Основная идея реализации двунаправленного списка заключается в том, что вместе с каждым элементом хранятся ссылки на следующий и предыдущий элементы. В случае реализации на базе массива ссылки представляют собой индексы его ячеек. Чаще, однако, элементы списка не располагают в каком-либо массиве, а просто размещают каждый по отдельности в *динамической памяти,* выделенной данной задаче. В качестве ссылок в этом случае используют адреса элементов в оперативной памяти.

Голова списка хранит ссылки на первый и последний элементы списка, но не хранит никакого элемента. Следующим за головой списка будет его первый элемент, а предыдущим — последний элемент. Когда список пуст, голова списка зациклена сама на себя.

ПРОБЛЕМЫ НЕПРЕРЫВНОСТИ

Пример неудачного использования непрерывных реализаций — файловые системы в некоторых старых операционных системах, например в ОС ЕС, в системе РАФОС (применявшейся в мини-компьютерах СМ ЭВМ). Если в современных файловых системах файлы фрагментированы (кусочки большого файла могут быть «раскиданы» по всему диску), то раньше это было не так. Файлы обязательно занимали непрерывный участок на диске. При постоянной работе они уничтожались и создавались заново на новом месте — так, всякое редактирование текстового файла приводило к его обновлению. В результате свободное пространство на диске становилось фрагментированным, т. е. состоящим из множества небольших кусков. Очень быстро возникла ситуация, когда большой файл невозможно было записать на диск: хотя свободного места в сумме оставалось много, не имелось достаточно большого свободного фрагмента.

В старых операционных системах приходилось выполнять процедуру сжатия, или, по-научному, дефрагментацию, диска. Файлы переписывались вплотную друг к другу, чтобы собрать свободное пространство диска в один непрерывный фрагмент. Во время выполнения этой длительной процедуры компьютер ничего другого делать не мог, т. е. фактически был недоступен для работы.

Указатель списка реализуется в виде пары ссылок на следующий и предыдущий элементы, он лишь отмечает некоторое место в цепочке элементов.

В случае однонаправленного списка вместе с каждым элементом хранится только ссылка на следующий элемент, таким образом, экономится память. Голова списка хранит ссылку на первый элемент, а последний элемент хранит нулевую ссылку (ссылка «в никуда»), так как в программах нулевой адрес никогда не используется.

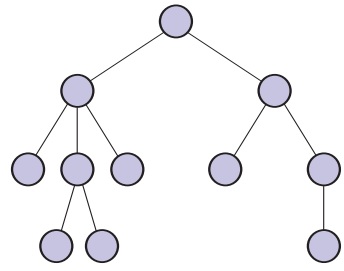
Процедуры добавления и удаления элементов списка не приводят к массовым операциям, потому что для их выполнения нужно всего лишь модифицировать ссылки в двух соседних элементах.

ДЕРЕВЬЯ И ГРАФЫ

Граф — это фигура, которая состоит из вершин и рёбер, соединяющих вершины. Примером графа является схема линий метро. Рёбра могут иметь направления, т. е. изображаться стрелочками; такие графы называют ориентированными.

Дерево — связный граф без циклов. Кроме того, в нём выделена одна вершина, которая называется корнем дерева. Остальные вершины упорядочиваются по длине пути, который ведёт от корня дерева к данной вершине.

Пусть выбрана некоторая вершина *X*. Вершины, соединённые рёбрами с вершиной *X* и расположенные дальше её от корня дерева, называют сыновьями вершины *X*. Они обычно упорядочены слева направо. Вершины, у которых нет сыновей, называются терминальными. Дерево изображают перевёрнутым, корнем вверх.



Деревья в программировании — наиболее часто используемый вид графа. Так, на их построении основаны многие алгоритмы сортировки и поиска. Компиляторы в процессе перевода программы с языка высокого уровня на машинный язык представляют фрагменты программы в виде деревьев, которые называются синтаксическими.

Деревья естественно появляются всюду, где имеются какие-либо иерархические структуры (структуры,

Динамическая память, или «куча», — это область оперативной памяти, в которой можно при необходимости захватывать куски нужного размера, а после использования освобождать, т. е. возвращать обратно в «кучу».

которые могут вкладываться друг в друга). Примером служит оглавление книги. Пусть книга состоит из частей, части — из глав, главы — из параграфов. Сама книга представляется корнем дерева, из которого выходят рёбра к вершинам, соответствующим частям книги. Из каждой вершины-части книги выходят рёбра к вершинам-главам, входящим в эту часть, и т. д.

Файловую систему компьютера также можно представить в виде дерева. Вершинам соответствуют каталоги (их ещё называют директориями или папками) и файлы. Файлы представляются терминальными вершинами дерева. Корню дерева соответствует корневой каталог диска.

Ссылочные реализации как будто специально придуманы для реализации деревьев. Вершина дерева представляется в виде объекта, содержащего ссылки на всех сыновей, а также некоторую дополнительную информацию, зависящую от конкретной задачи. Такой объект занимает область памяти фиксированного размера, которая обычно размещается в динамической памяти. Число сыновей, как правило, зависит от смысла решаемой задачи. Так, очень часто рассматриваются бинарные деревья, в которых число сыновей у произвольной вершины не превышает двух. Если один или несколько сыновей у вершины отсутствуют, то соответствующие ссылки содержат нулевые значения. У терминальных вершин все ссылки нулевые.

При работе с деревьями используются рекурсивные алгоритмы (алгоритмы, которые могут вызывать сами себя). При вызове алгоритма ему передаётся в качестве параметра ссылка на вершину дерева, которая рассматривается как корень поддерева, растущего из этой вершины. Если вершина терминальная, то алгоритм просто применяется к данной вершине. Если же у вершины есть сыновья, то он рекурсивно вызывается также для каждого из них. Порядок обхода поддеревьев зависит от сути алгоритма. Приведём пример простейшего рекурсивного алгоритма, подсчитывающего число терминальных вершин дерева. Его можно использовать для подсчёта числа файлов, содержащихся в заданном каталоге и во всех его подкаталогах.

```
алг цел число терминальных вершин
          (арг вершина V)
дано V – ссылка на корень дерева
надо Подсчитать число терминальных
       вершин дерева
нач
| если у вершины V нет сыновей
| | то знач := 1
| | иначе
| | знач := 0
| | нц для каждого сына X вершины V
| | | знач := знач + число
| | | терминальных вершин(X)
| | кц
| всё
кон
```

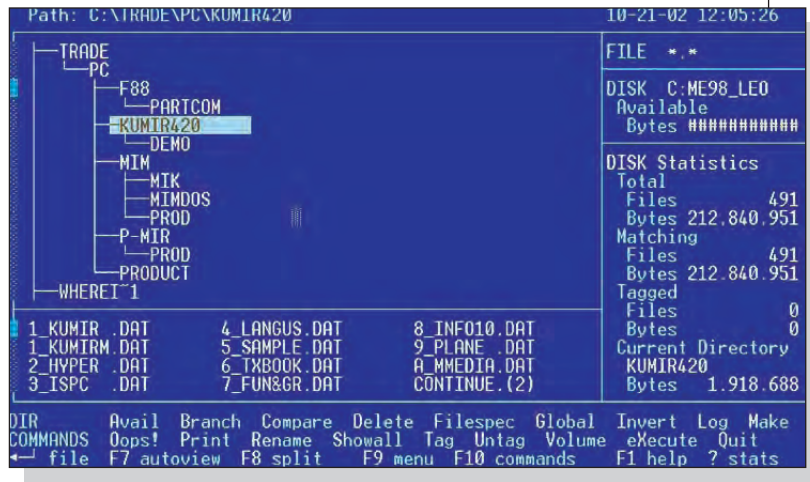
МНОЖЕСТВО

Множество — это структура данных, содержащая конечный набор элементов некоторого типа. Элементы множества никак не упорядочены и не равны между собой. В множество *M* можно добавить, а можно и удалить из него элемент *x*. Если при добавлении он уже содержится там, то ничего не происходит. Множество — потенциально не ограниченная структура, которая может содержать любое конечное число элементов.

В некоторых языках программирования накладывают ограничения на тип элементов и на максимальное количество элементов множества. Так, иногда рассматривают множество эле-

ментов дискретного типа (все существующие значения можно занумеровать целыми числами), где число элементов не должно превышать константы, задаваемой при его создании. Для таких множеств употребляют наименование Bitset (набор битов) или просто Set. Как правило, для их реализации используется так называемая битовая реализация множества на базе массива целых чисел. Каждое целое число рассматривается в двоичном представлении как набор битов, содержащий 32 элемента. Биты внутри одного числа нумеруются справа налево (от младших разрядов к старшим); нумерация битов продолжается от одного числа к другому, когда мы перебираем элементы массива. На пример, массив из десяти целых чисел содержит 320 бит, номера которых изменяются от 0 до 319. Множество в данной реализации может содержать любой набор целых чисел в диапазоне от 0 до 319. Число *N* содержится в множестве тогда и только тогда, когда бит с номером *N* равен единице (программисты говорят «бит установлен»). Соответственно если число *N* не содержится в множестве, то бит с номером *N* равен нулю (программисты говорят «бит очищен»).

Хотя в языке программирования Pascal слово Set (в переводе «множество») закреплено за ограниченным множеством элементов дискретного типа, такими множествами не исчерпываются потребности программирования. Например, для множества точек на плоскости битовая реализация не подходит.



В программировании очень часто рассматривают структуру чуть более сложную, чем просто множество, — «нагруженное множество». Пусть каждый элемент множества содержится в нём вместе с некоторой дополнительной информацией, которую именуют «нагрузкой» элемента. При добавлении элемента в множество нужно обязательно указывать нагрузку, которую он несёт. В разных языках программирования и в различных стандартных библиотеках такие структуры называют отображением (Map) или словарём (Dictionary). Действительно, элементы множества как бы отображаются на нагрузку, которую они несут. В интерпретации Словаря элемент

множества — это иностранное слово, нагрузка элемента — это перевод слова на русский язык.

Все элементы содержатся в нагруженном множестве в одном экземпляре, т. е. разные элементы множества не могут быть равны друг другу. В отличие от самих элементов их нагрузки могут совпадать (так, различные иностранные слова могут иметь одинаковый перевод). Поэтому иногда элементы нагруженного множества называют ключами, их нагрузки — значениями ключей. Каждый ключ уникален. Принято говорить, что ключи «отображаются» на их значения.

В качестве примера нагруженного множества наряду со словарём можно рассмотреть множество банковских счетов. Банковский счёт — это уникальный идентификатор, состоящий из 20 десятичных цифр. Нагрузка счёта — вся информация, которая ему соответствует, включающая имя и адрес владельца счёта, код валюты, сумму остатка и т. п.

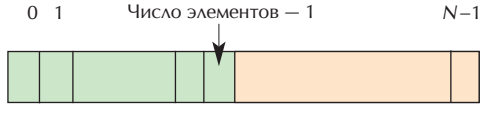
Наиболее часто применяемая операция в нагруженном множестве — определение нагрузки для заданного элемента x (значения ключа x). Реализация данной операции включает поиск элемента x в множестве. Поэтому эффективность любой реализации множества определяется прежде всего быстротой поиска элемента.

РЕАЛИЗАЦИИ МНОЖЕСТВА: ПОСЛЕДОВАТЕЛЬНЫЙ И БИНАРНЫЙ ПОИСК, ХЕШИРОВАНИЕ

Для простоты можно рассмотреть реализацию обычного, ненагруженного, множества. Его элементы хранятся в массиве, начиная с первой ячейки, в произвольном порядке, а в специальной переменной находится текущее число элементов множества, т. е. число используемых в данный момент ячеек массива. При поиске элемента x придётся последовательно перебрать все элементы, пока либо он найдётся, либо выяснится, что его там нет.

При добавлении элемента x к множеству сначала мы должны определить, не содержится ли там уже x. Для этого используется процедура поиска. Если элемент x принадлежит множеству, то выдаётся индекс ячейки массива, в которой он находится, в противном случае он просто дописывается в конец (в ячейку массива с индексом «число элементов») и переменная «число элементов» увеличивается на единицу. Для удаления элемента достаточно последний элемент множества переписать на место удаляемого и уменьшить переменную «число элементов» на единицу.

Такая реализация годится только для небольших множеств.



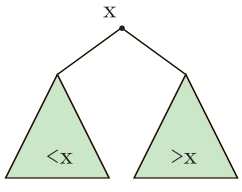
БИНАРНЫЙ ПОИСК

Допустим, можно сравнивать элементы множества друг с другом, определяя, какой из них больше. (Например,

для текстовых строк применяется лексикографическое сравнение: первые буквы сравниваются по алфавиту; если они равны, то сравниваются вторые буквы и т. д.) Тогда удаётся существенно ускорить поиск, применяя алгоритм бинарного поиска. Для этого элементы множества хранят в массиве в возрастающем порядке.

На каждом шаге поиска отрезок массива, в котором может находиться искомый элемент *x*, делится пополам. Рассматривается элемент *y* в середине отрезка. Если *x* меньше *y*, то выбирается левая половина отрезка, если больше, то правая. Поиск продолжается в новом отрезке.

Реализация бинарного поиска с программной точки зрения не представляет никакого труда, поэтому реализация множества с использованием бинарного поиска во всех отношениях лучше «наивной» реализации. Вместе с тем и она имеет недостатки: 1) при добавлении и удалении элементов в середине массива приходится «сдвигать» элементы в его конце на новое место; 2) поиск выполняется гарантированно быстро, но всё-таки не мгновенно. От первого из этих недостатков можно избавиться, применяя вместо непрерывной реализации на базе массива ссылочную, при которой элементы множества содержатся в узлах бинарного дерева. Элементы в вершинах упорядочены следующим образом: зафиксируем элемент *x* в произвольной вершине и рассмотрим два поддерева, растущих из этой вершины. Все элементы левого поддерева должны быть меньше, чем *x*, а все элементы правого поддерева — больше *x*. Чтобы поиск выполнялся быстро, дерево должно быть «сбалансированным», т. е. все его ветви должны иметь почти одинаковую длину. Процедуры добавления и удаления элементов должны сохранять свойство сбалансированности.



ХЕШИРОВАНИЕ

Идея хеширования состоит в том, что мы сводим работу с одним большим множеством к работе с массивом небольших подмножеств. Например, записная книжка содержит список фамилий людей с их телефонами (телефоны — это «нагрузка» элементов множества). Страницы записной книжки помечены буквами алфавита. Всё множество фамилий разбито на 28 подмножеств (буквы «ё», «й», «ь», «ы», «ъ» не учитываются), соответствующих буквам русского алфавита. При поиске фамилии сразу открывают записную книжку на нужной странице — и поиск значительно ускоряется.

Разбиение множества на подмножества осуществляется с помощью так называемой хеш-функции.

Хеш-функция — это функция, определённая на элементах множества и принимающая целые неотрицательные значения. Хеш-функция подбирается таким образом, чтобы:

Идея бинарного поиска иллюстрируется следующей шуточной задачей «Как поймать льва в пустыне?». Надо разделить пустыню забором пополам, затем ту половину, в которой находится лев, снова разделить пополам и т. д., пока лев не окажется пойманным.

1) её можно было легко вычислять;
2) она принимала всевозможные различные значения приблизительно с равной вероятностью. Например, хеш-функция для данной фамилии принимает значение номера первой буквы фамилии в русском алфавите (номера, как всегда в программировании, начинаются с нуля).

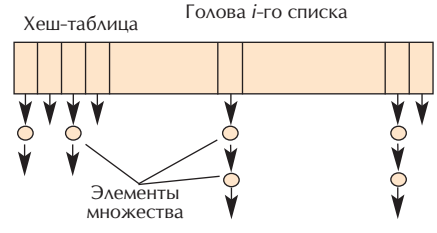
Прежде всего нужно построить массив подмножеств, называемый также хеш-таблицей. При этом важно определиться с числом подмножеств, т. е. размером хеш-таблицы. Для эффективной реализации каждое подмножество должно содержать по возможности не более одного элемента. Следовательно, размер хеш-таблицы (пусть он равен N) должен быть больше, чем среднее число элементов множества. Обычно N выбирают превышающим число элементов множества примерно в три раза. В примере с записной книжкой это означает, что должно быть записано около 10 фамилий. В таком случае большинство страниц либо пустые, либо содержат одну фамилию, хотя не исключены и коллизии, когда на одной странице записано несколько фамилий.

Итак, множество M разбивается на N непересекающихся подмножеств, занумерованных индексами от 0 до $N-1$. Подмножество с индексом i содержит все элементы x из M, значения хеш-функции для которых равняются i.

Теперь нужно подобрать хеш-функцию. Она должна принимать значения 0, 1, 2, ... , $N-1$. Если изначально у нас есть некоторая функция $H(x)$, принимающая произвольные целые значения, то в качестве хеш-функции надо использовать функцию $b(x)$, которая равна остатку от деления $H(x)$ на N.

При поиске элемента x сначала вычисляется значение его хеш-функции. Пусть оно равно b. Затем ищется x в подмножестве с индексом b. Поскольку это подмножество небольшое, то поиск осуществляется быстро.

Различные конкретные схемы хеш-реализации множества по-разному организуют массив подмножеств и борются с коллизиями. Наиболее универсальная схема, использующая ссылочную реализацию, выглядит так: каждое подмножество представляется в виде линейного однонаправленного списка, содержащего элементы подмножества. То есть имеется N списков. Головы всех списков хранятся в хеш-таблице. Элемент хеш-таблицы с индексом i представляет собой голову i-го списка. Голова содержит ссылку на первый элемент списка, первый элемент — на второй и т. д. Последний элемент в цепочке содержит нулевую ссылку. Если список пустой, то элемент хеш-таблицы с индексом i содержит нулевую ссылку.



«ЦИКЛЫ ДЛЯ КАЖДОГО» И ИТЕРАТОРЫ

Часто бывает необходимо перебрать все элементы структуры данных и для каждого элемента выполнить некоторое действие. При записи алгоритмов на неформальном языке используется конструкция «цикл для каждого»; например, в случае множества *M* можно записать следующий фрагмент программы:

```
нц для каждого элемента X
| из множества M
| действие(X)
кц
```

В большинстве структур данных такой цикл можно реализовать, пользуясь другими действиями, существующими для конкретного типа. Например, для списка применим следующий фрагмент программы:

```
установить указатель в начало списка
нц пока указатель не в конце списка
| действие (элемент за указателем)
| передвинуть указатель вперёд
кц
```

Множество отличается от всех других структур данных тем, что «цикл для каждого» невозможно смоделировать, пользуясь другими предписаниями множества. Поэтому выполнение такого цикла должно быть обеспечено реализацией множества. Метод построения «цикла для каждого» сильно зависит от используемого языка программирования и от конкретной библиотеки классов или подпрограмм. Обычно создаётся некоторый объект, позволяющий последовательно перебирать элементы структуры. Внутреннее устройство такого объекта зависит от структуры данных и скрыто от программиста. Данный объект обычно называют итератором или энумератором. С итератором возможны два действия:

- проверить, есть ли ещё элементы структуры данных, которые не были перебраны на предыдущих шагах;
- получить следующий элемент структуры данных.

В случае множества *M* «цикл для каждого» записывается с помощью итератора примерно так:

```
итератор := начать перебор элементов
      множества M
нц пока итератор. есть ещё элементы
| x := итератор. получить следующий
|      элемент множества
| действие (x)
кц
```

* * *

Конечно, рассмотренными примерами не исчерпывается всё многообразие структур данных. Однако большинство реальных примеров либо аналогичны им, либо получаются из комбинаций нескольких элементарных структур — например, реализация множества может использовать деревья. Искусство программиста состоит в том, чтобы выбрать структуру данных, наиболее подходящую для решаемой задачи, в которой отсутствуют массовые операции, память расходуется экономно и поиск выполняется быстро. Но, пожалуй, главный критерий — это простота и изящество программы, которая получается в результате; отсюда вытекает всё остальное. Стремительный прогресс современной вычислительной техники частично нивелирует описанные выше проблемы и позволяет плодить ресурсоёмкие программы и проекты.

СОРТИРОВКА

Герман Холлерит.

Джон Моучли.

Человеку свойственно стремление к порядку, гармонии. Это заметно повсюду, начиная от телефонной книги или каталога библиотеки и заканчивая расстановкой товаров на полке в магазине.

Вероятно, первым примером автоматизации процесса упорядочивания данных является электромеханический табулятор Германа Холлерита, применённый в 1890 г. для обработки результатов переписи населения в США. Впоследствии аналогичные устройства получили самое широкое распространение. Известно, что в 1930 г. в Англии статистические и бухгалтерские фирмы охотно брали в аренду сортировочные машины — всего за 9 фунтов стерлингов в месяц. Были изобретены и более сложные, так называемые подборочные, машины, выполнявшие за один проход слияние двух отсортированных колод карт.

И всё-таки возможность быстрого переупорядочивания действительно больших массивов данных появилась только после изобретения компьютеров, и именно тогда задача сортировки приобрела важное практическое значение. Первым сортировку данных на компьютере, применяя свой собственный метод, выполнил в 1946 г. Джон Моучли, создатель компьютера ENIAC. Необходимость использования компьютеров для сортировки данных не казалась в то время очевидной. Не случайно в своей лекции о сортировке в том же году Моучли сказал: «Требование в одной машине объединить возможности вычислений и сортировки кое-кому может показаться требованием использовать один прибор и как консервный нож, и как авторучку»...

Здесь рассматриваются методы только так называемой *внутренней сортировки* (это название возникло исторически, поскольку она применялась для переупорядочивания массивов, все элементы которых располагались во внутренней, оперативной, памяти компьютера с произвольным доступом. Кроме того, известна *внешняя сортировка,* алгоритмы которой используют вспомогательную память на внешних носителях). Методы сортировки могут быть разбиты на несколько групп, в зависимости от основной идеи переупорядочивания данных, лежащей в их основе:

- сортировка с помощью включения;
- сортировка с помощью выбора;
- сортировка с помощью обмена.

Пусть дана последовательность из n элементов $a_1, a_2, \ldots, a_n$; сортировкой называют операцию перестановки этих элементов в последовательность $a_{i_1}, a_{i_2}, \ldots, a_{i_n}$, на которой выполняются отношения $f(a_{i_k}) \leqslant f(a_{i_{k+1}})$, $k = 1, \ldots, n-1$.

Чаще всего значения функции f не вычисляются, а задаются в явном виде и связаны с определёнными элементами последовательности. Эти значения называют ключами; обычно ключи — целые или действительные числа. Вообще говоря, можно упорядочивать не сами элементы, а только их ключи, ведь перестановка соответствующего элемента вместе с ключом на сложность алгоритма не повлияет.

СОРТИРОВКА С ПОМОЩЬЮ ПРЯМОГО ВКЛЮЧЕНИЯ

Суть метода состоит в том, что исходный массив условно делится на две части — уже отсортированную и ещё не отсортированную (первоначально отсортированная часть состоит лишь из первого элемента массива). Затем на каждом шаге первый элемент неотсортированной части удаляется из неё и помещается в отсортированную часть на своё место — с учётом упорядоченности.

Число сравнений элементов при включении i-го элемента на своё место в отсортированную часть в среднем равно $i/2$. Поскольку эта операция выполняется для каждого из $n-1$ элементов, то алгоритм имеет сложность порядка n^2. Если элементы исходного массива уже упорядочены, то количество сравнений будет минимальным, а если массив упорядо-

чен в обратном порядке, то максимальным.

На первый взгляд кажется, что алгоритм можно улучшить. Действительно, последовательность, в которую надо включить очередной элемент, упорядочена, так что возможен *двоичный поиск* места включения (включаемый элемент сравнивается с элементом в середине уже отсортированной последовательности, и процесс деления пополам идёт до тех пор, пока не будет найдено место включения). Модифицированный таким образом алгоритм называется *сортировкой с двоичным включением*. Включение элемента в отсортированную часть требует сдвига всех элементов, расположенных справа от него. Так что, хотя количество сравнений явно сокращается, число перестановок элементов остаётся тем же и сложность алгоритма по-прежнему имеет порядок n^2.

Этот результат показывает, что далеко не всегда очевидные на первый взгляд усовершенствования алгоритма дают существенный эффект. Вообще метод прямого включения не кажется слишком привлекательным, ведь сдвиг элементов массива в памяти достаточно длительная операция.

СОРТИРОВКА С ПОМОЩЬЮ ПРЯМОГО ВЫБОРА

Общая схема алгоритма следующая:

1) выбирается наименьший из n элементов массива;

2) выбранный элемент меняется местами с первым элементом.

Считая первый элемент отсортированной частью массива, повторяем эти действия для n-1 оставшихся неотсортированных элементов, затем для $n-2$ и т. д. до тех пор, пока в неотсортированной части массива не останется один элемент — самый большой в исходном массиве.

Этот метод в некотором смысле является антиподом предыдущего. При прямом включении на каждом шаге рассматриваются один элемент исходного массива и все элементы отсортированной части. Здесь же просматриваются все неупорядоченные элементы исходного массива и выбирается один, помещаемый в конец отсортированной части.

Число сравнений элементов не зависит от их исходного расположения и пропорционально n^2, значит, алгоритм также имеет квадратичную сложность. А вот среднее

О ВАЖНОСТИ ПОРЯДКА

Идея удобного расположения объектов весьма древняя. Возможно, самый ранний пример её воплощения — вавилонская таблица (она датируется приблизительно 200 г. до н. э.), которая содержит более 100 пар значений шестидесятеричных чисел и их обратных величин, записанных для упрощения поиска в порядке возрастания.

Сравнение чисел издавна кажется делом естественным. Но в Древней Греции для обозначения чисел использовались буквы (и некоторые дополнительные знаки), так что, вероятно, мысль о внесении определённого порядка в расположение слов также не должна была казаться нелепой.

Археологи много раз находили упорядоченные по алфавиту (и даже по первым двум буквам) списки жителей древнегреческих поселений. Упорядочивали слова в словарях многие авторы античности и раннего Средневековья, например знаменитый римский врач и естествоиспытатель Клавдий Гален (около 130 — около 200) в «Словаре Гиппократа».

А то, что сегодня называют лексикографическим порядком, было впервые описано в 1286 г. неким Джованни из Генуи. В предисловии к своему словарю он приводит примеры, когда порядок слов определяется не только второй, но и последующими буквами, вплоть до шестой. Однако спустя ещё 300 лет этот метод требовал дополнительных пояснений. Так, в словаре английского языка «Table Alphabetical», изданном в Лондоне в 1604 г., объяснялось, что слова, начинающиеся с *a*, следует искать в начале книги, а начинающиеся с *v* — в её конце...

Клавдий Гален.

Разумеется, расстановка слов в необходимом порядке выполнялась составителями словарей вручную и требовала достаточной усидчивости и немалого внимания (упомянутый словарь английского языка содержал немало ошибок).

СОРТИРОВКА С ПОМОЩЬЮ ПРЯМОГО ОБМЕНА

Хотя в обоих описанных алгоритмах элементы меняются местами, всё-таки перестановки играют в них вспомогательную роль. Данный метод, напротив, основан именно на них. Алгоритм заключается в проходе по неупорядоченной части массива справа налево; при этом соседние элементы попарно сравниваются и, если необходимо, меняются местами. В результате каждого прохода минимальный элемент неупорядоченной части массива сдвигается к её левому концу. Можно сказать, что более «лёгкие» элементы всплывают, как пузырьки воздуха в воде, поэтому метод иногда называют *пузырьковой сортировкой*. Проходы повторяются до тех пор, пока не будут упорядочены все элементы. Если во время прохода не произошло ни одной перестановки, выполнение алгоритма можно закончить.

Хотя в наилучшем случае (исходный массив полностью упорядочен) число перестановок элементов равно нулю, в среднем оно всё-таки пропорционально n^2. А вот число сравнений постоянно и равно $(n^2-n)/2$. Алгоритм, следовательно, также имеет квадратичную сложность.

Минимальный элемент всплывает за один проход, а вот «тяжёлые» элементы опускаются при каждом проходе только на одну позицию. Поэтому если наибольший элемент расположен

число перестановок меньше, чем для прямого включения, поэтому сортировка с помощью прямого выбора чаще всего предпочтительнее. Однако если элементы исходного массива достаточно хорошо упорядочены, прямое включение всё же несколько эффективнее (в этом случае при равном количестве перестановок оно требует меньше сравнений).

в исходном массиве слева, то потребуется максимальное количество проходов. Избежать этого можно, чередуя направления просмотра массива (этот метод назвали *шейкерной сортировкой*). К сожалению, сложность алгоритма при этом не уменьшается: число перестановок остаётся прежним.

УЛУЧШЕННЫЕ АЛГОРИТМЫ СОРТИРОВКИ

Итак, все рассмотренные до сих пор методы сортировки (их называют *прямыми*) имеют квадратичную сложность.

Возникает принципиальный вопрос: а можно ли и до какого предела улучшать алгоритм решения некоторой задачи? Интересно, что им задавался ещё Чарлз Бэббидж, который считал, что, получив с помощью аналитической машины какой-либо результат, надо обязательно выяснить, не существует ли метод вычислений, дающий тот же результат за меньшее время.

Легко заметить, что во всех прямых методах сортировки за один элементарный шаг элемент массива перемещается только на одну позицию. Доказано, что в среднем каждый из n элементов массива должен за время сортировки сдвинуться на $n/3$ позиции. Это означает, что для упорядочивания n элементов требуется порядка n^2 элементарных шагов. Чтобы уменьшить сложность алгоритма, на каждом элементарном шаге надо стремиться перемещать элементы на большие расстояния. Именно эта идея лежит в основе построения всех улучшенных алгоритмов сортировки.

СОРТИРОВКА ШЕЛЛА

Она является усовершенствованием сортировки с помощью прямого включения. Её суть заключается в неоднократном разбиении исходного массива на группы и их независимой сортировке.

Например, сначала надо разбить исходный массив на четыре группы элементов, отстоящих друг от друга на расстояние 4, и отсортировать каждую группу (она называется *четверной*) независимо, любым известным методом. Затем сформировать две группы, в которых элементы отстоят друг от друга на две позиции. Их сортировку вновь произвести отдельно внутри каждой из групп *(двойная сортировка)*. На третьем этапе *(одинарная сортировка)* сортируется весь массив. В результате получается упорядоченный массив, и, хотя число этапов велико, на каждом из них либо сортируется небольшое число элементов, либо элементы уже достаточно упорядочены и требуется сравнительно немного перестановок. Выигрыш достигается за счёт того, что меняются местами элементы, расположенные на значительном расстоянии друг от друга.

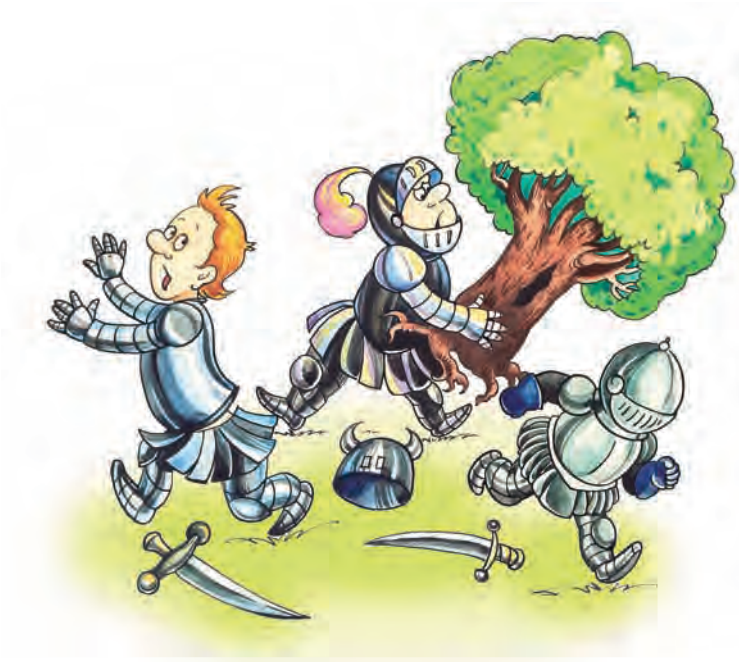
Интересную математическую проблему представляет выбор расстояний между элементами для последовательных сортировок. Доказано, что если массив отсортировать сначала с шагом i, а затем с шагом j, то его упорядоченность сохранится. Расстояния, используемые последовательными сортировками, не обязательно должны быть степенями двойки (лишь на последнем этапе расстояние должно быть равно единице). Более того, желательно, чтобы они не были множителями друг друга — в этом случае при каждом очередном проходе элементы групп будут перемешиваться, что даёт значительный эффект. В одной из работ предлагается, например, такая последовательность расстояний: 1, 4, 13, 40, 121, ... (применяемая, разумеется, в обратном порядке). В случае последовательности расстояний 1, 3, 7, 15, 31, ... сложность алгоритма оценивается как $n^{1,2}$. Но, хотя эта оценка значительно лучше, чем n^2, есть и более предпочтительные алгоритмы сортировки.

СОРТИРОВКА С ПОМОЩЬЮ ДЕРЕВА

Метод является аналогом метода прямого выбора, так как он тоже сводится к нахождению сначала минимума

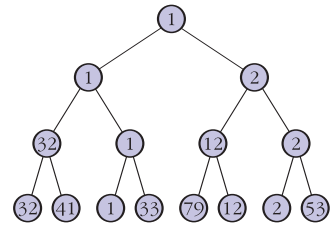
Сложность задачи — это сложность наилучшего алгоритма, известного для её решения.

Дональд Шелл.

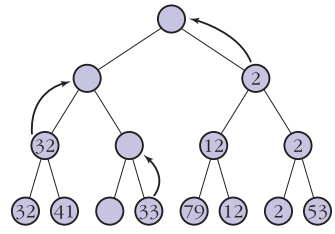


среди n элементов, затем минимума среди $n-1$ элементов и т. д. Всего, таким образом, требуется произвести $(n^2-n)/2$ сравнений. Однако эту оценку можно улучшить. Дело в том, что во время поиска наименьшего элемента получена масса информации, которая затем теряется. Но её можно сохранить, если представить массив с помощью специально построенного *минимизирующего дерева*. Сравнить попарно элементы массива и выбрать каждой паре меньший элемент, затем также сравнить их попарно и т. д.

В результате, выполнив $n-1$ сравнений, найдём наименьший элемент (он находится в *корне* дерева). Это начальный этап сортировки. На рисунке показано минимизирующее дерево для массива {32, 41, 1, 33, 79, 12, 2, 53}.



Теперь нужно вернуться из корня дерева назад вдоль пути, содержащего минимальный элемент, исключая его из дерева и заменяя на пустой элемент. После этого снова произвести подъём, помещая в пустые промежуточные вершины меньшие из сравниваемых элементов. Элемент, передвинувшийся в корень дерева после выполнения этих действий, — вновь наименьший из оставшихся.



Снова исключить его. Спустя n таких шагов дерево станет пустым и процесс сортировки на этом закончится. На каждом из n шагов выбора минимального элемента требуется только $\log_2 n$ сравнений (по одному сравнению, уточняющему значение некоторых вершин на каждом уровне дерева). Поэтому сложность всего процесса составляет $n\log_2 n$. Таким образом, получился алгоритм, даже лучший, чем сортировка Шелла!

БЫСТРАЯ СОРТИРОВКА

Этот метод сортировки используют в основном обмены. Главная его идея заключается в том, чтобы осуществлять перестановки элементов на как

можно большее расстояние. Последовательность действий при этом такая:

1) выбрать наугад (например, посередине массива) элемент x;

2) просмотреть массив слева от x, пока не обнаружен элемент a, который больше его; если такого элемента нет, то $a=x$ (т. е. за искомый a принимается x);

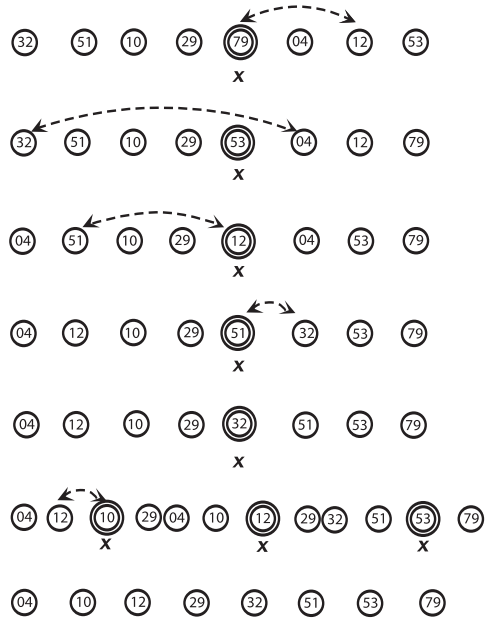
3) просмотреть массив справа от x, пока не обнаружен элемент b, который меньше его; если такого элемента нет, то $b=x$ (т. е. за искомый b принимается x);

4) поменять местами элементы a и b, если $a>b$;

5) продолжить процесс просмотра и обмена до тех пор, пока исходный массив не окажется разбитым на две части. Левая содержит элементы, не превосходящие любой из элементов, составляющих правую. Теперь описанную процедуру (шаги 1 — 4) применить для обеих частей массива.

Доказано, что если в качестве границы x выбрана *медиана* сортируемого массива (так называется элемент, не больший половины и не меньший другой половины элементов), то общее число сравнений составляет $n\log_2 n$, а перестановок — $n\log_2 n/6$. Вообще выбор границы представляет интересную математическую задачу. Установлено, что ей может быть любой элемент массива, в том числе первый или последний; есть даже предположение, что её надо выбирать случайным образом. В целом сложность алгоритма составляет $n\log_2 n$.

Для решения одной и той же задачи возможны самые разные подходы. Анализ алгоритмов сортировки вскрывает сущность понятия сложности алгоритмов, а их сравнение даёт богатый материал для поиска путей совершенствования. (Возможно получить существенный выигрыш в эффективности, перейти от квадратичной сложности тривиальных алгоритмов к сложности порядка $n\log_2 n$ эффективных алгоритмов.)



Почти все алгоритмы сортировки представляют несомненный практический интерес. Даже, казалось бы, не слишком эффективные прямые методы, и те в ряде случаев могут быть с успехом применены.

ПОСТРОЕНИЕ ИНФОРМАЦИОННЫХ МОДЕЛЕЙ

Информация, обрабатываемая на компьютере, должна быть представлена в виде значений величин используемого языка программирования, массивов, таблиц, переменных. Набор величин, содержащий всю необходимую информацию об исследуемых объектах и процессах, называют *информационной моделью*.

Информационная модель хранит не всё о моделируемых явлениях, а только ту часть, которая нужна для рассматриваемых задач, остальное при моделировании отбрасывается. Если круг решаемых задач расширяется, то приходится расширять и модель, включать в неё больше информации.

Информационную модель можно строить разными способами, даже когда круг решаемых задач фиксирован.

Билетная касса в кинотеатре.

Пусть надо организовать предварительную продажу билетов на один сеанс в кинотеатр, где в зале 40 рядов по 50 мест. Для этого необходимо знать, на какие места билеты уже проданы. Раньше перед кассиром лежал план с изображением всех мест в кинозале, где крестиками отмечали места, билеты на которые уже проданы. Таблица заменила план в информационной модели:

```
целтаб a[1:40, 1:50],
```

`a[i, j]=1`, если место продано, и `0` — если нет.

Перед началом продажи, пока ни один билет не продан, таблицу необходимо заполнить нулями:

```
нц для i от 1 до 40
| нц для j от 1 до 50
| | a[i, j]:=0
| кц
кц
```

При продаже билета на место 1 в ряду 20 нужно выполнить команду:

```
a[20, 1]:=1
```

Зал кинотеатра.

С помощью этой таблицы можно получить информацию о ходе продажи и подсчитать количество билетов, оставшихся в кассе (цел остаток):

```
остаток:=0
нц для i от 1 до 40
| нц для j от 1 до 50
| | если a[i, j]=0
| | | то остаток:=остаток+1
| | всё
| кц
кц
```

Для того чтобы определить общую стоимость непроданных билетов, надо знать цену каждого билета, однако в кинотеатре цена билетов разная, в зависимости от ряда. Следовательно, в модель логично включить информацию о ценах на билеты, представив её в виде линейной таблицы или вспомогательного алгоритма

```
алг цел цена (цел i)
```

Алгоритм возвращает стоимость в рублях без копеек (поэтому цел, а не вещ) билета в ряде `i`, например: `цена(1)` вернёт 100, `цена(20)` вернёт 250, так как 20-й ряд дороже первого.

Перед началом продажи этот вспомогательный алгоритм уже должен быть определён, т. е. цена билетов должна быть задана. Чтобы изменить стоимость билетов, достаточно изменить вспомогательный алгоритм `цена` или создать новый, например для выходных и праздничных дней либо, наоборот, для будней.

Теперь в модели достаточно информации и для подсчёта выручки в любой момент:

```
s:=0
нц для i от 1 до 40
| нц для j от 1 до 50
| | если a[i, j]=1
| | | то  s:=s + цена(i)
| | всё
| кц
кц
```

В модели кинозала можно было бы пронумеровать места подряд во всех рядах 40×50=2000 и вместо прямо-

угольной таблицы ввести линейную таблицу `a[1:2000]`. Но это неудобно, так как в реальном кинозале никакой сквозной нумерации мест нет, места задаются номером ряда и номером кресла в ряду. К тому же при постоянном преобразовании информации из одной формы в другую избежать ошибок практически невозможно.

Алгоритмы можно написать поиному уже после того, как модель выбрана. Если в таблице `a[1:40, 1:50]` проданным местам соответствуют единицы, а непроданным — нули, то количество проданных мест подсчитывают просто как сумму всех элементов таблицы, так же как и выручку:

```
n := 0; s := 0
нц для i от 1 до 40
| нц для j от 1 до 50
| | n := n+ a[i, j]
| | s := s+ a[i, j]*цена(i)
| кц
кц
```

Предложенная модель не учитывает возможности бронирования мест, продажи абонементов, скидок пенсионерам и т. п. В ней также отсутствует информация о количестве входов и выходов из кинозала, наличии проходов в зале, количестве буфетов и пр. Для подсчёта выручки или числа проданных билетов данная информация, конечно, не нужна, но если требуется найти два места рядом, это важно.

Значит, придётся расширить модель. Пусть в кинозале есть проход от первого до последнего ряда, тогда информацию можно задать заранее созданной линейной таблицей `p[1:49]` (49=50 – 1), так как количество проходов не превышает 49, если кресел

ИНФОРМАЦИОННАЯ МОДЕЛЬ ТРАНСПОРТНОЙ СЕТИ

В некой стране 100 городов, между которыми проложены авиатрассы. Из одного города можно перелететь (при необходимости — с пересадками) в любой другой. В каждом городе только один аэропорт.

Требуется построить информационную модель, которая позволит ответить на вопросы, как перелететь из одного города в другой с минимальным числом пересадок и какая при этом будет длина пути.

```
цел n; n := 100                    | количество городов

литтаб название[1:n]               | название города
вещтаб расстояние[1:n, 1:n]        | расстояние между
                                   | городами i и j

| расстояние[i, i]=0 для всех i и расстояние[i, j]=0,
| если нет прямой дороги
```

Используя эту модель, можно, например, проанализировать, как добраться из города `i1` в город `i2`, совершив не более одной пересадки:

```
вывод "из города", название[i1], "в город", название[i2]
если расстояние[i1, i2]>0
| то вывод "рейс есть, расстояние =", расстояние[i1, i2]
| иначе p := 0; найден := "нет"
| нц пока p<n и найден = "нет"
| | p := p+1
| | если расстояние[i1, p]>0 и расстояние[p, i2]>0
| | | то найден := "да"
| | всё
| кц
| если найден = "да"
| | то вывод "есть маршрут с пересадкой в", название [p]
| | вывод "расстояние =", расстояние[i1, p]+
| |       + расстояние[p, i2]
| всё
всё
```

В разработке компьютерной системы обычно принимают участие заказчик и исполнитель. Задача последнего — выполнить требования заказчика. Однако на практике заказчик не всегда в состоянии сразу сформулировать все требования к компьютерной системе. Часто он добавляет условия в процессе разработки и даже в процессе эксплуатации системы. Поэтому опытный исполнитель заранее предугадывает возможные запросы заказчика и старается построить информационную модель так, чтобы в процессе развития системы модель только расширялась, а её готовые части и работающие с ними алгоритмы менялись как можно реже.

в ряду 50). Если $p[i]=1$, то в ряду имеется проход между i-м и $(i+1)$-м местами. Как найти два места рядом, решает следующий алгоритм.

```
нашли := "нет"
i := 1; j := 1
нц пока нашли="нет" и i<=40
| если a[i, j]=0 и a[i, j+1]=0 и
| |                    p[j]=0
| | то нашли := "да"
| | иначе j := j + 1
| | если j=50
| | | то i := i + 1; j := 1
| | всё
| всё
кц
```

Пусть в кинотеатре билеты пенсионерам продаются без очереди и со скидкой 50 %. Очерёдность никакого отношения к компьютерной модели и алгоритмам не имеет, а вот продажа билетов со скидкой 50 % затрагивает и модель, и алгоритмы. При подсчёте выручки придётся учитывать скидку, поэтому в модель должна быть добавлена эта информация. Тогда в таблице $a[1:40, 1:50]$: непроданным местам по-прежнему будут соответствовать нули, а проданному месту — процент с оплаты стоимости билета: $a[i, j]=100$, если билет оплачен на 100 %, $a[i, j]= 50$, если билет оплачен на 50 %.

Алгоритм подсчёта числа проданных мест и выручки теперь выглядит так:

```
n := 0; s := 0
нц для i от 1 до 40
| нц для j от 1 до 50
| | если a[i, j]<>0
| | | то n := n + 1
| | | s := s + a[i, j]/100*цена(i)
| | всё
| кц
кц
```

Достоинства данной модели в том, что при необходимости она позволяет варьировать размеры скидки или продавать билеты с учётом разных скидок, например детям до 12 лет со скидкой 70 %. Однако выдавать в ней совсем бесплатно билеты нельзя, потому что при скидке 100 % в таблицу $a[1:40, 1:50]$ нужно занести 0 % от стоимости билета, а такое место нельзя будет отличить от свободного.

ПАРАЛЛЕЛЬНОЕ ПРОГРАММИРОВАНИЕ

В докомпьютерное время в кассах кинотеатров при продаже билетов на один и тот же сеанс обычно кассиры делили между собой билеты поровну, и каждый продавал свою часть. Иногда у одной из касс выстраивалась очередь, а у другой не было никого, потому что либо всё было продано, либо оставались билеты только на плохие места. Подобное происходило и при продаже билетов на поезда дальнего следования. Если железнодорожная касса располагалась не на вокзале, а в другом месте, то кассир заранее брал бланки билетов и дозванивался до вокзала, чтобы узнать о наличии мест в поездах или чтобы сообщить о том, какой билет куплен. Естественно, что работа шла очень медленно, образовывались многочисленные очереди, возникали всевозможные путаницы, например несколько билетов продавалось разным пассажирам на одно и то же место. Работа касс была организована параллельно (как на современных компьютерах можно параллельно выпол-

ИНФОРМАЦИОННЫЕ МОДЕЛИ В ГЕОМЕТРИИ

Основы числового кодирования геометрической информации заложил великий французский учёный XVII в. Рене Декарт (1596–1650): любую точку плоскости можно задать парой чисел (декартова система координат). Числовое кодирование точек позволяет кодировать более сложные геометрические объекты планиметрии. Любая компьютерная система, разрешающая проводить геометрические операции, базируется на моделях подобного типа.

Для всякого геометрического объекта можно придумать много моделей. Для каких-то операций над объектами будут удобны одни модели, а для других — другие. Поэтому для каждого объекта нужно создать несколько информационных моделей и при необходимости переходить от одной к другой. Вот несколько моделей хорошо известных геометрических объектов:

```
Точка
вещ x, y      | координаты точки

Окружность
вещ r         | радиус окружности
вещ a, b      | координаты центра

Прямая
вещ x1, y1    | координаты двух
вещ x2, y2    | точек на прямой

Отрезок
вещ x1, y1    | координаты начала отрезка
вещ x2, y2    | координаты конца отрезка

Треугольник
вещ x1, y1    | координаты
вещ x2, y2    | вершин
вещ x3, y3    | треугольника

n-угольник
вещтаб x[1:n]       | (x[i], y[i]) координаты
вещтаб y[1:n]       | i-й вершины
```

Рене Декарт.

Пользуясь этими моделями, можно переводить геометрические понятия на алгоритмический язык:

```
Длина отрезка: (x1-x2)**2 + (y1-y2)**2

Точка внутри окружности:
(x-a)**2 + (y-b)**2 < r**2

Периметр n-угольника:
p:= 0
нц для i от 1 до n-1
| p:=p+ длина(i, i+1)
кц
p:=p+ длина(n, 1)
```

Один и тот же геометрический объект можно кодировать по-разному, например, отрезок задаётся и так:

```
вещ a, b   | координаты середины отрезка
вещ l      | длина отрезка
вещ ϕ      | угол наклона отрезка к оси абсцисс.
```

нять сразу несколько программ), при этом работа каждого кассира рассматривалась как независимый *процесс,* а список билетов — как *общий ресурс.*

ПРОБЛЕМА ВЗАИМНОГО ИСКЛЮЧЕНИЯ

Требуется решить проблему синхронизации выполнения двух процессов при доступе к общему ресурсу. Для этого нужно выделить в каждом процессе часть, называемую *критической секцией* или *критическим интервалом,* где происходит доступ к общему ресурсу. Например, алгоритм «процесс» выполняется циклически: начинается с обращения к вспомогательному алгоритму «критическая секция», и заканчивается выполнением вспомогательного алгоритма «окончание».

Когда число процессов невелико и они фиксированы заранее, разделить их общий ресурс не представляет особой трудности. Так, легко синхронизировать два процесса: драйвер клавиатуры (программа, осуществляющая ввод символов при нажатии клавиш) и текстовый редактор (программа, запрашивающая очередной символ). Сложнее обеспечить синхронизацию работы процессов, число которых заранее неизвестно.

```
алг процесс
нач
| нц
| | критическая секция
| | окончание
| кц
кон
```

```
алг параллельные процессы
нач цел переключатель
| переключатель := 1
| процесс1 : процесс 2 | параллельное выполнение процессов
кон
```

```
алг процесс 1
нач
| нц
| | нц пока переключатель = 2
| | кц
| | критическая секция1
| | переключатель := 2
| | окончание 1
| кц
кон
```

```
алг процесс 2
нач
| | нц
| | нц пока переключатель = 1
| | кц
| | критическая секция2
| | переключатель := 1
| | окончание 2
| кц
кон
```

```
алг параллельные процессы
нач лог флаг1, флаг2
| флаг1 := "нет"
| флаг2 := "нет"
| процесс1 : процесс 2 | параллельное выполнение процессов
кон
```

```
алг процесс 1
нач
| нц
| | флаг1 := "да"
| | нц пока флаг2
| | кц
| | критическая секция1
| | флаг1 := "нет"
| | окончание 1
| кц
кон
```

```
алг процесс 2
нач
| нц
| | флаг2 := "да"
| | нц пока флаг1
| | кц
| | критическая секция2
| | флаг2 := "нет"
| | окончание 2
| кц
кон
```

Если модифицировать алгоритмы процессов так, чтобы только один процесс мог вызывать свою критическую секцию, то это решит данную проблему. Первое выполнение синхронизации процессов состоит в добавлении целой величины переключатель, которая будет принимать одно из двух значений: «1» или «2» — в зависимости от того, какой из процессов может запускать свою критическую секцию.

Если переключатель равен 1, процесс 2 находится в состоянии ожидания, пока переключатель не станет равен 2. Переключатель примет такое значение, когда первый процесс пройдёт свою критическую секцию. Аналогично поступит и второй процесс, заставив, в свою очередь, ждать первый. Таким образом, в каждый момент времени лишь один процесс может находиться в своём критическом интервале. Но для того чтобы процесс повторно вошёл в свою критическую секцию, ему необходимо дождаться другого процесса, т. е. процессы входят в свои критические секции последовательно: первый, второй, первый, второй... Это излишне жёстко связывает выполнение двух параллельных процессов.

Поэтому к алгоритму синхронизации требуется добавить обязательное условие: задержка любого процесса вне его критического интервала не должна влиять на ход остальных процессов.

Алгоритм синхронизации соответственно изменится: вместо простого переключателя будут введены логические величины флаг1 и флаг2, которые принимают значение да, когда процесс входит в свою критическую секцию, т. е. занимает общий ресурс.

Данный алгоритм более удачен: он не требует соблюдения чёткой последовательности выполнения критических секций. Если один из процессов короче другого, то он может чаще попадать в критическую секцию. Однако и этот алгоритм не лишён недостатков. В случае когда оба процесса готовы войти в свои критические секции, т. е. флаг1 и флаг2 приняли значение да, следующие за командой присвоения циклы ожидания будут выполняться бесконечно,

```
алг параллельные процессы
нач лог флаг1, флаг2, цел переключатель
| флаг1 := "нет"
| флаг2 := "нет"
| процесс1 : процесс 2 | параллельное выполнение процессов
кон
```

```
алг процесс 2
нач
| нц
| | флаг1 := "да"
| | переключатель := 2
| | нц пока флаг2 и
| | переключатель = 2
| | кц
| | критическая секция1
| | флаг1 := "нет"
| | окончание 1
| кц
кон
```

```
алг процесс 2
нач
| нц
| | флаг2 := "да"
| | переключатель := 1
| | нц пока флаг1 и
| | переключатель = 1
| | кц
| | критическая секция2
| | флаг2 := "нет"
| | окончание 2
| кц
кон
```

и ни один из процессов не сможет войти в свою критическую секцию. Это напоминает любезность Чичикова и Манилова в «Мёртвых душах» Н. В. Гоголя, когда они, стоя перед открытой дверью, пропускают друг друга вперёд: «...Только после вас».

И в том и другом случае возникает необходимость выполнения ещё одного условия: решение, какой процесс должен войти в свой критический интервал, необходимо принять за конечное время.

В таком алгоритме как бы объединяются первый и второй алгоритмы. Для синхронизации требуются и величина `переключатель`, и величины `флаг1` и `флаг2`. `Переключатель` по-прежнему равен номеру процесса, который выполняется в критической секции.

Можно заметить, что величина `переключатель` вообще не требует инициализации (присваивания начального значения), так как устанавливается в значении «1» или «2» до цикла ожидания. В отличие от переключателя, флаги сбрасываются (значение `нет`). Их нельзя не инициализировать, так как какой-нибудь из процессов может «вырваться» вперёд, и другой не успеет присвоить свой флаг. В начале процессов флаг поднимается (значение `да`), процесс готов войти в критическую секцию. Переключатель присваивает номер другого процесса, как бы «пропуская» соседа вперёд. Цикл ожидания прервётся, если другой процесс покинет свою критическую секцию (сбросив свой флаг). Переключатель нужен лишь для того, чтобы не возникала ситуация, подобная случаю с Чичиковым и Маниловым.

Проблема взаимного исключения возникла в начале 60-х гг. XX в. Одним из первых создателей алгоритма считается Доналд Кнут. Известный голландский программист Эдсгер Дейкстра в 1965 г. представил наиболее полное и оригинальное решение проблемы синхронизации параллельных процессов. Однако эти алгоритмы не работали, если один из процессов ломался (останавливался) при исполнении своего «окончания». Непревзойдённый по простоте и красоте алгоритм предложил в 1981 г. американец Гарри Петерсен (версия этого алгоритма и приведена в качестве решения проблемы критической секции выше в примере). Алгоритм Петерсена настолько прост, что, как говорят, вообще не нуждается в пояснениях.

СЕМАФОРНАЯ ТЕХНИКА

Эдсгер Дейкстра не только предложил алгоритм, но и придумал оригинальное средство синхронизации параллельных процессов. Основная сложность при синхронизации состоит в том, что проверка значений

Эдсгер Дейкстра.

СЕМАФОРЫ

Семафоры — одна из простейших техник синхронизации параллельных процессов. Не менее красивые и эффективные алгоритмы получаются при использовании так называемых *условных критических интервалов*. (Общие ресурсы представлены переменными специального типа, или *общими*, которые процесс может использовать только внутри специальных операторов — в условных критических интервалах.) Подобные операторы были внедрены в специальную версию языка Pascal. Развитие техники синхронизации параллельных процессов привело к появлению программ-*мониторов*, вобравших в себя и управляющие синхронизацией переменные, и программы захвата и освобождения ресурсов.

Алгоритм, решающий проблемы взаимного исключения, легко расширяется на *N* параллельных процессов. Для синхронизации по-прежнему потребуются переключатель и по одному флагу на каждый процесс.

величин и изменение этих величин не зависят друг от друга. Пока один процесс проверяет условие, разрешающее или не разрешающее ему войти в критический интервал, он не знает, начал ли другой процесс изменять это значение, т. е. не изменилось ли условие, пока его проверяли.

Дейкстра определил величины специального типа — *семафоры* и две операции, которые можно с ними производить. Семафоры могут принимать целые неотрицательные значения. Существуют и так называемые *бинарные семафоры*, соответственно принимающие только значения «0» и «1».

Операция ***V*** увеличивает значение семафора на единицу. Операция ***P*** уменьшает на единицу, если значение семафора больше нуля, или останавливается и ждёт, пока какой-нибудь процесс не увеличит значение семафора (сем означает тип «семафор»):

```
алг V (сем s)
нач
| s := s+1
кон
```

```
алг P (сем s)
нач
| нц пока s=0
| кц
| s := s-1
кон
```

Существенно, что операции ***P*** и ***V*** считаются неделимыми. Это значит, что с начала выполнения операции ***V*** до её окончания доступ к семафору запрещён. В случае ожидания несколькими процессами в операции ***P*** (когда семафор равен нулю) и изменения его другим процессом (применение к нему операции ***V***) только одна операция ***P*** сможет завершиться. Остальные останутся ждать следующего выполнения операции ***V***. В соответствии со своим названием один семафор организует движение у одного разделяемого ресурса. То есть для решения проблемы синхронизации процессов потребуется всего один семафор независимо от количества процессов.

```
алг параллельные процессы
нач сем семафор
| семафор := 1
| процесс1 : процесс 2 | параллельное выполнение процессов
кон
```

```
алг процесс 1
нач
| нц
| | P(семафор)
| | критическая секция1
| | V(семафор)
| | окончание алгоритма1
| кц
кон
```

```
алг процесс 2
нач
| нц
| | P(семафор)
| | критическая секция2
| | V(семафор)
| | окончание алгоритма2
| кц
кон
```

ЧИТАТЕЛИ — ПИСАТЕЛИ

С помощью семафоров решаются многие проблемы синхронизации. Интересным примером является задача **читатели – писатели**, типичная для реальных операционных систем. Несколько процессов-писателей выполняют алгоритм записи в общий ресурс — память. Из памяти читают другие процессы, которые и называют процессами-читателями. По условию задачи одновременно лишь один писатель может находиться в своём критическом интервале, и тогда из памяти нельзя читать. Напротив, сразу несколько читателей могут обращаться к общей памяти.

Задача имеет две версии. В первой ожидание писателя произвести запись в буфер не мешает очередному читателю присоединиться к тем, кто уже занят чтением. Во второй писатели имеют приоритет по сравнению с читателями при доступе к общей памяти: если какой-то писатель готов сделать

запись, то новые читатели не допускаются. Последняя версия задачи несомненно сложнее первой. Для простоты будет приведено решение первой версии.

Для решения задачи применяются два семафора. Первый семафор — `память` — предназначен для синхронизации доступа писателей к памяти. Семафор не должен разрешать писателю осуществить запись, если хотя бы один читатель производит чтение. Для этого вводится целая величина `число читателей`, показывающая число процессов, которые осуществляют чтение в данный момент. Когда первый процесс хочет начать чтение, то он увеличивает число читателей — значение равно «1». При этом следующее условие становится истинным и выполняется операция *P* с семафором `память`. Если ресурс свободен, то он «захватывается» читателями или ждёт, когда писатель закончит свою критическую секцию — `запись в память`. Аналогичные действия осуществляются после критической секции `чтение из памяти`, когда последний читатель освобождает ресурс и разрешает запись или чтение из него.

Семафор `читатель` служит, чтобы ограничить доступ к величине `число читателей` только одному процессу. Увеличение (или уменьшение) этой величины происходит неразрывно с его проверкой в условии. Подобную технику удобно применять, когда нужно ограничивать доступ процессов к любой величине. Важно обратить внимание, что для решения данной задачи использовались бинарные семафоры. Дейкстра в своих работах показал, что все задачи, решённые посредством семафоров, легко выполнить, применяя лишь бинарные семафоры. Вообще говоря, все семафоры можно представить как целую величину и бинарный семафор, что, собственно, и продемонстрировано с величиной `число читателей` и семафором `читатель`.

```
алг читатели – писатели
нач сем читатель, память, цел число читателей
| число читателей := 0
| память := 1
| читатель := 1
| читатель : ... | параллельное выполнение процессов-читателей
| ... : писатель | параллельное выполнение процессов-писателей
кон
```

```
алг читатель
нач
| P(читатель)
| число читателей := число читателей + 1
| если число читателей=1
| | то P(память)
| все
| V(читатель)
| чтение из памяти
| P(читатель)
| число читателей := число читателей - 1
| если число читателей=0
| | то V(память)
| всё
| V(читатель)
кон
```

```
алг писатель
нач
| P(память)
| запись в память
| V(память)
кон
```

ЗАДАЧА ПРО ОБЕДАЮЩИХ ФИЛОСОФОВ

В повседневной жизни часто встречаются задачи параллельного доступа. Самый простой пример: если в ванной комнате уже кто-то моется, а кому-то ещё вдруг понадобилась ванна, то последний лучше поторопит первого, чем станет мыться вместе с ним. Ещё одна задача про параллельные процессы — хорошо знакомая всем водителям проблема проезда нерегулируемого перекрёстка. По правилам

Во многих языках программирования также существует поддержка параллельного программирования. Так, в SmallTalk есть специальный класс Process, в C++, Java синхронизация и параллелизм осуществляются с помощью специальных библиотек классов.

```
алг обедающие философы
нач семтаб вилка [1:5], цел к
| нц для к от 1 до 5
| | вилка[к]:=1
| кц
| философ(1): ... :философ(5)| параллельный запуск процессов-
|                               философов
кон
```

```
алг философ(цел  n)
 дано n
 надо
нач
| нц
| | P(вилка[n])
| | P(вилка [след(n)])
| | ОБЕД
| | V(вилка [n])
| | V(вилка [след(n)])
| | РАЗМЫШЛЕНИЯ
| кц
кон
```

```
алг цел след (цел n)
 дано n
 надо
нач
| n:=n+1
| если n>5
| | то n:=1
| все
| знач:=n
кон
```

дорожного движения водитель должен уступить дорогу машине справа. Но что делать, если к перекрёстку с разных сторон одновременно подъехали четыре автомобиля?

Классическая задача синхронизации — задача про обедающих философов. В замке жили пять философов. Они всё время проводили в размышлениях о сущности бытия. Редкие встречи проходили у них за обеденным столом. Ели философы только вилками, причём двумя одновременно. В замке имелось всего пять вилок, которые и лежали на столе. Когда кто-то из учёных спускался в трапезную, то садился на свободное место за пятиугольный стол и брал вилку справа и вилку слева, если те были свободны, и ел. Потом клал вилки на место и уходил к себе в комнату. Обед и размышления, обед и размышления — так проводили своё время философы в замке.

Для упрощения записи алгоритма используется таблица из пяти семафоров семтаб, где для каждой вилки имеется свой семафор ресурса. Алгоритм написан для произвольного числа философов. Вспомогательный алгоритм след реализует операцию $(n+1) \bmod 5$, $\bmod 5$ — остаток от деления числа на 5.

Однако теоретически вполне реальна ситуация, когда на обед придут сразу все философы, одновременно возьмут в правую руку по вилке и будут ждать, когда освободится вилка слева. Такое положение называется deadlock (в переводе с английского «смертельное объятие»), оно создаёт взаимную блокировку процессов. При блокировке философы просто умрут от голода. Избежать этой проблемы поможет следующий алгоритм. Теперь вилка — логическая величина, равная да, если её никто не взял. Изменение этой величины «охраняет» обычный семафор (сем). Что делать, когда философ не может есть? Он должен спать, прямо за столом. Семафор сон блокирует соответствующий процесс: если философ собрался поесть, а вилка занята, то в ожидании он засыпает. Его будит другой философ, освободивший нужную вилку. Вообще, когда кто-то пообедал, он на всякий случай начинает будить соседей. При этом опять требуется использовать бинарный семафор, чтобы значение семафора никогда не превышало единицы. Приведённый алгоритм позволяет избе-

жать блокировки процессов, однако возможна ситуация, когда один философ всё время спит, так как ему постоянно недостаёт то левой, то правой вилки. Такую ситуацию называют starvation (в переводе с английского «отталкивание»). Правда, существует решение и этой проблемы, однако оно достаточно сложно для данной энциклопедии.

```
алг обедающие философы
нач семтаб сон [1:5], сем, цел к, лог таб вилка [1:5]
| нц для к от 1 до 5
| | сон [к]:=1
| | вилка [к]:="да"
| кц
| философ(1): ... :философ(5) | параллельный запуск процессов-
|                               философов
кон
```

```
алг философ (цел n)
 дано n
 надо
нач
| нц
| | P(сем)
| | нц пока не (вилка[n] и вилка[след(n)])
| | | V(сем)
| | | P(сон[n])
| | | P(сем)
| | кц
| | вилка[n]:="нет"
| | вилка[след(n)]:="нет"
| | V(сем)
| | ОБЕД
| | P(сем)
| | вилка[n]:="да"                        | n — вилка справа
| | вилка[след(n)]:="да"                  | n+1 — вилка слева
| | V(сон[след(n)])                       | n+1 — сосед слева
| | V(сон[след(след(след(след(n))))]) | n+4 — сосед справа
| | V(сем)
| | РАЗМЫШЛЕНИЯ
| кц
кон
```

ТЕОРИЯ ПРОГРАММИРОВАНИЯ

Логика (от *греч.* «логос», означающего одновременно «слово» и «смысл») — наука о законах, формах и операциях правильного мышления. Её основная задача заключается в нахождении и систематизации правильных способов рассуждения.

МАТЕМАТИЧЕСКАЯ ЛОГИКА

КЛАССИЧЕСКАЯ ЛОГИКА

Наука *логика* известна с глубокой древности, её отцом считается великий древнегреческий философ Аристотель (384—322 до н. э.).

Если идёт снег, то сейчас зима; идёт снег; следовательно, сейчас зима.

Это рассуждение состоит из трёх утверждений. Первые два утверждения называют *посылками*, а третье — *заключением*.

Ещё античные философы заметили, что очень часто рассуждения имеют значительное сходство в своём строении, которое не зависит от содержания входящих в него утверждений.

Вот другой пример.

Если воду нагреть до 100 °С, то она закипает; воду нагрели до 100 °С; следовательно, вода кипит.

Несмотря на различие содержания, форма (схема) у обоих рассуждений одинакова.

Если есть первое, то есть и второе; первое имеет место; следовательно, имеет место и второе.

Разумно оценивать правильность рассуждения теми результатами, к которым оно приводит. Рассуждение с использованием этой схемы останется правильным, о каких бы объектах ни велась речь. Если рассуждение правильное, то из его посылок следует заключение. В таких случаях говорят, что схема рассуждения представляет собой *логический закон*.

Существует множество других схем правильного рассуждения, например:

Есть или первое, или второе; первого нет; следовательно, есть второе.

Важность логики заключается в том, что она даёт возможность вырабатывать новые знания без обращения к опыту или интуиции, с помощью формальных рассуждений.

За два тысячелетия со времён Аристотеля традиционная логика не слишком далеко ушла вперёд. Великий немецкий философ Иммануил Кант (1724 — 1804) даже считал, что эта наука уже полностью завершила своё развитие. Однако постепенно в логике назревала революция. И подготовил её другой немецкий философ, математик, физик, изобретатель, юрист, историк, лингвист — Готфрид Вильгельм Лейбниц (1646—1716).

Ещё в 1672 г., посетив работавшего в Париже голландского изобретателя и физика Христиана Гюйгенса (1629—1695), Лейбниц был обескуражен тем, как много сил и времени расходует тот на осуществление необходимых математических расчётов. Возможно, эти впечатления и стали отправной точкой в размышлениях Лейбница над проблемой вычислений. «Недостойно таких замечательных людей, подобно рабам, терять время на вычислительную работу, которую безусловно можно было бы поручить любому лицу при использовании машины», — писал он позднее. Такую машину Лейбницу удалось создать спустя много лет, в 1694 г.

Лейбниц изучал проблему вычислений и теоретически. В 1703 г. он впервые привлёк внимание учёного мира к двоичной системе счисления, доказывая её преимущества перед десятичной системой, так как в последовательностях нулей и единиц гораздо легче обнаружить закономерности их поведения (он называл их «чудесным порядком»).

Лейбниц первым заметил, что выполнение арифметических действий ведётся с помощью простых правил, для которых не имеет значения смысл чисел, а важно лишь то, как они записаны. Правила эти достаточно очевидны, а их применение даёт однозначный результат, следовательно, эти

Лейбниц полагал, что единица представляет божественное начало, а нуль — небытие. Добавление единиц к нулю, позволяющее выразить любое число, он отождествлял с актом божественного творения всего сущего из ничего.

результаты нельзя поставить под сомнение или оспорить. После этого у Лейбница возникла идея: нельзя ли сделать столь же неоспоримыми и производимые человеком умозаключения, представив их как вычисления? Правда, проблему он исследовал больше как философ, предложив девиз: «Будем вычислять».

Идеи Лейбница о математизации умозаключений, значительно опередив своё время, не оказали заметного влияния на научные воззрения его современников. Потребовалось ещё целых полтора столетия, пока в трудах английского математика Джорджа Буля (1815—1864) не были заложены основы математического подхода к логике. Интересно, что Буль не имел математического образования, но это не помешало ему в 1849 г. стать профессором кафедры математики в Куинс-колледже в Корке (Ирландия). Его труды «Математический анализ логики» (1847 г.), «Логическое исчисление» (1848 г.), «Исследование законов мышления» (1854 г.) вошли в фундамент современной математической логики.

Иммануил Кант.

Готфрид Вильгельм Лейбниц.

АЛГЕБРА ЛОГИКИ

Школьная алгебра изучает *формулы*, которые состоят из букв, знаков арифметических операций и скобок.

Формулы можно преобразовывать. Любые допустимые (эквивалентные) преобразования алгебраических формул задаются законами:

Джордж Буль.

• ассоциативным

$$(a+b)+c=a+(b+c),$$

$$(a\cdot b)\cdot c=a\cdot(b\cdot c);$$

• коммутативным

$$a+b=b+a,$$

$$a\cdot b=b\cdot a;$$

• дистрибутивным

$$a(b+c)=ab+ac,$$

а также получаемыми с их помощью важными следствиями (например, формулами сокращённого умножения), такими, как

$$a^2-b^2=(a+b)(a-b).$$

Кроме того, формулы могут использоваться для вычислений; в этом случае вместо букв подставляют их числовые значения и производят соответствующие арифметические действия. Каждая формула задаёт арифметическую функцию, количество аргументов которой равно количеству различающихся букв в формуле.

Алгебра логики также изучает формулы. Но переменные в них могут принимать только одно из двух возможных значений — ложь и истина. Эти значения называют *логическими*. Они обозначаются также, как 0 и 1, Л и И, *false* и *true*. Вводятся специальные знаки логических операций.

И в этом случае каждая формула задаёт функцию. Эти функции называют *логическими*. Их аргументами являются логические переменные, а сами функции могут принимать только логические значения.

Любую такую функцию можно задать в виде таблицы, в которой указаны все возможные наборы значений

x	φ_1	φ_2	φ_3	φ_4
0	0	0	1	1
1	0	1	0	1

аргументов и соответствующие значения функции.

В случае функции от одного аргумента, $\varphi(x)$, такая таблица будет выглядеть крайне просто.

Первая функция, φ_1, принимает значение 0 при любых значениях аргумента, т. е. это *константа* 0, аналогично функция φ_4 — *константа* 1. Значения этих функций не зависят от x, поэтому говорят, что для них переменная x *несущественна*. Функция φ_3, заменяющая значение своего аргумента на противоположное, называется *отрицанием*. Обозначается отрицание по-разному: $\varphi_3(x)=\neg x$ или, более часто, $\varphi_3(x)=\bar{x}$. И наконец, функция $\varphi_2(x)=x$ называется *повторением*.

Логических функций одной переменной насчитывается всего четыре. В общем же случае число логических функций n переменных будет равно 2^{2^n}. (2^n — это число возможных наборов длины n, состоящих из 0 и 1, т. е. число строк таблицы. Тогда 2^{2^n} равно количеству различных расстановок 0 и 1 в столбце с таким числом строк.) Значит, логических функций двух переменных должно быть $2^{2^n}=16$.

Как и в случае функций одной переменной, функции φ_1 и φ_{16} — это константы, соответственно 0 и 1. У них обе переменные несущественны. Для функций $\varphi_4(x_1, x_2)$ и $\varphi_{13}(x_1, x_2)$ несущественна переменная x_2, поскольку $\varphi_4(x_1, x_2)=x_1$, а $\varphi_{13}(x_1, x_2)=\bar{x}_1$. А для функций $\varphi_6(x_1, x_2)$ и $\varphi_{11}(x_1, x_2)$ несущественна переменная x_1, поскольку $\varphi_6(x_1, x_2)=x_2$, а $\varphi_{11}(x_1, x_2)=\bar{x}_2$.

Функцию $\varphi_2(x_1, x_2)$ называют *конъюнкцией* x_1 и x_2. Её принято обозначать $x_1\&x_2$ или же $x_1\wedge x_2$. Конъюнкция

x_1	x_2	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	φ_8	φ_9	φ_{10}	φ_{11}	φ_{12}	φ_{13}	φ_{14}	φ_{15}	φ_{16}
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

равна 1 только в том случае, когда и x_1, и x_2 равны 1. Поэтому конъюнкцию часто называют ещё функцией И. Называют её также *логическим умножением*, так как значения функции совпадают с результатом умножения аргументов по правилам арифметики.

Функцию $\varphi_8(x_1, x_2)$ называют *дизъюнкцией* x_1 и x_2. Обозначают дизъюнкцию $x_1 \vee x_2$. Её называют также функцией ИЛИ, так как она равна 1, если единице равен или x_1, или x_2.

Ещё одна функция, $\varphi_7(x_1, x_2)$, получила название *сложения по модулю* 2. Для неё используется обозначение $x_1 \oplus x_2$. Функция равна 1, если значения аргументов различаются, и равна 0, если они совпадают. Отсюда происходит другое название этой функции — *неравнозначность*.

Напротив, функция $\varphi_{10}(x_1, x_2)$ в случае равенства аргументов равна 1, а в случае неравенства — равна 0. Поэтому она названа *эквивалентностью* или *равнозначностью*. Обозначается эта функция $x_1 \sim x_2$ или же $x_1 \equiv x_2$.

Имеют названия ещё три функции:

импликация — $\varphi_{14}(x_1, x_2)$, обозначается $x_1 \to x_2$ или $x_1 \supset x_2$; читается «если x_1, то x_2»;

стрелка Пирса (другое название НЕ-ИЛИ) — $\varphi_9(x_1, x_2)$. Обозначается $x_1 \downarrow x_2$.

штрих Шеффера (другое название НЕ-И) — $\varphi_{15}(x_1, x_2)$, обозначается $x_1 \mid x_2$.

Остальные функции используются достаточно редко.

Если формула содержит только введённые знаки операций, то её значение на любом наборе переменных легко вычислить. Например, функцию $\varphi(x_1, x_2, x_3)$, которая задана формулой $x_1 \wedge \bar{x}_2 \vee (x_1 \vee x_3)$, можно задать следующей таблицей:

x_1	x_2	x_3	$\varphi(x_1, x_2, x_3)$
0	0	0	1
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	1
1	1	1	1

Вообще таблицу соответствующей функции можно построить для каждой формулы, подставляя в неё значения переменных на всех наборах. Табличное представление функции (а следовательно, и формулы) единственно. Но одна и та же функция может быть записана с помощью разных формул.

Например, функцию штрих Шеффера можно записать и как $\varphi_{15}(x_1, x_2) = \bar{x}_1 \vee \bar{x}_2$, и как $\varphi_{15}(x_1, x_2) = \overline{x_1 \wedge x_2}$. Такие формулы, представляющие одну и ту же функцию, называются *эквивалентными* или *равносильными*.

Для записи формул нет необходимости использовать все функции. Например, от штриха Шеффера в формулах можно избавиться, выразив его через конъюнкцию, дизъюнкцию и отрицание, поэтому функция $\varphi(x_1, x_2, x_3)$ может быть записана в виде $\overline{\overline{x_1 \wedge x_2}} \vee \overline{x_1 \vee x_3}$. Существуют наборы логических функций, посредством которых можно выразить любые другие логические функции. Такие наборы называют *функционально полными*, или *базисами*.

Так же как и штрих Шеффера, через конъюнкцию, дизъюнкцию и отрицание можно выразить и другие логические функции двух переменных:

$$x_1 \to x_2 = \bar{x}_1 \vee x_2,$$

$$x_1 \sim x_2 = x_1 x_2 \vee \bar{x}_1 \bar{x}_2,$$

$$x_1 \oplus x_2 = x_1 \bar{x}_2 \vee \bar{x}_1 x_2,$$

$$x_1 \downarrow x_2 = \overline{x_1 \vee x_2} = \bar{x}_1 \bar{x}_2.$$

Если кит идёт по улице, то полночь!

Формулы, содержащие только знаки этих трёх операций (а также переменные и скобки), называют *булевыми формулами* (в честь Дж. Буля). Справедлива теорема, гласящая, что любая логическая функция может быть представлена в виде булевой формулы. Поэтому набор из конъюнкции, дизъюнкции и отрицания образует базис. Множество логических функций, на которых определены эти три операции, называется *булевой алгеброй*. Операции булевой алгебры обычно называют *булевыми операциями*.

Для булевых операций характерны общие основные свойства.

- Законы ассоциативности:

$$(x_1 \vee x_2) \vee x_3 = x_1 \vee (x_2 \vee x_3),$$

$$(x_1 \wedge x_2) \wedge x_3 = x_1 \wedge (x_2 \wedge x_3).$$

- Законы коммутативности:

$$x_1 \vee x_2 = x_2 \vee x_1,$$

$$x_1 \wedge x_2 = x_2 \wedge x_1.$$

- Законы дистрибутивности:

$$x_1 \wedge (x_2 \vee x_3) = (x_1 \wedge x_2) \vee (x_1 \wedge x_3),$$

$$x_1 \vee (x_2 \wedge x_3) = (x_1 \vee x_2) \wedge (x_1 \vee x_3).$$

- Правила де Моргана:

$$\overline{x_1 \wedge x_2} = \bar{x}_1 \vee \bar{x}_2,$$

$$\overline{x_1 \vee x_2} = \bar{x}_1 \wedge \bar{x}_2.$$

- Идемпотентность:

$$x \wedge x = x,$$

$$x \vee x = x.$$

(Это значит, что в логических формулах, в отличие от алгебраических, не бывает степеней и коэффициентов.)

- Двойное отрицание:

$$\bar{\bar{x}} = x.$$

- Закон противоречия:

$$x \vee \bar{x} = 0.$$

- Закон исключённого третьего:

$$x \vee \bar{x} = 1.$$

- Свойства констант:

$$x \wedge 1 = x, \quad x \wedge 0 = 0,$$

$$x \vee 1 = 1, \quad x \vee 0 = x,$$

$$\bar{0} = 1, \quad \bar{1} = 0.$$

Справедливость каждого из этих *эквивалентных соотношений* может быть легко проверена, достаточно вычислить их левые и правые части на всех возможных наборах значений входящих в них переменных.

При записи булевых формул знаки операции конъюнкции чаще всего опускают, как и знак умножения в алгебре. Аналогично при последовательном выполнении нескольких конъюнкций или дизъюнкций опускают скобки; опускаются и внешние скобки у конъюнкции:

$$(x_1 \wedge (x_2 \wedge x_3)) \vee ((x_4 \vee x_5) \wedge x_6) =$$

$$= x_1 x_2 x_3 \vee (x_4 \vee x_5) x_6.$$

Основные эквивалентные преобразования бывают двух типов. Пер-

вый — упрощение формул. Оно позволяет получать эквивалентные формулы меньшей длины.

• Поглощение:

$$x \vee xy = x; \quad x(x \vee y) = x.$$

• Склеивание:

$$xy \vee x\bar{y} = x,$$

$$(x \vee y)(x \vee \bar{y}) = x.$$

• Обобщённое склеивание:

$$xz \vee y\bar{z} \vee xy = xz \vee y\bar{z}.$$

Второй тип — приведение к дизъюнктивной нормальной форме (ДНФ).

Конъюнкция переменных или их отрицаний, в которой каждая переменная встречается не более одного раза, называется *элементарной конъюнкцией*. ДНФ — это формула, имеющая вид дизъюнкции элементарных конъюнкций. Из соотношения обобщённого склеивания видно, что ДНФ формулы не обязательно единственна.

Если функция n переменных $\varphi(x_1, x_2, ..., x_n)$ задана таблицей, то её булеву формулу можно построить так. Выписываются конъюнкции всех переменных, на наборе которых функция равна 1. Переменные, равные 0, берутся с отрицанием. Все конъюнкции соединяются знаками дизъюнкции. Получившаяся формула называется *совершенной дизъюнктивной нормальной формой* (СДНФ). СДНФ для каждой функции единственна (что следует из самого метода построения СДНФ), они могут отличаться только порядком следования переменных в конъюнкциях и порядком конъюнкций. Не имеет СДНФ лишь функция константа 0.

Используя таблицу, построим СДНФ функции $\varphi(x_1, x_2, x_3)$:

$$\bar{x}_1\bar{x}_2\bar{x}_3 \vee \bar{x}_1x_2\bar{x}_3 \vee x_1x_2\bar{x} \vee x_1x_2x_3.$$

Получившаяся формула достаточно громоздка, но её можно упростить, используя эквивалентные преобразования. В данном случае дважды применив склеивание (соответственно к первой и второй, а также к третьей и четвертой конъюнкциям):

$$\bar{x}_1\bar{x}_3 \vee x_1x_2.$$

Полученная ДНФ значительно короче.

Наборы конъюнкция и отрицание, дизъюнкция и отрицание также являются базисами. С помощью отрицания конъюнкцию можно выразить через дизъюнкцию и наоборот:

$$x_1 \vee x_2 = \overline{\bar{x}_1 \vee \bar{x}_2} = \bar{\bar{x}}_1\,\bar{\bar{x}}_2,$$

$$x_1x_2 = \overline{\bar{x}_1\bar{x}_2} = \bar{\bar{x}}_1 \vee \bar{\bar{x}}_2.$$

Существуют и другие базисы. Наиболее известен среди них базис Жегалкина, включающий конъюнкцию, сложение по модулю 2 и константу 1. В самом деле,

$$\bar{x} = x \oplus 1,$$

$$x_1 \vee x_2 = x_1x_2 \oplus x_1 \oplus x_2.$$

Базис может состоять даже из одной функции. Например, и штрих Шеффера, и стрелка Пирса также образуют базисы, поскольку любая из функций — отрицание, конъюнкция и дизъюнкция — может быть выражена через них:

$$\bar{x} = x|x = x \downarrow x,$$

$$x_1x_2 = \overline{\bar{x}_1\bar{x}_2} = \overline{x_1|x_2} = (x_1|x_2)|(x_1|x_2),$$

$$x_1 \vee x_2 = \overline{x_1 \downarrow x_2} = (x_1 \downarrow x_2) \downarrow (x_1 \downarrow x_2).$$

БУЛЕВА АЛГЕБРА И КОМПЬЮТЕР

М. А. Гаврилов.

Наличие в программах проверок условия, условных переходов — именно то, что определяет чёткую грань между вычислительными устройствами (вроде арифмометра или карманного калькулятора) и компьютером. Они позволяют программисту предусмотреть все возможные варианты выполнения разрабатываемой им программы.

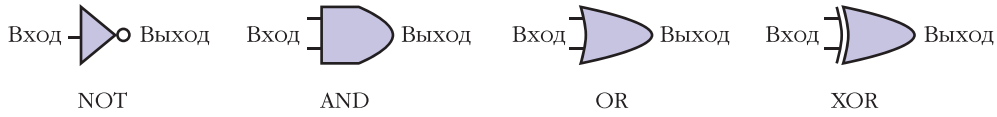
Для установления истинности условия требуется составление логических формул и вычисление логических выражений с использованием правил алгебры логики. Поэтому во многих языках программирования, начиная с самых ранних, таких, как FORTRAN или Algol-60, определены логические значения и операции с ними. Для логических значений в языках программирования часто используются обозначения false и true, позволяющие не путать их с числами 0 и 1 в тексте программы (хотя фактически истинным считается любое ненулевое значение логической переменной).

Вентили есть не только в схемах.

Интересное применение алгебры логики в программировании — это использование так называемых *предикатных вычислений*. Компилятор снабжает некоторые машинные команды предикатами (логическими условиями). Значения предикатов хранятся на специальных регистрах, а в систему команд компьютера вводятся команды над предикатами. Команды под управлением предикатов выполняются независимо от значения последних. Но при этом если значение предиката есть true, то команда выполняется обычным образом, а если значение предиката равно false, то она не изменяет состояние процессора. Использование предикатного механизма позволяет компилятору переупорядочить команды в программе для более эффективного её исполнения.

Ещё одним важнейшим применением алгебры логики в информатике является создание *логических схем*. Над возможностями применения логики в технике учёные и инженеры задумывались уже давно. Например, голландский физик Пауль Эренфест (1880—1933), кстати несколько лет работавший в России, писал ещё в 1910 г.: «...пусть имеется проект схемы проводов автоматической телефонной станции. Надо определить: 1) будет ли она правильно функционировать при любой комбинации, могущей встретиться в ходе деятельности станции; 2) не содержит ли она излишних усложнений. Каждая такая комбинация является посылкой, каждый маленький коммутатор есть логическое “или-или”, воплощённое в эбоните и латуни; всё вместе — система чисто качественных... “посылок”, ничего не оставляющая желать в отношении сложности и запутанности... Правда ли, что, несмотря на существование алгебры логики, своего рода “алгебра распределительных схем” должна считаться утопией?». Созданная позднее М. А. Гавриловым (1903—1979) теория релейно-контактных схем показала, что это вовсе не утопия.

Основой построения сложных логических схем являются *вентили* — так называют простейшие устройства, на входы которых поступают начальные данные, а на выходе полу-



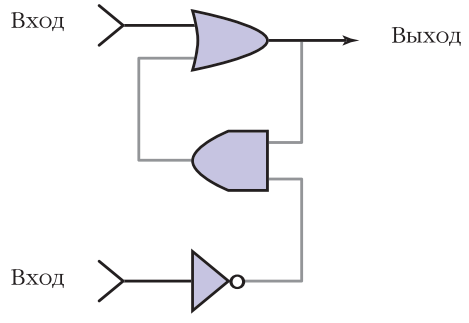
чается результат некоторой булевой операции. На разных этапах развития техники вентили строились с использованием доступной технологии, например зубчатых колёс или электромеханических реле. В современном компьютере вентиль — это небольшая электронная цепь, в которой значения 0 и 1 соответствуют разным уровням электрического напряжения.

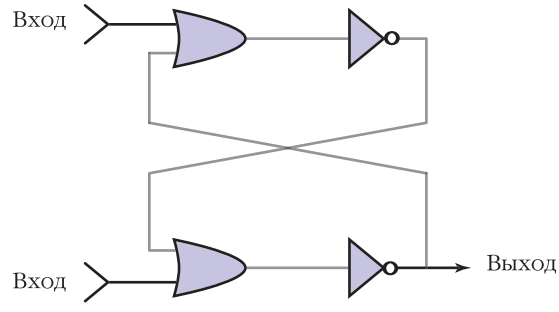
На рисунке показаны используемые в логических схемах изображения вентилей, соответствующих логическим функциям NOT (отрицание), AND (логическое И), OR (логическое ИЛИ) и XOR (неравнозначность, исключающее ИЛИ).

Из вентилей составляют более сложные схемы, в частности *триггеры*. Триггером называется схема, постоянно выдающая значение 0 или 1 до тех пор, пока одиночный импульс от другой схемы не переведёт её в противоположное состояние. Следовательно, триггер позволяет хранить в компьютере один бит информации.

Возможны разные варианты построения триггера (на рисунке представлены два из них).

Несмотря на различную структуру, оба этих триггера имеют абсолютно одинаковые свойства. Отсюда видно,





как велико значение формального представления логической схемы. Оно необходимо для того, чтобы разработчик имел возможность выбрать наиболее подходящий ему вариант построения схемы из вентилей. Не только триггеры, но и другие базовые логические схемы (в частности, *элемент задержки*, имеющий один вход и один выход, на который с определённой задержкой поступает сигнал, поданный на вход) используют для построения всё более и более сложных схем. Процесс разработки общей логической схемы устройства (в том числе и компьютера в целом), таким образом, становится иерархическим, причём на каждом следующем уровне в качестве «кирпичиков» используются логические схемы, созданные на предыдущем этапе.

Алгебра логики дала в руки конструкторам мощное средство разработки, анализа и совершенствования логических схем. В самом деле, гораздо проще, быстрее и дешевле изучать свойства и доказывать правильность работы схемы с помощью выражающей её формулы, чем создавать реальное техническое устройство. Именно в этом состоит смысл любого математического моделирования.

Набор операций, выполняемых логическими элементами, должен быть функционально полным (вспомним понятие базиса). В этом случае любая логическая схема заведомо реализуема. А применение эквивалентных преобразований формул позволяет существенно упрощать их. В итоге удаётся строить логические схемы из минимально возможного количества элементов, что, в свою очередь, обеспечивает большую скорость работы и увеличивает надёжность устройства.

ВОЗМОЖНА И ДРУГАЯ ЛОГИКА

«Человек не знал двух слов — “да” и “нет”. Он отвечал туманно: “Может быть, возможно, мы подумаем”...». Эту запись находим на страницах знаменитых «Записных книжек» замечательного писателя Ильи Ильфа (одного из соавторов романов «Двенадцать стульев» и «Золотой телёнок»).

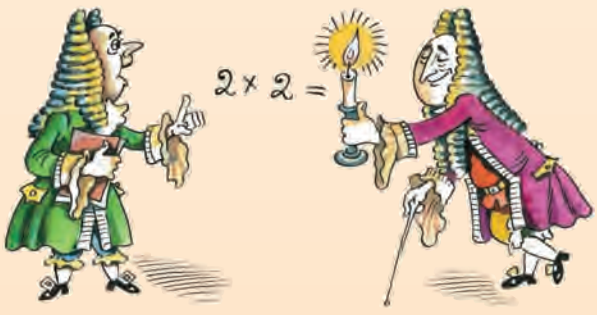
И в самом деле, часто нам явно не хватает двух известных слов, точнее, двух логических значений. Ведь то и дело мы слышим высказывания, про которые нельзя сказать, истинны они или ложны. «Возможно, я получу на экзамене отличную оценку». Или, например, обычной является ситуация, когда мы должны принять решение — делать что-либо или нет, не имея при этом всей необходимой информации либо не зная степени её достоверности.

Учёные давно пытались преодолеть ограничения классической аристотелевой логики. Например, русский логик Н. А. Васильев в 1910 г. разработал оригинальную систему, назвав её «воображаемая логика». Согласно Васильеву, каждое суждение может быть *утвердительным*, *отрицательным* или *акцидентальным*.

Если акцидентальное суждение истинно, то и утвердительное, и отрицательное суждения являются ложными. Тем не менее одно и то же суждение не может быть одновременно и истинным, и ложным. Логика Васильева не имела большой известности и только в последние годы учёные вновь стали обращаться к его идеям.

Зато самое широкое распространение получили так называемые *многозначные логики*. В них значения истинности переменных и функций располагаются в диапазоне от 0 до $k-1$ (тогда 0 можно понимать как абсолютную ложь, а $k-1$ — как абсолютную истину).

Основоположником новой науки стал польский логик и математик Ян Лукасевич (1878—1956), предложивший в 1920 г. трёхзначную логику. В логике Лукасевича значения могли быть *истинными*, *ложными* и *нейтральными*. Спустя год американский учёный Эмиль Пост (1897—1954) создал её обобщённую модель — *k-значную логику*. Ещё позднее, в 1930 г., Ян Лукасевич и Альфред Тарский (1902—1983) разработали *бесконечнозначную логику*.



Для многозначных логик также можно определять алгебры, подобные булевой. Для k-значной логики операции отрицания, конъюнкции и дизъюнкции можно задать следующим образом:

$$\overline{x} = (k-1) - x,$$

$$x_1 \wedge x_2 = \min(x_1, x_2),$$

$$x_1 \vee x_2 = \max(x_1, x_2).$$

Для двузначной логики, т. е. для случая $k=2$, это определение приводит к уже известным булевым операциям.

Применяется в логике и так называемый лингвистический подход, при котором истинность переменных задаётся не числовыми значениями, а упорядоченным набором словесных характеристик. Например, набор значений лингвистической переменной «Рост человека» можно упорядочить (по возрастанию): лилипут, очень маленький, маленький, ниже среднего, средний, выше среднего, высокий, очень высокий, великан.

Лингвистические переменные имеют прямое отношение к так называемой *нечёткой математике*, построенной на базе понятий *нечёткого множества* и *нечёткого вывода*. С её помощью решаются многие важные практические задачи. Например, в области искусственного интеллекта, где разрабатываемые для разных аналитических и диагностических целей технические, медицинские и другие *экспертные системы* в основе механизма принятия решения содержат нечёткий вывод.

Основоположником нечёткой математики является американский учёный Лофти Заде, развивающий свою теорию с 60-х гг. XX в.

Существует ещё и совершенно особая женская логика! Великий русский писатель Иван Сергеевич Тургенев так охарактеризовал её, вложив в уста Пигасова (персонаж романа «Рудин») слова: «...мужчина может, например, сказать, что дважды два не четыре, а пять или три с половиною, а женщина скажет, что дважды два — стеариновая свечка».

Но тут уж пасуют и математика, и информатика...

ТЕОРИЯ АЛГОРИТМОВ

ФОРМАЛИЗАЦИЯ ПОНЯТИЯ АЛГОРИТМА

Теория алгоритмов строит и изучает конкретные модели алгоритмов. Эти модели различаются наборами преобразуемых объектов и элементарных шагов, способами выбора следующего шага и т. д.

Можно выделить три главные линии построения теоретических моделей алгоритмов.

Первая линия связана с тем очевидным фактом, что любые данные могут быть закодированы числами. Каждое преобразование данных сводится к арифметическим вычислениям, а каждый алгоритм находит значение некоторой числовой функции. Элементарными шагами алгоритма являются выполняемые арифметические операции. Последовательность элементарных шагов определяется в таких моделях либо с помощью *суперпозиции* функций, либо с помощью *рекурсии*.

Вторая линия связана с предположением, что любой алгоритм можно представить как последовательное преобразование входных слов, составленных из символов конечного алфавита. Таким образом возникло понятие *нормальных алгоритмов*, предложенных и изученных А. А. Марковым (1903—1979).

Третья линия достаточно близка ко второй. Только в отличие от неё использует не абстрактные подстановки, а определяет какие-либо конкретные механизмы. При этом предполагается: каждый шаг алгоритма таков, что его может выполнить достаточно простое устройство (машина). Желательно, чтобы оно было *универсальным*, т. е. чтобы на нём можно было выполнять любой алгоритм.

Значит, надо представить механизм работы машины в виде стандартной схемы, максимально простой по логической структуре, но настолько точной, чтобы она могла служить предметом математического исследования. Впервые это было сделано американским математиком Эмилем Постом в 1936 г.,

Суперпозиция — подстановка одних функций в другие. Она порождает математические формулы. Например, если в функцию $xy+z$ вместо y подставить функцию $t-u$, а вместо z функцию vw, получим формулу $x(t-u)+vw$.

НОРМАЛЬНЫЕ АЛГОРИТМЫ МАРКОВА

А. А. Марков

Для нормальных алгоритмов задаются конечный алфавит и конечное множество *подстановок* (правил замены одних частей слова на другие). Работа алгоритма состоит из выполнения в определённом порядке двух операторов: *операторов-распознавателей* и *операторов подстановки*. Оператор-распознаватель находит вхождение левой части подстановки в слово, а оператор подстановки заменяет это вхождение на соответствующую правую часть.

Пусть заданы алфавит $A=\{x, y\}$ и множество подстановок $yyx\to y$; $xx\to y$; $yyy\to x!$ (Последняя — заключительная подстановка, после выполнения которой процесс останавливается.) Исходное слово $p = x\,y\,x\,x\,y\,y$.

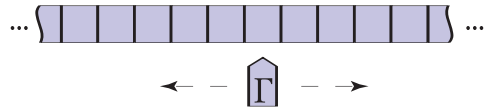
Если с помощью оператора-распознавателя выяснилось, что первая подстановка к этому слову неприменима, надо перейти к анализу второй подстановки. Оператор-распознаватель выделяет первое вхождение её левой части (xx) в p: $p=xy(xx)\to xyy$. Оператор подстановки выполнит подстановку, в результате чего получается новое слово $p_1=xyyxyy$. К нему применима первая подстановка: $p_1=x(yyx)yy\to xyyy=p_2$. К полученному слову применима лишь третья подстановка: $p_2=x(yyy)\to xx=p_3$. Это была заключительная подстановка, поэтому слово p_3 — окончательный результат работы алгоритма.

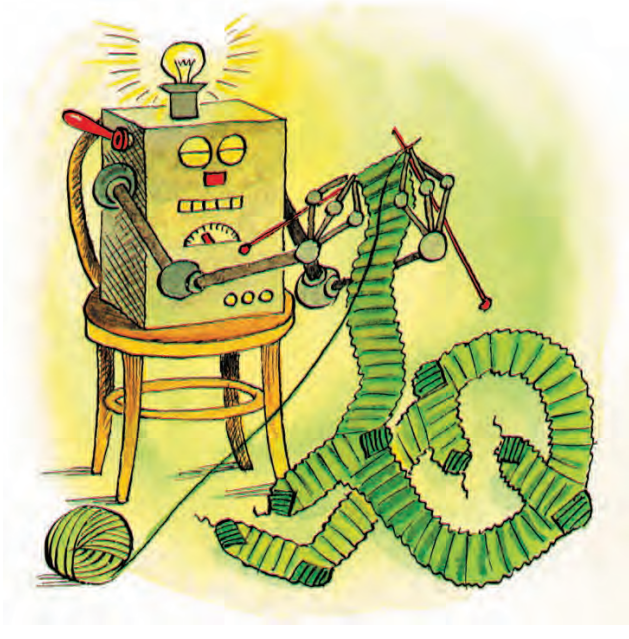
Рекурсией называют вычисление очередного значения функции с использованием ранее полученных её значений. Примером рекурсивного задания функции является определение факториала: $f(0) = 1$; $f(n+1) = f(n)(n+1)$. Здесь значение функции задаётся для начального значения аргумента, а каждое последующее определяется через предыдущее.

ещё до создания современных вычислительных машин, и практически одновременно английским математиком Аланом Тьюрингом.

МАШИНА ПОСТА

Машина Поста состоит из ленты и головки. Лента разделена на клетки и считается *условно бесконечной* (это означает, что при необходимости число клеток может быть увеличено). В каждой клетке может быть записан один *символ* из фиксированного *алфавита*. В любой конкретный момент головка обозревает одну клетку и способна работать только с ней.





Работают машины Поста под управлением *программ*. Программы состоят из *команд*, имеющих по три поля, в которых записываются номер команды, операция и отсылка. Команды нумеруются начиная с 1; выполнение программы начинается с команды номер 1. Отсылка показывает, какую команду нужно выполнять после данной. Для машины Поста определены операции пяти видов:

- движение головки на одну клетку вправо;
- движение головки на одну клетку влево;
- запись символа S в обозреваемую клетку, при этом прежнее содержимое клетки пропадает;
- СТОП — завершение работы машины Поста, в поле отсылки не нуждается;
- команда разветвления, записываемая так: $? \begin{cases} S_1, & M_1 \\ S_2, & M_2 \\ \dots & \\ S_n, & M_n \end{cases}$

Здесь $S_1, S_2, \dots$ — некоторые символы из алфавита машины Поста, M_1, $M_2, \dots$ — отсылки. По этой команде анализируется символ из обозреваемой клетки. Если он совпадает с S_i, следующей будет выполняться команда с номером M_i. При этом головка никуда не перемещается и на ленту ничего не записывается. Если в клетке стоит символ, не упомянутый в команде разветвления, то происходит так называемая безрезультатная остановка машины Поста (иначе говоря, машина ломается).

Другими возможными исходами выполнения программы на машине Поста являются бесконечная работа и результативная остановка, которая вызывается выполнением команды СТОП.

Вот пример. Пусть алфавит машины Поста состоит из символов {-, 0, 1} (символом «-» обозначим пробел), а на ленте записана последовательность символов 1 и 0. Головка «смотрит» на самый левый элемент этой последовательности. Требуется заменить все символы 1 на 0, а 0 на 1. Программа (вместе с комментариями, которые размещаются между % ... %) будет выглядеть так:

1. ?
   - 1 2 % если символ 1, то выполнить команду 2%;
   - 0 3 % если символ 0, то выполнить команду 3%;
   - - 5 % если символ «-», то выполнить команду 5%
2. 0 4 % записать 0 в обозреваемую клетку, выполнить команду 4%;
3. 1 4 % записать 1 в обозреваемую клетку, выполнить команду 4%;
4. → 1 % сдвинуть головку вправо, выполнить команду 1%;
5. СТОП % если встретили символ «-», то остановка машины %.

Исходные данные в примере записываются в виде конечной последовательности непустых символов из входного алфавита машины. Такие последовательности называются *словами* в этом алфавите. Известно *предложение Поста*: «Для всякого алгоритма, исходными данными и результатами которого являются слова, существует эквивалентная ему программа для машины Поста». Поэтому алгоритмы, работающие со словами, часто называют *постовыми*.

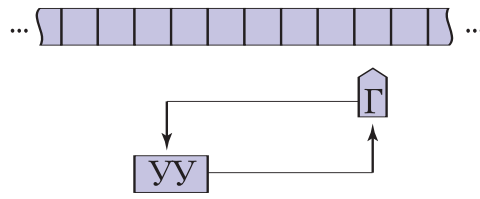
МАШИНА ТЬЮРИНГА

Машина Тьюринга состоит из трёх частей:

- неограниченная в обе стороны лента, разделённая на ячейки;
- управляющее устройство (УУ);
- головка (Г).

С каждой машиной Тьюринга связаны два конечных алфавита. Один из них называют алфавитом внешних символов $S=\{s_0, s_1, s_2, \dots, s_k\}$, а другой — алфавитом внутренних состояний $Q=\{q_0, q_1, \dots, q_n\}$, причём особо выделяют два состояния — начальное q_0 и заключительное q_f. В каждый момент времени каждая ячейка ленты содержит только одну букву из алфавита S. Отсутствие записи в ячейке интерпретируется как запись буквы s_0. Устройство управления может находиться в одном из состояний $q_i \in Q$. Головка «смотрит» на одну из ячеек.

В каждый момент машина Тьюринга описывается своей *конфигурацией*, включающей сведения о состоянии управляющего устройства, запись на ленте и указание на обозреваемую ячейку. Работа машины Тьюринга состоит из тактов; во время $(i+1)$-го такта конфигурация машины преобразуется из текущей, K_i, в новую, K_{i+1}. Преобразование зависит лишь от состояния УУ и содержимого обозреваемой ячейки. Оно заключается в следующих действиях:



- изменить состояние q_i на состояние q_j;
- заменить букву s_l, которая записана в обозреваемой ячейке, буквой s_p;
- изменить положение головки — сдвинуть её на одну ячейку влево или вправо (или же не сдвигать вовсе).

Такое преобразование называют *командой* машины Тьюринга. Команды записываются в виде $q_i s_l \rightarrow q_j s_p R$.

R — это одна из букв Л, Н или П, означающих: Л — обозревать соседнюю слева ячейку; Н — продолжать обозревать ту же ячейку; П — обозревать соседнюю справа ячейку. Совокупность всех команд, выполняемых машиной Тьюринга, называется *программой*.

Для каждой буквы $s_l \in S$ и состояния $q_i \in Q$ программа содержит всего одну команду с левой частью $q_i s_l$. Поэтому если зафиксировать конфигурацию K_1, с которой начинается работа, то работа машины Тьюринга определяется однозначно. В первом такте конфигурация K_1 (в ней применима единственная команда) преобразуется в конфигурацию K_2. Аналогично во втором такте K_2 единственным способом преобразуется в конфигурацию K_3 и т. д.

Работа машины Тьюринга может продолжаться неограниченно долго, причём начиная с любой конфигурации, поэтому необходимо определить правила окончания процесса. Можно

Алан Тьюринг.

АЛГОРИТМИЧЕСКИ НЕРАЗРЕШИМЫЕ ПРОБЛЕМЫ

Задачу называют алгоритмически неразрешимой, если не существует машины Тьюринга (или нормального алгоритма Маркова, или машины Поста), которая её решает.

Вот несколько примеров из области теории алгоритмов.

Пусть по описанию алгоритма А и аргументу *x* необходимо выяснить, остановится ли алгоритм А, если *x* подаётся на его вход. Эта задача (её называют проблемой остановки) является алгоритмически неразрешимой. В частности, доказана теорема о том, что нельзя построить машину Тьюринга T_0, решающую проблему остановки для произвольной машины Тьюринга *T*. Важным практическим следствием этого факта является невозможность создать универсальный (пригодный для любой программы) алгоритм отладки программ. В самом деле, одним из наиболее распространённых последствий ошибок программиста является зацикливание програмы, при котором она не доходит до конца выполнения и может работать «бесконечно долго». Алгоритмическая неразрешимость проблемы остановки означает, что невозможно определить, результативно ли будет работать произвольный алгоритм.

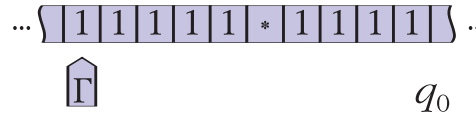
В теории алгоритмов известно утверждение — теорема Райса: «Никакое нетривиальное свойство вычислимых функций не является алгоритмически разрешимым». Смысл данной теоремы таков: нельзя построить алгоритм (т. е. машину Тьюринга, машину Поста, нормальный алгоритм Маркова и др.), определяющий, обладает или нет функция, вычисляемая произвольным алгоритмом А, тем или иным свойством. Согласно теореме Райса, нельзя построить алгоритм, устанавливающий, является ли функция постоянной, периодической, ограниченной, есть ли среди её значений данное число, равна ли она другой функции и т. д. Можно даже подумать, что вообще всё на свете неразрешимо!

Но мы не случайно говорили о необходимости различать понятия функции и реализующего её алгоритма, ведь в теореме Райса присутствуют и те и другие. Теорему Райса тогда можно понимать как невозможность из описания алгоритма узнать что-либо о свойствах вычисляемой им функции. Установление неразрешимости той или иной задачи отнюдь не является неуспехом. Напротив, это новый научный результат, такой же как и в любой естественно-научной дисциплине. Даже если задача алгоритмически неразрешима, её частные случаи вполне могут иметь решение. Так, признано, что не существует общего алгоритма отладки произвольной программы. Но ведь конкретные программы каждый программист отлаживал неоднократно!

считать, что её работа прекращается на некотором такте, если в нём — а соответственно и во всех последующих — конфигурация не изменяется. Такую конфигурацию называют *заключительной*.

Программу можно задавать в виде таблицы, отражающей связь нового содержимого обозреваемой ячейки, дальнейшего движения головки и нового состояния.

Пусть алфавит внешних символов машины Тьюринга состоит из двух символов: {*, 1}; кроме того, пусть символ 0 соответствует пустой ячейке. На ленту подаются два слова (каждое из которых содержит только единицы), разделённых символом *. Надо убрать этот разделитель и слить слова в одно. Следовательно, начальную конфигурацию машины Тьюринга можно изобразить так:

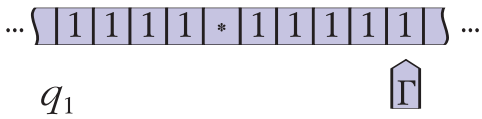


а конечное состояние ленты должно быть таким:

Соответствующая программа имеет вид:

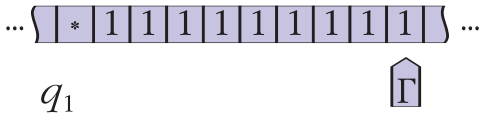
	q_0	q_1	q_2
1	0 П q_2	Л	П
0	П	П q_0	1 q_1
*	0!	Л	П

В момент начала работы головка обозревает крайнюю левую единицу, а машина находится в состоянии q_0. В обозреваемую ячейку записывается 0, головка сдвигается вправо на одну ячейку, и происходит переход к состоянию q_2. В этом состоянии головка последовательно сдвигается вправо до конца слова, пропуская символы * и 1. Обнаружив символ 0, головка в эту ячейку записывает вместо него 1, а машина переходит в состояние q_1:



q_1

Машина остаётся в этом состоянии, пока головка не вернётся снова к первой единице (она сдвигается влево, а достигнув «пустой» ячейки, сдвинется на одну ячейку вправо). После этого конфигурация установится в состояние q_0. Вместо первой единицы будет записан 0, а машина переходит в состояние q_2. Теперь весь предыдущий цикл повторится четыре раза, пока, наконец, машина не примет вид



q_1

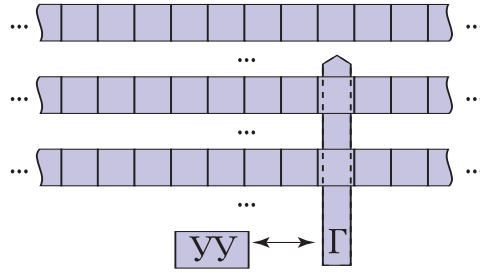
Символы 1 и «*» в состоянии q_1 вызывают сдвиг головки влево, так что, лишь дойдя до первого символа 0, машина выполнит команду $Пq_0$, и её состояние изменится на q_0. Однако в этом состоянии вместо «*» в ячейку записывается символ 0 и машина останавливается.

Легко заметить некоторое сходство между машиной Тьюринга и вычислительными машинами, ведь и те и другие работают над исходными данными задачи по некоторой программе её решения. Можно даже сказать, что машина Тьюринга является в каком-то смысле идеальной моделью компьютера. Однако она намного проще даже самых первых, весьма примитивных с современной точки зрения вычислительных машин. Разумеется, машина Тьюринга не предназначена для широкого практического применения. Не слишком удобна она и при чисто теоретических построениях из-за её громоздкости.

В связи с этим получили распространение многочисленные усовершенствованные образцы, одним из которых является *многоленточная машина Тьюринга.* Она имеет несколько входных лент и только одну выходную, на которую записываются символы результата. Содержимое ячеек выходной ленты нельзя читать и корректировать. В такой машине

удобно предполагать, что головка неподвижна, а ленты движутся, причём, возможно, в разных направлениях.

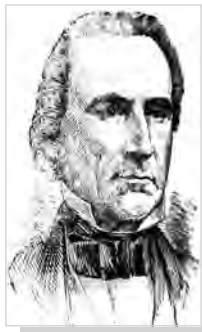
Такт работы этой машины включает следующие действия: одновременное считывание текущего символа со всех лент (кроме выходной), смена состояния в зависимости от считанного набора символов, запись новых символов, сдвиг лент.



ВЫЧИСЛИМЫЕ И НЕВЫЧИСЛИМЫЕ ФУНКЦИИ

Возникает естественный вопрос: а можно ли в принципе говорить об общих свойствах алгоритма, не теряется ли эта общность при такой конкретизации?

Оказалось, и это строго доказано в теории алгоритмов, что любой алгоритм, описанный в одной модели, может быть описан и в другой. Такая взаимная сводимость алгоритмических моделей позволила создать систему понятий, не зависящую от выбора



Алонзо Черч.

Эмиль Пост.

конкретной формализации. И её основой является понятие *вычислимой функции*, т. е. функции, для вычисления которой существует алгоритм.

Какие же функции могут быть вычислены с помощью алгоритмов? Первым в истории математики ответ на этот вопрос дал Алонзо Черч (1903—1995). В 1930 г. он сформулировал такое утверждение: «Всякая функция, вычислимая некоторым алгоритмом, является рекурсивной». Оно известно сейчас как *тезис Черча*.

Спустя несколько лет А. Тьюринг предложил другое утверждение, названное позднее *тезисом Тьюринга*: «Всякий алгоритм может быть реализован машиной Тьюринга». Тезис Тьюринга не является теоремой, его невозможно доказать, но он проверен многолетней практикой, и до сих пор не приведено ни одного примера алгоритма, который нельзя было бы реализовать с помощью машины Тьюринга.

Сравнение двух тезисов позволяет записать их в обобщённом виде: «Функция вычислима на машине Тьюринга тогда и только тогда, когда она является рекурсивной» (*тезис Тьюринга — Черча*). Таким образом, тезис Тьюринга — Черча говорит об эквивалентности двух введённых моделей алгоритма. Но при этом он является утверждением, которое можно строго доказать. Доказываются и другие теоремы о сводимости одного определения алгоритма к другому. Например, доказаны сводимость машин Тьюринга к нормальным алгоритмам Маркова, предложение Поста об эквивалентности алгоритмов над словами и машин Поста. То есть действительно можно говорить об алгоритмах в самом общем смысле, не конкретизируя используемую модель.

Фактически алгоритм — это способ задания функции. Но функция может задаваться разными способами, например таблицей или формулой. Однако существуют зависимости, которые невозможно представить ни в том ни в другом виде. Известная формула,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!}$$

лишь позволяет посчитать это значение с устраивающей нас точностью, но не является точным представлением функции e^x.

Этот пример показывает, что, вероятно, существуют и *невычислимые функции*, что имеются и такие задачи, в которых связь между входными и выходными величинами столь сложна, что нельзя задать строго определённый пошаговый процесс преобразования исходных данных в результат. Иными словами, нельзя построить алгоритм их решения. Такие задачи называются *алгоритмически неразрешимыми*.

СЛОЖНОСТЬ АЛГОРИТМОВ

ПОНЯТИЕ СЛОЖНОСТИ

По мере накопления опыта работы в той или иной области знаний человек начинает интуитивно чувствовать, что одна задача более трудна, чем другая. Разумеется, такого субъективного, не выраженного строгой количественной мерой понятия, как «легче» или «труднее», недостаточно. Как же можно (и можно ли?) определить действительную сложность задачи? Для этого существуют разные подходы.

Очень часто задачу называют сложной из-за того, что она обладает разветвлённой логической структурой, содержит большое количество проверок условий, переходов и т. д. Программа, реализующая алгоритм её решения, также кажется нам сложной, поскольку нельзя понять, что же она делает. А вот для компьютера выполнение этой программы никакой трудности не представляет! Потому что он просто выполняет одну за другой команды, которые там написаны. Для него абсолютно неважно, какие они именно — сложения или проверки условия и перехода, циклы или команды выбора.

Более сложной может показаться программа с текстом, состоящим из большего числа операторов. Но и такой взгляд на сложность оказывается не вполне оправданным. В самом деле, для компьютера две программы — состоящая из одного оператора цикла, 100 раз повторяющего сложение двух чисел, и состоящая из 100 операторов сложения, выписанных подряд, — практически одинаковы.

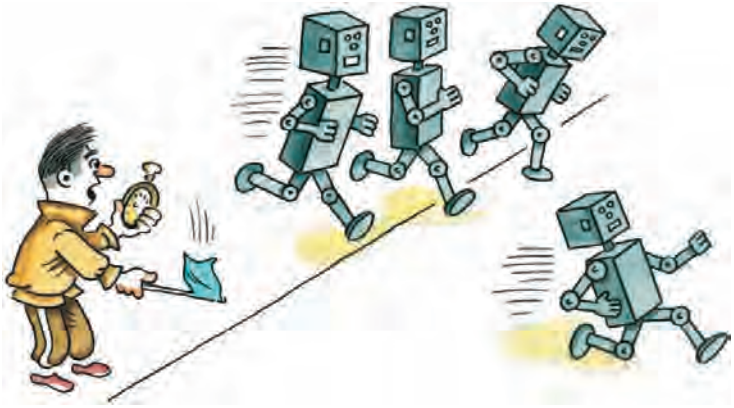
В то время, когда начинала складываться теория алгоритмов, вычислительные возможности человека всё ещё оставались крайне скромными. Поэтому в течение длительного времени разработка алгоритмов на практике не представлялась возможной. Лишь с появлением первых компьютеров учёные стали уделять особое внимание вопросам создания алгоритмов. С началом их практического использования выяснилось, что огромное значение имеет ещё одно свойство — *эффективность*. Не только весь алгоритм, но и каждый его шаг должны быть такими, чтобы исполнитель был способен выполнить их за конечное время, лежащее в некоторых разумных пределах. А если речь идёт о вычислительной машине, то и в пределах, обусловленных конкретным запасом памяти. С эффективностью алгоритма интуитивно связано понятие *стоимости*, понимаемое как количественная оценка требуемых для его реализации ресурсов. Более точный смысл понятию стоимости придаёт понятие *сложности алгоритмов*. Оно овеществляет стоимость во времени и в пространстве, позволяет оценить срок выполнения и объём требуемой при этом памяти.

Временна́я сложность алгоритма — это время T, необходимое для его выполнения. Оно равно произведению числа элементарных шагов алгоритма на среднее время выполнения одного шага, $T=kt$. Поскольку t зависит от исполнителя, реализующего алгоритм, то естественно считать, что сложность алгоритма в первую очередь зависит именно от k. *Пространственная сложность алгоритма* выражается количеством единиц памяти, требуемых для его реализации.

Пространственная сложность не может расти быстрее временной сложности, ведь на использование каждой дополнительной единицы памяти обязательно потребуется дополнительное время. Поэтому для простоты часто ограничиваются рассмотрением лишь временной сложности.

На практике сложность алгоритма обычно связывают с основным его параметром, существенно влияющим на количество выполняемых действий. Как правило, это размер массива обрабатываемых данных (размер задачи). Например, в задаче умножения двух квадратных матриц это количество их строк, в задаче упорядочивания в алфавитном порядке слов некоторого текста — количество слов и др. Сложность алгоритма (число операций, необходимых для его реализации) выражают в виде функции от объема входных данных.

Например, время работы алгоритма, выполняющего только операции чтения и занесения n данных



в оперативную память, определяется формулой $T=an+b$, где a — время чтения записи, b — суммарное время вспомогательных операций. Коэффициенты a и b неизвестны, но они зависят от характеристик компьютера, транслятора и т. д. Однако наиболее важен характер зависимости от основного параметра, поскольку именно он определяет сложность. Здесь имеет место линейная зависимость от n, и сложность алгоритма поэтому называется *линейной*. Записывается она как $O(n)$, а читается «O большое от n».

Можно рассмотреть процесс упорядочивания по возрастанию n элементов массива с помощью так называемого прямого выбора. Он выполняется так: определяется наименьший элемент во всём массиве и меняется местами с первым элементом; затем определяется наименьший элемент в оставшейся части массива и меняется местами со вторым элементом и т. д. В первый раз при поиске минимального элемента выполняются $(n-1)$ сравнений, затем $(n-2)$ сравнений. Всего необходимо выполнить $(n-1)+(n-2)+...+1=(n^2-n)/2$ сравнений. Для больших n можно считать, что сложность определяется только старшим членом этого полинома, поэтому её записывают как $O(n^2)$. Говорят, что этот алгоритм имеет *квадратичную сложность*.

А вот сложность алгоритма умножения матриц (таблиц) размера $n \times n$ будет уже *кубической*, $O(n^3)$. Ведь для вычисления каждого элемента результирующей матрицы требуется n умножений и $(n-1)$ сложений, а всего этих элементов n^2.

На практике время выполнения алгоритма может зависеть от набора данных, к которым он применяется. Например, для некоторых алгоритмов сортировки оно существенно сокращается, если первоначально эти данные были частично упорядочены. Чтобы учитывать это, сохраняя возможность анализировать алгоритм независимо от данных, различают:

максимальную сложность $T_{max}(n)$ — время выполнения алгоритма в том случае, когда выбранный набор n данных наиболее увеличивает срок выполнения алгоритма;

среднюю сложность $T_{ср.}(n)$ — среднее время выполнения алгоритма, применённого к n произвольных данных.

Сложность задачи — это сложность наилучшего алгоритма, известного для её решения, поэтому она зависит от уровня развития методов решения. Например, известна задача проверки планарности графа (т. е. проверки того, можно ли изобразить граф на плоскости так, что его рёбра не пересекаются). Первый алгоритм её решения, предложенный в 1930 г., имел сложность $O(n^6)$. Затем его неоднократно улучшали, а в 1974 г. удалось построить алгоритм сложности $O(n)$! Для решения задачи могут быть разработаны алгоритмы разной сложности. Логично воспользоваться лучшим среди них, обладающим наименьшей сложностью, но одновременно хочется знать: нельзя ли его ещё улучшить? Для этого надо разделить задачи (и алгоритмы) на группы — *классы сложности*.

Ясно, что лучшими являются *линейные* алгоритмы, имеющие сложность

Пусть есть алгоритм А. Почти всегда существует параметр n, характеризующий объём обрабатываемых алгоритмом данных. Пусть функция $T(n)$ — время выполнения А, а f — функция от n. Говорят, что алгоритм А имеет теоретическую сложность $O(f(n))$, если

$$\frac{T(n)}{f(n)} \underset{n\to\infty}{\to} k,$$ где k — константа.

Если алгоритм выполняется за фиксированное время, не зависящее от размера задачи, говорят, что его сложность равна $O(1)$.

порядка $O(n)$, где n — размерность входных данных. Такие алгоритмы действительно существуют. Например, сложение двух чисел столбиком в случае, если одно из них состоит из n, а другое — из m цифр, требует не более $max(n, m)$ сложений и не более $max(n, m)$ запоминаний, т. е. данный алгоритм имеет сложность порядка $O(n+m)$. Это выражение показывает только порядок величины — постоянные факторы в нём не учитываются. К сожалению, алгоритмов, имеющих линейную сложность (а тем более таких, сложность которых не зависит от размерности обрабатываемых данных), крайне мало.

NP-СЛОЖНЫЕ ЗАДАЧИ

Вот известная задача, называемая *задачей коммивояжёра*. Коммивояжёр (разъездной торговец) выезжает из дома, он должен посетить всех своих клиентов, проживающих в разных населённых пунктах. Расстояния между ними известны. Коммивояжёру нужно выбрать такой маршрут, чтобы посетить каждого клиента ровно один раз и при этом проехать минимальное расстояние (или проехать не больше, заданного расстояния). Эти ограничения легко понять. Действительно, коммивояжёр хочет максимально сократить затраты на бензин (или не желает истратить на него больше определённой суммы).

Решение можно искать следующим образом. По очереди рассматриваем все вероятные варианты маршрутов, выбирая среди них тот, который имеет минимальную длину. Количество вариантов маршрутов легко подсчитать — оно равно $n!$ (конечно, это факториал, а не знак восклицания). Значит, чтобы найти самый короткий маршрут, мы должны проверить $n!$ вариантов решения. Но $n!$ растёт даже быстрее, чем 2^n. Так что, если коммивояжёру требуется посетить много клиентов, решение этой задачи становится весьма трудоёмким.

В то же время, перебирая маршруты, можно обнаружить устраивающий вариант очень быстро — если повезёт, то практически сразу. Так что алгоритм поиска решения состоит из двух этапов:

предъявление некоторого маршрута (этап угадывания);

проверка, является ли этот маршрут решением (этап проверки); если является, то выполнение алгоритма прекращается.

Этап проверки имеет полиномиальную сложность (он может

ПОЛИНОМИАЛЬНЫЕ И ЭКСПОНЕНЦИАЛЬНЫЕ АЛГОРИТМЫ

Полиномиальным (или алгоритмом полиномиальной временной сложности) называют алгоритм, у которого временная сложность есть $O(p(n))$, где $p(n)$ — полином от n. Задачи, имеющие алгоритм решения, сложность которого составляет полином заданной, постоянной и не зависящей от n степени, относят к классу P. Это фактически означает, что их можно решить с использованием компьютера. Ведь даже для очень больших значений n общее число операций, необходимых для реализации полиномиального алгоритма, всё равно остаётся таким, что современный компьютер без труда сможет выполнить их за приемлемое время.

Существует множество других задач, сложность которых составляет не менее $O(x^n)$, где x — константа. Такие задачи относятся к классу E и называются *экспоненциальными*. Среди них задача построения всех подмножеств заданного множества или всех поддеревьев заданного графа. В самом деле, если множество M содержит n элементов, то общее количество подмножеств, которые можно образовать из них, равно 2^n. При больших значениях n эта задача становится практически нерешаемой, поскольку алгоритм её решения должен включать в себя 2^n шагов. Соответственно и алгоритмы, в оценку сложности которых n входит в показатель степени, относятся к *экспоненциальным*.

Считают, что задача не является *хорошо решаемой*, пока для неё не получен полиномиальный алгоритм. Если же для её решения не существует подобного алгоритма, то задачу называют *трудно решаемой*. (Правда, при небольших значениях n экспоненциальный алгоритм может быть даже менее сложным, чем полиномиальный. Тем не менее различие между алгоритмами этих типов весьма велико и проявляется при больших значениях n.)

Известно понятие полиномиальной сводимости задач. Говорят, что задача A полиномиально сводится к задаче B, если существует функция полиномиальной сложности, преобразующая решение задачи A в решение задачи B. Установлено, что среди *NP*-сложных задач существуют такие, которые можно полиномиально свести к любой другой задаче из класса *NP*. Их назвали *NP*-полными задачами (к ним относится, в частности, и задача коммивояжёра). Если бы удалось найти детерминированный алгоритм полиномиальной сложности для решения хотя бы одной *NP*-полной задачи, это означало бы, что существуют алгоритмы полиномиальной сложности решения любой задачи из класса *NP*.

заключаться в выполнении всего лишь одного сравнения). Этап угадывания состоит в нахождении очередного маршрута и вычислении его длины; очевидно, что время его выполнения пропорционально количеству населённых пунктов n, т. е. этап тоже имеет полиномиальную сложность. Такие алгоритмы называют *недетерминированными*.

При некоторой доле удачи можно решить задачу коммивояжёра за полиномиальное время. Задачи, имеющие недетерминированное решение, которое удаётся получить за полиномиальное время, были названы *NP-сложными* (от *nondeterministic — «недетерминированные»* и *polinomial — «полиномиальные»*).

Таких задач очень много, они имеют огромное практическое значение. Поэтому крайне важно знать ответ на вопрос: а не существуют ли детерминированные алгоритмы их решения, имеющие полиномиальную сложность? Мнения учёных на этот счёт диаметрально противоположны. Подавляющее большинство из них уверены, что найти детерминированные алгоритмы решения этих задач никогда не удастся.

Существуют алгоритмы решения многих *NP*-сложных задач, которые для значительного диапазона исходных данных обеспечивают приемлемое время решения. В основе таких алгоритмов часто лежит метод *ветвей и границ*, динамически отсекающий бесперспективные направления перебора вариантов решений. Поэтому главным моментом в практическом решении *NP*-сложных задач является умение строить приближённые, или эвристические, алгоритмы, базирующиеся на некоторых догадках *(эвристиках)*. При построении приближённых алгоритмов обычно требуют, чтобы они имели полиномиальную сложность. Известно, что иногда удаётся найти решение, лишь на 10—15 % хуже оптимального. Однако, несмотря на такую хорошую статистику, задачи по-прежнему остаются *NP*-сложными и, к сожалению, не имеют гарантированных оценок времени решения. Даже незначительное изменение исходных данных способно привести к его резкому скачку. В каких-то случаях это время может быть недопустимо велико.

ТЕОРИЯ ФОРМАЛЬНЫХ ЯЗЫКОВ. ПЕРЕВОД И КОМПИЛЯЦИЯ

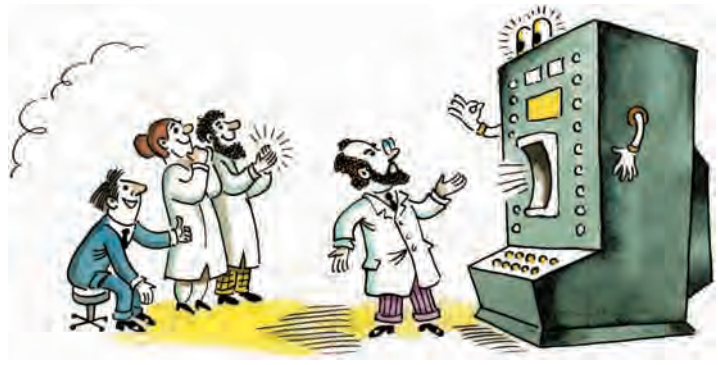
Программы для самых первых компьютеров записывались как последовательность машинных команд, представленных в виде двоичных или восьмеричных кодов. Конечно, составлять и отлаживать большие программы

в таком виде было очень тяжело. Поэтому программисты всегда стремились научить компьютер понимать язык, приближенный к человеческому. Так родились Ассемблер, FORTRAN, Алгол-60, C++ и другие *языки программирования.* Но все они в одном отношении кардинально отличаются от языков естественных.

В устной и письменной речи всегда допустима определённая неточность и недосказанность — слушатель или читатель восстановит смысл из контекста (в общем случае этот контекст включает весь культурный багаж человечества). Компьютер же нуждается в точных и однозначных инструкциях. Значит, запись компьютерных программ на языках программирования должна подчиняться строгим правилам, не допускающим никакой двусмысленности, а компилятор должен безошибочно превращать их в последовательности машинных команд. Здесь не обойтись без помощи математики — самой точной из наук. Неудивительно, что математические методы анализа «машинных» языков были разработаны и широко применяются при создании компиляторов.

ОПРЕДЕЛЕНИЕ ФОРМАЛЬНОГО ЯЗЫКА

Предложение в любом языке состоит из элементарных символов: букв, цифр, пробелов и знаков препинания. Поэтому можно было бы формально определить предложение как последовательность или цепочку символов. Однако такое определение слишком широкое. В огромном количестве всевозможных цепочек символов имеет смысл выделить подмножество *правильно построенных цепочек* — тех, которые состоят из «правильных слов, расставленных в правильном порядке». Например, цепочки «рорылво ргнцои орвылора иыврпфг» и «По карета дороге плывёт» не являются правильно составленными предложениями русского языка, предложение же «Карета плывёт по дороге» правильно составлено, хотя и сомнительно с точки зрения смысла.



В искусственных языках, к которым относятся и языки программирования, множество правильно построенных предложений может быть заранее точно описано с помощью формальных математических конструкций. Поэтому такие языки называют *формальными языками*.

Формальный язык считается определённым, если заданы: 1) *алфавит*, который включает не только буквы, но и все символы данного языка, в том числе цифры, пробелы, знаки препинания; 2) *критерий*, позволяющий однозначно определить, является данная цепочка символов из алфавита языка правильно построенной или нет. Каким образом задан этот критерий, не имеет значения. В простейшем случае можно просто перечислить все допустимые цепочки. Или составить свод правил «создания» подобных цепочек. Такой свод правил называется *грамматикой,* а строение цепочек языка — *синтаксисом.*

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ

Контекстно-свободные грамматики были придуманы вовсе не математиками или программистами — впервые их использовали лингвисты для описания естественных языков.

Составные части цепочки обозначают либо буквами, не входящими в алфавит языка, либо названиями понятий — *метасимволами,* или *нетерминалами,* заключёнными в угловые скобки. Можно рассмотреть наиболее простые фразы русского языка,

Грамматика называется контекстно-свободной, поскольку при замене нетерминала на правую часть правила никак не учитывается, в каком месте текущей цепочки он стоит: интерпретация понятия не зависит от контекста.

в которых используются нетерминалы: предложение, подлежащее, группа подлежащего, сказуемое, группа сказуемого, определение, обстоятельство. Грамматика состоит из *правил*. В левой части правила записывается нетерминал, в правой части — произвольная цепочка из терминалов и нетерминалов, где как бы раскрывается смысл понятия, указанного в левой части. Правая и левая части правила разделяются стрелочкой, которая читается «это есть».

Первое правило этой грамматики определяет, что предложение состоит из двух частей — группы подлежащего, за которой следует группа сказуемого.

⟨предложение⟩→
⟨группа подлежащего⟩ ⟨группа сказуемого⟩

В грамматике должны быть расписаны правила для всех понятий языка. В примере можно ввести два правила для группы подлежащего:

⟨группа подлежащего⟩→
⟨подлежащее⟩
⟨группа подлежащего⟩→
⟨определение⟩ ⟨группа подлежащего⟩.

Во втором правиле нетерминал группа подлежащего встречается как в левой, так и в правой частях. Это важная черта контекстно-свободной грамматики, позволяющая строить сколь угодно длинные фразы. В данном случае второе правило позволяет добавлять слева любое количество определений к подлежащему.

Если в грамматике имеется несколько правил с одинаковой левой частью, то их можно свести в одну запись, соединив все правые части при помощи вертикальной черты (она читается как «или»):

⟨группа подлежащего⟩→
⟨подлежащее⟩ | ⟨определение⟩⟨группа подлежащего⟩.

Вот все правила грамматики:

⟨предложение⟩→
⟨группа подлежащего⟩ ⟨группа сказуемого⟩;
⟨группа подлежащего⟩→
⟨подлежащее⟩ | ⟨определение⟩ ⟨группа подлежащего⟩;
⟨подлежащее⟩→лодка | карета;
⟨определение⟩→зелёная | белая | большая;
⟨группа сказуемого⟩→
⟨сказуемое⟩ ⟨обстоятельство⟩;
⟨сказуемое⟩→катится | плывёт;
⟨обстоятельство⟩→ по дороге | по реке.

В грамматике выделяется начальный нетерминал; цепочки языка выводятся из него. В примере это нетерминал ⟨предложение⟩. На первом шаге к нему применяют одно из правил, левая часть которого совпадает с данным нетерминалом. Происходит замена символа на правую часть правила. Последовательность таких замен называется *выводом*. Вывод закончен, когда в цепочке не останется ни одного нетерминала.

Рассмотрим пример вывода. Используем единственное правило для нетерминала ⟨предложение⟩:

⟨предложение⟩→
⟨группа подлежащего⟩ ⟨группа сказуемого⟩.

Затем к нетерминалу ⟨группа подлежащего⟩ применяется второе правило:

⟨группа подлежащего⟩⟨группа сказуемого⟩→
⟨определение⟩⟨группа подлежащего⟩⟨группа сказуемого⟩.

Продолжим вывод:

⟨определение⟩⟨группа подлежащего⟩⟨группа сказуемого⟩→
зелёная ⟨группа подлежащего⟩⟨группа сказуемого⟩→
зелёная ⟨подлежащее⟩ ⟨группа сказуемого⟩→
зелёная карета ⟨группа сказуемого⟩→
зелёная карета ⟨сказуемое⟩ ⟨обстоятельство⟩→
зелёная карета катится ⟨обстоятельство⟩→
зелёная карета катится по дороге.

На последнем шаге получился законченный вывод. Аналогично этому можно построить и другие варианты ⟨предложения⟩:

большая белая лодка плывёт по реке;
карета плывёт по дороге;
белая зелёная карета катится по реке
и т. п.

Два последних примера показывают, что грамматика никак не учитывает смысл фразы.

ЯЗЫК АРИФМЕТИЧЕСКИХ ФОРМУЛ

Можно рассмотреть пример, более важный с точки зрения программирования, определив понятие *арифметического выражения,* включающего имена переменных, целые константы, знаки арифметических операций, а также скобки. Имена переменных состоят из одной латинской буквы.

Задана грамматика:

⟨формула⟩→⟨элементарная формула⟩|⟨формула⟩ + ⟨формула⟩|⟨формула⟩ - ⟨формула⟩|⟨формула⟩ * ⟨формула⟩|⟨формула⟩ / ⟨формула⟩;
⟨элементарная формула⟩→⟨переменная⟩|⟨константа⟩|(⟨формула⟩);
⟨переменная⟩→ *a*|*b*|*c*|... *z*;
⟨константа⟩→⟨цифра⟩|⟨константа⟩⟨цифра⟩;
⟨цифра⟩→ 0|1|2|...|9.

Язык, заданный этой грамматикой, состоит из всех правильных арифметических выражений, включающих однобуквенные имена переменных и целые константы. Вот пример вывода цепочки *x*+7:

⟨формула⟩→⟨формула⟩ + ⟨формула⟩→
⟨формула⟩ + ⟨элементарная формула⟩→
⟨формула⟩ +⟨константа⟩→⟨формула⟩ + ⟨цифра⟩→
⟨формула⟩ +7→
⟨элементарная формула⟩ + 7→
⟨переменная⟩ + 7 → *x* + 7.

На каждом шаге вывода применяется замена по одному из правил к самому правому нетерминалу текущей цепочки. Такой вывод называют *правым*. Существует ещё и левый вывод. Всякую цепочку можно вывести с их помощью. Правда, в некоторых грамматиках выводы бывают не единственными. *Недетерминированность* неформально означает неоднозначность трактовки предложений языка. Приведённая грамматика языка арифметических формул недетерминированная: формула 1-2-3 имеет два разных правых вывода:

1) ⟨формула⟩→⟨формула⟩ - ⟨формула⟩→
⟨формула⟩ - ⟨элементарная формула⟩→ ...
⟨формула⟩ - 3 → ... → 1-2-3;
2) ⟨формула⟩→⟨формула⟩ - ⟨формула⟩→
⟨формула⟩ - ⟨формула⟩ - ⟨формула⟩→ ...
⟨формула⟩ - 2-3 → ... → 1-2-3.

Первый вывод соответствует тому, что операции выполняются слева направо, т. е. операция 1-2 сначала представлена в виде подформулы, а затем от неё вычитают 3. Второй вывод: операции осуществляют справа налево,

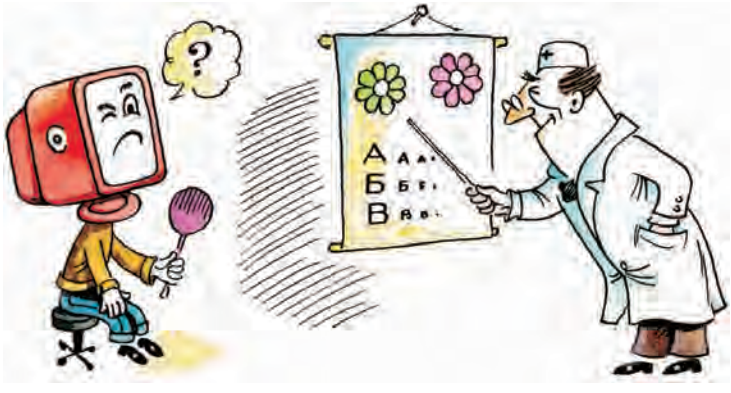
т. е. 2-3 — подформула, которую вычитают из единицы. При реальных вычислениях в первом случае результат (1-2)-3=-4, во втором случае 1-(2-3)=2. Формула трактуется неоднозначно, т. е. грамматика неудачна.

Грамматику называют *детерминированной*, если правый вывод произвольной цепочки единственный. В этом примере грамматику можно сделать детерминированной, введя дополнительные понятия ⟨терм⟩ и ⟨множитель⟩:

⟨формула⟩→⟨формула⟩+⟨терм⟩ | ⟨формула⟩ - ⟨терм⟩ | ⟨терм⟩;
⟨терм⟩→⟨терм⟩ * ⟨множитель⟩ | ⟨терм⟩ / ⟨множитель⟩ | ⟨множитель⟩;
⟨множитель⟩→⟨элементарная формула⟩.

Остальные правила такие же, как и в первом случае. Здесь формула представляется в виде суммы или разности термов, каждый терм — это произведение или частное, множитель — это либо переменная или константа, либо произвольная формула, заключённая в скобки. Данная грамматика уже детерминированная, она соответствует выполнению операций слева направо, как это и принято в математике при отсутствии скобок. Кроме того, она учитывает, что приоритет операций умножения и деления выше, чем сложения и вычитания.

Современные алгоритмические языки описываются с помощью контекстно-свободных грамматик. Более того, учёные предполагают, что и естественные языки можно задать таким же образом. Интересно, что контекстно-свободный язык, заданный грамматикой языка С, помимо правильных С-программ обязательно включает в себя и некоторые ошибочные программы. Однако это связано не с синтаксисом, а с неописанными переменными и неправильными типами выражений. Точно так же контекстно-свободная грамматика русского языка, если бы её кто-нибудь смог выписать, позволяла бы наряду с нормальными предложениями строить фразы, лишённые смысла.



АВТОМАТНЫЕ И РЕГУЛЯРНЫЕ ЯЗЫКИ

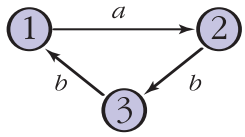
Контекстно-свободная грамматика обычно описывает синтаксис языка на уровне строения предложения. Например, с точки зрения синтаксиса русского языка нет никакой разницы между прилагательными «зелёный» и «жёлтый», точно так же в любом алгоритмическом языке нет грамматического различия между разными целыми константами. Поэтому грамматику алгоритмического языка удобно задавать, беря в качестве элементарных понятия «переменная», «константа», а не отдельные буквы и цифры, составляющие их.

«Язык слов» отличается от «языка предложений», его строение значительно проще и не требует применения контекстно-свободных грамматик. В большинстве алгоритмических языков отдельные слова задаются с помощью *конечных автоматов* или *регулярных выражений.*

Конечный автомат — это ориентированный граф, т. е. фигура, состоящая из вершин, соединённых между собой стрелочками-*рёбрами* (допустимы рёбра-петли, соединяющие вершины сами с собой). Рёбра помечены буквами исходного алфавита. Одна из вершин графа обозначается как начальная, и несколько вершин — как конечные. Вершины графа называют также *состояниями* конечного автомата, а рёбра — *переходами.*

Язык, заданный конечным автоматом, состоит из всевозможных цепочек, которые можно прочесть, двигаясь из начальной вершины в одну из конечных вершин по рёбрам

графа. Рассмотрим, например, конечный автомат



где вершина *1* — начальная, а вершина *2* — конечная. Язык, заданный этим автоматом, состоит из цепочек *a, abba, abbabba, abbabbabba, ...*, соответствующих путям, начинающимся в вершине *1* и заканчивающимся в вершине *2*.

Всякий *автоматный* язык является контекстно-свободным. Обратное неверно, класс контекстно-свободных языков шире, чем класс автоматных языков. Например, язык, состоящий из всевозможных цепочек нулей и единиц, в которых количество нулей равно количеству единиц, является контекстно-свободным, но не автоматным.

Существует другой способ задания автоматных языков — с помощью *регулярных выражений*. Регулярные выражения включают в себя несколько

простых операций, позволяющих из уже построенных регулярных языков строить новые языки. Это операции *конкатенации* (соединения двух цепочек), *выбора* из двух вариантов, а также *итерации* (повторения цепочки произвольное число раз). Процесс построения начинается с элементарных языков, содержащих единственную цепочку из одной буквы или пустую цепочку. Теорема Клини

ПРОГРАММЫ LEX И GREP

Операционная система UNIX содержит несколько стандартных программ, работа которых основана на теории регулярных языков. Прежде всего это программа LEX, осуществляющая лексический анализ текста, т. е. разбивающая текст на слова в соответствии с определённым набором регулярных выражений. Также это программа GREP (*англ.* Get Regular Expression Pattern — «выделить образец по регулярному выражению»), осуществляющая контекстный поиск в одном или нескольких файлах. В качестве образца поиска можно задать либо произвольный фрагмент текста, либо регулярное выражение. Причём поиск происходит очень быстро, он занимает время, линейно зависящее от длины входного файла и не зависящее от длины искомого фрагмента. Высокая скорость работы объясняется тем, что программа GREP строит для поиска конечный автомат, на который последовательно подаются символы из входного потока.

```
[root@linus root]# cat russian.hash | tr '\000' '\n' | grep -i '^[фу]..с.[ло]$'
ФИАСКО
ФРЕСНО
УМЫСЕЛ
ФАРСЕЛ
ФРИСКО
УЖАСНО
УЛЬСИО
ФОССЕЛ
[root@linus root]#
```

Результат работы программы GREP.
Требовалось найти в словаре слова из шести символов, начинающиеся с букв «Ф» или «У», оканчивающиеся на «Л» или «О», с четвёртой буквой «С».

Задача разбора предложения естественного языка, вообще говоря, не имеет однозначного решения. Например, по фразе «запорожец обогнал мерседес» не вполне ясно, кто кого обогнал, т. е. что в этом предложении является подлежащим, а что — дополнением. Смысл фразы иногда можно понять по контексту, но это недоступно компьютеру.

утверждает, что классы автоматных и регулярных языков совпадают, т. е. всякий автоматный язык можно задать посредством регулярного выражения и, наоборот, всякий регулярный язык можно задать с помощью конечного автомата. В примере регулярное выражение следующее:

$$a(bba)^*,$$

где звёздочкой обозначена итерация, а скобки используются для группировки.

Задание языка с помощью регулярного выражения более удобно для человека, а представление языка с помощью конечного автомата — для компьютера.

ОСНОВНЫЕ ЭТАПЫ РАБОТЫ КОМПИЛЯТОРА

Работа любого компилятора состоит из несколько шагов.

Шаг 1: разбор исходной программы и перевод её в некоторое внутреннее представление, удобное для дальнейшей работы. На этом этапе выявляются также ошибки в тексте исходной программы.

Шаг 2: перевод программы на промежуточный язык, не зависящий от системы команд конкретного компьютера.

Шаг 3: оптимизация кода программы.

Шаг 4: генерация выходного кода на Ассемблере или непосредственно на языке машинных команд.

Программисты научились быстро и эффективно выполнять шаги 1, 2 и 4. Основным достижением стало использование теории формальных языков и разработанных на её основе программ автоматической генерации компиляторов.

Шаг 3 является наиболее трудоёмким и наименее формализованным и представляет собой основное поле битвы между различными компаниями, выпускающими компиляторы.

Бóльшая часть усилий разработчиков компилятора приходится именно на оптимизацию кода.

ЗАДАЧА РАЗБОРА

На первом шаге работы компилятора (разбор исходной программы) требуется определить, является ли заданная цепочка символов правильно построенным предложением, и если является, то восстановить её вывод. Процесс восстановления вывода называют *разбором* или *синтаксическим анализом,* а программу, осуществляющую разбор, — *парсером* (от *англ.* to parse — «разбирать»).

Можно рассмотреть, систему автоматического перевода с английского языка на русский. Перевод отдельных слов не даёт результата, потому что синтаксис английских и русских предложений не совпадает. Поэтому нужно сначала осуществить разбор английского предложения, т. е. понять, что является подлежащим, сказуемым, дополнением и т. д., и только потом составить соответствующее русское предложение по правилам русского языка.

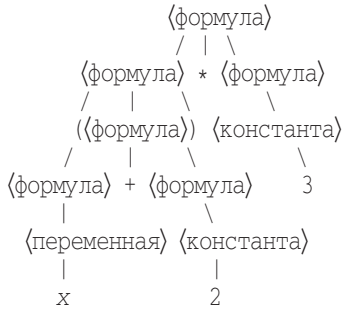
В отличие от естественных языков, задача разбора для алгоритмических языков всегда имеет однозначное решение. Транслятор, переводящий программу с алгоритмического языка высокого уровня на Ассемблер или язык

машинных команд, решает задачу разбора на первом шаге работы. При этом он преобразует программу в некоторое внутреннее представление; как правило, это представление с помощью *синтаксических деревьев*.

Понятие синтаксического дерева можно проиллюстрировать на примере. Пусть дана грамматика

⟨формула⟩→⟨формула⟩ + ⟨формула⟩ |
⟨формула⟩ * ⟨формула⟩ | (⟨формула⟩) |
⟨переменная⟩ | ⟨константа⟩

(для краткости опущены правила для нетерминалов ⟨переменная⟩ и ⟨константа⟩). Вот синтаксическое дерево для выражения (*x*+2)*3:



Каждая нетерминальная вершина синтаксического дерева соответствует одному шагу вывода цепочки. Вершины нумеруются нетерминалами, которые заменяются на правую часть соответствующего правила при выводе. Из вершины выходят ветви ко всем символам, составляющим правую часть правила.

По выводу цепочки однозначно восстанавливается синтаксическое дерево и, наоборот, по синтаксическому дереву однозначно выписываются правый и левый выводы.

Теоретически для всех контекстно-свободных грамматик задача разбора имеет алгоритмическое решение, но общий алгоритм работает слишком медленно, за время порядка N^3, где N — длина входной цепочки. Это неприемлемо для трансляторов: при увеличении длины входной программы в 10 раз время трансляции возрастает в 1000 раз! Поэтому на практике работают только со специальными классами грамматик, допускающими более быстрый разбор. Важнее всего два алгоритма разбора.

• *Нисходящий (рекурсивный)* разбор. Класс грамматик, разрешающих нисходящий разбор, называют LL(1)-грамматиками.

• *Восходящий* разбор, или разбор с помощью конечного автомата со стеком (иногда говорят «автомата с магазинной памятью»). Соответствующий класс грамматик называется LR(1)-грамматиками.

В обоих случаях входная цепочка просматривается слева направо. Первая буква L в названии класса означает направление просмотра (от *англ.* left — «левый»). В первом случае строятся левые выводы, во втором — правые, на что указывает вторая буква в названии (от *англ.* right — «правый»). Единица в названии поясняет, что алгоритм просматривает входную цепочку на один символ вперёд.

При нисходящем разборе синтаксическое дерево строится сверху вниз, начиная с корневой вершины, а при восходящем разборе — восстанавливается снизу вверх.

В грамматике арифметических формул имелось правило:

⟨формула⟩→⟨формула⟩ + ⟨терм⟩.

Попытка выполнить нисходящий разбор в этой грамматике приведёт к бесконечной рекурсии, поскольку подпрограмма ⟨формула⟩ будет постоянно обращаться сама к себе, не прочитав ни одного символа, содержащегося во входной цепочке. Следовательно, эта грамматика не допускает нисходящего разбора (но допускает восходящий разбор!).

Всякая грамматика, обладающая свойством LL(1), располагает также свойством LR(1), т. е. класс LR(1)-грамматик включает в себя класс LL(1). В то же время есть LR(1)-грамматики, которые не являются LL(1): например, детерминированная грамматика языка арифметических формул и грамматика языка C. Вообще, построить леворекурсивную грамматику всегда сложно: даже если это удаётся сделать, результат выглядит крайне неестественно. Напротив, LR(1)-грамматика обычно очень легко выписывается для большинства алгоритмических языков.

Зачем же вообще рассматривать LL(1)-грамматики и рекурсивные алгоритмы? Единственная причина состоит в том, что понимание алгоритма восходящего разбора, использующегося в большинстве случаев, требует некоторой программистской культуры. Нужно хотя бы знать, что такое стек.

Поэтому многие пособия по программированию для начинающих ограничиваются только рекурсивными алгоритмами разбора.

НИСХОДЯЩИЙ (РЕКУРСИВНЫЙ) РАЗБОР

Пусть в грамматике имеется правило

$$A \to BtD,$$

где A, B, D — нетерминалы, а t — терминал (т. е. обычный символ). Надо создать подпрограмму с именем «A», которая пытается выделить из входной цепочки начальный отрезок, выводимый из нетерминала A. По указанному правилу этот отрезок должен начинаться с отрезка, выводимого из нетерминала B; затем следует обычный символ t и далее отрезок, выводимый из нетерминала D. Подпрограмма A сначала вызывает подпрограмму B, которая читает и разбирает начальный отрезок цепочки, соответствующий нетерминалу B. Затем подпрограмма A проверяет, совпадает ли очередной символ цепочки с символом t, пропускает его и вызывает подпрограмму D. Таким образом, обработка отрезка, представляющего нетерминал A, сводится к обработке более коротких отрезков, представляющих нетерминалы B и D.

Если имеются два разных правила для нетерминала A, например

$$A \to BtD,$$
$$A \to EFg,$$

то подпрограмма A по первому символу входной цепочки должна разобраться, какое из этих двух правил нужно применить. Это можно сделать, только если любые цепочки, выводимые из нетерминалов B и E, обязательно начинаются с разных символов.

Слово «рекурсивный» в названии алгоритма употребляется потому, что в определении нетерминала B снова может встретиться нетерминал A (либо в грамматике будет правило, которое содержит A как в левой, так и в правой частях). Следовательно, подпрограмма B, в свою очередь, вызывает A (либо подпрограмма A вызывает сама себя). Поэтому при нисходящем разборе возможна рекурсия.

ВОСХОДЯЩИЙ РАЗБОР

В алгоритме восходящего разбора используется стек. Его можно представлять как слово, состоящее из терминалов и нетерминалов, которое наращивается или изменяется только с правого конца. Для примера можно рассмотреть грамматику арифметических формул:

$$F \to F+F \mid V \mid C$$
$$V \to x \mid y$$
$$C \to 1 \mid 2 \mid \ldots \mid 9,$$

где нетерминал F обозначает формулу, V — переменную, C — константу. Тут ограничились только операцией сложения, двумя переменными x, y и цифрами в качестве констант.

Разберём цепочку $x+y+1$. Алгоритм моделирует так называемый LR-процесс. На любом шаге состояние LR-процесса определяется стеком и непрочитанной частью исходной цепочки. Стек как бы повёрнут навстречу входной цепочке. Вначале стек пуст, а входная цепочка совпадает с исходной:

$$[\ \Leftarrow x+y+1$$

Алгоритм может выполнять два действия. Первое называется *сдвигом* и состоит в переносе первого символа входной цепочки в стек (как на арабских счётах, символ словно «перещёлкивается» справа налево):

$$[\ \Leftarrow x+y+1 \ \to\ [x \Leftarrow +y+1$$

Пусть вершина стека, т. е. конец слова, представляющего стек, совпадает с правой частью некоторого правила, в данном случае правила $V \to x$. Тогда вершину стека можно заменить на левую часть правила. Это действие называется *свёрткой* по соответствующему правилу:

$$[x \Leftarrow +y+1 \ \to\ [V \Leftarrow +y+1$$

Теперь выполняется свёртка по правилу $F \to V$, т. е. символ V в вершине стека заменяется на F:

$$[V \Leftarrow +y+1 \ \to\ [F \Leftarrow +y+1$$

Затем дважды выполняется сдвиг:

$$[F \Leftarrow +y+1 \ \to\ [F+ \Leftarrow y+1 \to\ [F+y \Leftarrow +1$$

Теперь выполняется свёртка по правилу $V \to y$, затем свёртка по правилу $F \to V$:

$$[F+y \Leftarrow +1 \ \to\ [F+V \Leftarrow +1 \to\ [F+F \Leftarrow +1$$

и, наконец, свёртка по правилу $F \to F+F$, т. е. производится замена цепочки $F+F$ в вершине стека на левую часть правила — на букву F:

$$[F+F \Leftarrow +1 \quad \to\ [F \Leftarrow +1$$

Далее дважды выполняется сдвиг, затем свёртка по правилу $C \to 1$, потом свёртка по правилу $F \to C$ и на последнем шаге свёртка по правилу $F \to F+F$:

$$[F \Leftarrow +1 \to\ [F+ \Leftarrow 1 \ \to\ [F+1 \Leftarrow \to$$

$$[F+C \Leftarrow \ \to\ [F+F \Leftarrow \ \to\ [F \Leftarrow$$

По окончании исходная цепочка полностью прочитана, а стек LR-процесса содержит начальный нетерминал грамматики. Следовательно, разбор доведён до успешного завершения, а исходная цепочка принадлежит языку.

При восходящем разборе может возникнуть неоднозначность. Так, в некоторый момент можно выполнить сдвиг или свёртку либо свёртки по разным правилам. В алгоритме разбора

строится конечный автомат, по вершине стека и по первому символу входной цепочки определяющий, какое именно действие следует выполнить.

Существует чёткая процедура построения этого автомата. Однако для реальных алгоритмических языков число его состояний может измеряться тысячами. Для человека это очень много, и кто хоть раз попробовал вручную выписать все состояния и заполнить таблицы разбора, скорее напишет программу, делающую это автоматически, чем согласится снова выполнить подобную работу. Впрочем, писать такую программу не нужно, она уже давным-давно написана и называется YACC.

ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА РАЗРАБОТКИ КОМПИЛЯТОРОВ LEX И YACC

Для решения задачи синтаксического разбора текста предназначены программы LEX и YACC (*англ.* Yet Another Compiler Compiler — «ещё один компилятор компиляторов»). Синтаксический разбор выполняется в два этапа. На первом символы разбиваются на слова (лексемы) с помощью сканера (*англ.* Scaner; это программа на языке C, текст которой автоматически воспроизводится программой LEX). Входной файл для программы LEX содержит перечень регулярных выражений, представляющих всевозможные типы слов исходного языка.

На втором этапе сканер передаёт парсеру очередное слово, указывая его тип (переменная, константа, ключевое слово и т. п.) и необходимую дополнительную информацию — например, для констант передаются их числовые значения. Парсер реализует алгоритм восходящего разбора (LR-разбора). Параллельно стеку LR- процесса парсер поддерживает семантический стек, позволяющий программисту хранить любую дополнительную информацию, связанную с понятиями разбираемого языка, и таким образом решать необходимую ему задачу. Например, в случае перевода с английского языка на русский для слов можно хранить их значение, для определений — ссылку на определяемое существительное, чтобы согласовать род, и т. д.

Текст программы парсера на языке C создаётся в результате работы YACC. Созданный YACC файл с программой на C затем используется как часть большого проекта, реализующего компилятор.

Такая технология разработки компиляторов имеет огромные преимущества по сравнению с примитивным «ручным» программированием. Во-первых, она экономит силы и время и, следовательно, уменьшает возможность ошибок. Во-вторых, контекстно-свободная грамматика входного языка просто переписывается из соответствующего стандарта. Это гарантирует, что все созданные подобным образом компиляторы будут соответствовать стандарту и даже на разных компьютерах станут работать одинаково (по крайней мере, на стадии синтаксического разбора). А переносимость программ с одной платформы на другую по-прежнему является одной из главных болевых точек в программировании. Хотя текст парсера весьма изящен и вообще всегда приятно использовать достижения науки на практике, подчас приходится реализовывать компиляторы вручную, отодвинув в сторону два замечательных инструмента LEX и YACC. К сожалению, за красоту приходится платить потерей производительности.

ДОКАЗАТЕЛЬСТВО ПРАВИЛЬНОСТИ ПРОГРАММ

Доказательство правильности программ напоминает доказательство математических теорем. Каждая программа рассматривается как теорема такого вида: «Если исходные данные удовлетворяют некоторым условиям, то после выполнения данной программы для её результатов будут справедливы некоторые другие условия».

Доказательство обычно включает следующие этапы:

1. В программе выделяют несколько контрольных точек, для каждой из них формулируется утверждение, описывающее состояние переменных программы в данной точке. Эти утверждения используются при доказательстве как промежуточные опорные точки (в математике аналогичную роль играют *леммы*).

2. Исходя из логики программы составляется её блок-схема, показывающая, из какой точки в какую возможен переход, при каком условии и как при этом изменяются значения переменных.

3. Для каждого перехода доказывается следующее положение: если в исходной точке выполнялось записанное в ней утверждение, то и утверждение, записанное в конечной точке перехода, выполняется. Если это действительно для всех переходов, выполнение программы дойдёт до заключительной точки, заключительное утверждение будет верно.

Перечисленные пункты показывают, что, если программа завершится, результат её будет правильным, однако они не гарантируют, что она будет завершена. Поэтому полное доказательство обязательно включает ещё один пункт.

4. Доказывается что заключитиельная точка программы рано или поздно будет достигнута, т. е. программа завершится.

Пример 1. Есть мешок арбузов. С помощью чашечных весов (без гирь) необходимо найти самый тяжёлый арбуз. Возможное решение этой задачи представлено на рисунке в виде блок-схемы. Алгоритмически его можно записать так:

```
(1) самый тяжёлый арбуз — в мешке
нц пока в мешке больше одного арбуза
положить два любых арбуза из мешка на чашки весов
(2) самый тяжёлый арбуз — в мешке или на весах
более лёгкий арбуз отложить в сторону
более тяжёлый арбуз вернуть в мешок
кц
(3) в мешке ровно один арбуз — самый тяжёлый
взять из мешка самый тяжёлый арбуз
```

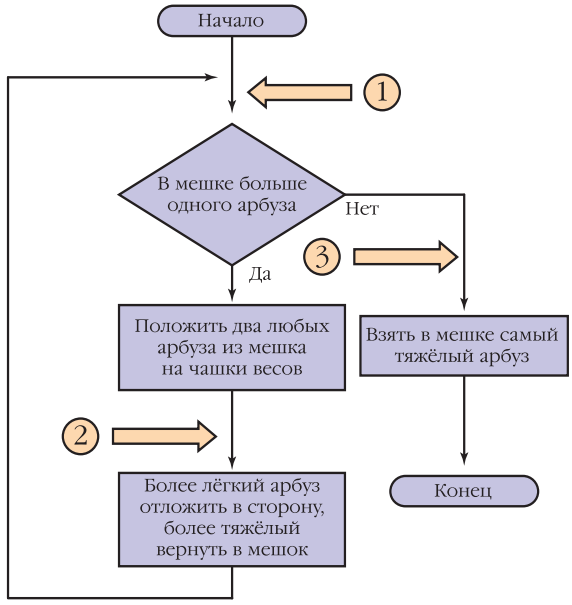
Примечание: предполагается, что в мешке нет одинаковых арбузов.

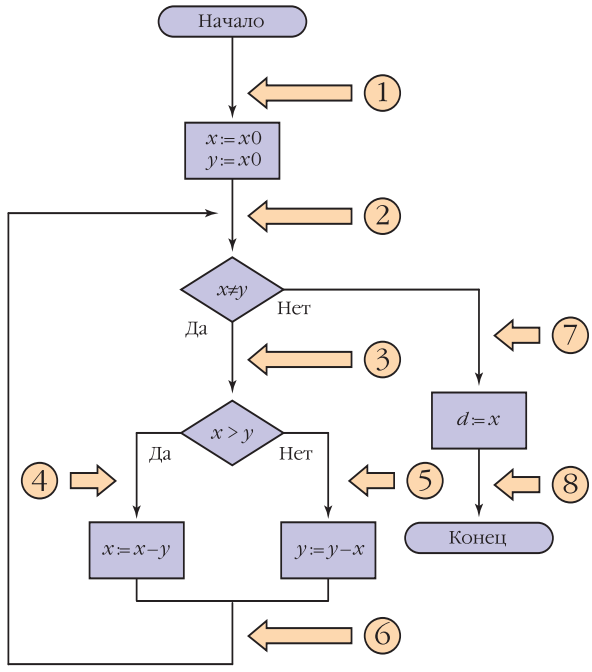
Контрольные точки отмечены на блок-схеме кружками, соответствующие утверждения записаны в алгоритме в виде комментариев.

Доказательство правильности этого алгоритма:

1. В начальный момент условие (1) верно: самый тяжёлый арбуз находится в мешке.

2. Если верно условие (1), то при входе в цикл верно условие (2). Это очевидно: в промежутке между этими точками арбузы из мешка «уходили» только на весы.





3. Если верно условие (2), то при следующем проходе цикла верно условие (1). То есть если в точке (2) самый тяжёлый арбуз находился в мешке, то он там и остался: между точками (2) и (1) из мешка ничего не вынимали. Если же в точке (2) самый тяжёлый арбуз был на весах, то при выполнении следующих действий он был возвращён в мешок.

4. Если верно условие (1), то при выходе из цикла верно условие (3). Выполнение условия (1) означает, что мешок не пуст. Переход в (3) возможен, только если не выполняется условие цикла, т. е. в мешке не больше одного арбуза. Следовательно, в мешке ровно один арбуз, и в соответствии с условием (1) он — самый тяжёлый.

5. Выполнение алгоритма обязательно завершится, так как при каждом исполнении цикла в мешке становится на один арбуз меньше.

Пример 2. На рисунке изображены блок-схема алгоритма нахождения наибольшего общего делителя двух целых чисел (алгоритм Евклида) и процедура на языке Pascal.

Правильность этого алгоритма можно доказать следующим образом:

1. Выделяем контрольные точки и записываем соответствующие им утверждения.

2. Возможные переходы, условия и изменения переменных видны на рисунке.

3. Все утверждения в контрольных точках верны:

(1) x0 > 0, y0 > 0

Данное положение верно по определению:

(2) $x > 0$, $y > 0$, НОД(x,y) = = НОД($x0$, $y0$)

В эту точку возможен переход из двух мест: из точек (1) и (6). Переход из (1) происходит в начале работы программы, из (6) — на каждом очередном витке цикла. Справедливость утверждения (2) надо доказывать отдельно для каждого перехода.

Между точками (1) и (2) выполняются присваивания: $x := x0$; $y := y0$.

После этих присваиваний $x = x0$, $y = y0$. Если было выполнено (1), то справедливость (2) очевидна.

Между (6) и (2) не происходит вообще никаких действий, а утверждения в этих точках совпадают. Следовательно, если выполняется утверждение (6), то выполняется и утверждение (2).

Утверждение (2), расположенное при входе в цикл, играет особую роль в доказательстве правильности программ.

$$(3)\ x > 0,\ y > 0,\ x <> y,$$
$$НОД(x, y) = НОД(x0, y0)$$

При переходе из (2) в (3) никаких изменений в значениях переменных нет, значит, все условия, входящие в утверждение (2), остаются верными. В точке (3) к ним добавилось условие из заголовка цикла — $x <> y$. Здесь оно обязательно верно, так как в противном случае тело цикла не выполнялось бы.

$$(4)\ x > y > 0,$$
$$НОД(x, y) = НОД(x0, y0)$$

В точку (4) совершается переход из (3) при условии $x > y$. Если добавить это условие к утверждению (3), как раз и получится утверждение (4).

$$(5)\ y > x > 0,$$
$$НОД(x, y) = НОД(x0, y0)$$

В точку (5) переходим из (3), если не выполняется предыдущее условие $x > y$, т. е. при $y >= x$. Случай $y = x$ исключается, так как в (3) утверждается, что $x <> y$.

$$(6)\ x > 0,\ y > 0,$$
$$НОД(x, y) = НОД(x0, y0)$$

В точку (6) переходим из (4) и из (5).

При переходе из (4) выполняется присваивание $x := x - y$. Предположим, что старое значение x (до присваивания) было x_c, а новое (после присваивания) — $x\_н$.

В соответствии с (4) $x\_c > y$, поэтому $x\_н = x\_c - y > 0$.

Пусть d — произвольный общий делитель x_c и y. Очевидно, что разность $x\_c - y$ делится на d, значит, d — общий делитель $x\_н$ и y.

Пусть теперь d — произвольный общий делитель $x\_н$ и y. Очевидно, что сумма $x\_н + y$ делится на d, значит, d — общий делитель x_c и y.

Отсюда следует, что множество общих делителей x_c и y совпадает со множеством общих делителей $x\_н$ и y. А значит, и наибольшие общие делители этих пар совпадают, т. е. $НОД(x\_н, y) = НОД(x\_c, y)$.

Утверждение (6) доказано.

$$(7)\ x = y > 0,$$
$$НОД(x, y) = НОД(x0, y0)$$

Переход в (7) происходит из (2) при нарушении условия $x <> y$, т. е. при $x = y$. Утверждение (7) — это и есть утверждение (2), к которому добавилось условие $x = y$.

$$(8)\ d = НОД(x0, y0)$$

В (7) утверждается, что $x = y$. $НОД(x, x) = x$, значит, после присваивания $d := x$ условие (8) выполнено.

4. Программа обязательно завершится, поскольку при каждом исполнении большее из чисел x уменьшается, а числа при этом должны оставаться положительными.

Евклид.

Утверждение, которое всегда выполняется перед проверкой условия цикла, называется *инвариантом цикла.*

АЛГОРИТМ ЭВКЛИДА НА ЯЗЫКЕ PASCAL

```
procedure Euclid (x0, y0: integer; var d: integer)
  var x, y: integer
begin
  {(1) x0 > 0, y0 > 0}
  x:=x0; y:=y0;
  {(2) x > 0, y > 0, НОД(x, y) = НОД(x0, y0)}
    while x <> y do begin
    {(3) x > 0, y > 0, x <> y,
    НОД(x, y) = НОД(x0,y0)}
    if x > y
      then {(4) x > y > 0, НОД(x, y) = НОД(x0, y0)}
      x:=x - y
      else {(5) y > x > 0, НОД(x, y) = НОД(x0, y0)}
      y:=y - x
    {(6) x > 0, y > 0, НОД(x, y) = НОД(x0, y0)}
  end
  {(7) x = y > 0, НОД(x, y) = НОД(x0, y0)}
  d:=x;
  {(8) d = НОД(x0, y0)}
end
```

ЯЗЫКИ ПРОГРАММИРОВАНИЯ

ЯЗЫК ПРОГРАММИРОВАНИЯ ЛОГО

Язык программирования ЛОГО создал в 1969 г. выдающийся американский учёный Сеймур Пайперт, известный своими работами в области педагогики, математики, психологии и информатики. Главная идея языка — программирование деятельности исполнителя, управляемого с помощью компьютера. По существу, ЛОГО — это язык управления исполнителем. Среди многочисленных реализаций систем программирования ЛОГО есть и такие, где роль программно-управляемых исполнителей играют аппаратные роботы. Однако наибольшую популярность получил программно-реализованный исполнитель Черепашка, который перемещается по плоскости с поднятым или опущенным пером, что позволяет ему оставлять след — нарисованную траекторию движения.

Система команд, непосредственно воспринимаемых и выполняемых Черепашкой, как правило, невелика и проста. В эту систему команд включаются, в частности, перемещения вперёд и назад на задаваемую пользователем длину, поворот на указанный угол, а также операции с рисующим инструментом — «поднять» и «опустить».

ЛОГО ориентирован на обучение программированию младших школьников. Тот факт, что этот язык не умер (как случилось с многими искусственными языками), а уже треть века живёт и продолжает развиваться, свидетельствует, что выбранная

предметная область (раннее обучение детей информатике) и предложенные разработчиками языка средства удачно сочетаются друг с другом.

В язык управления Черепашкой встроены управляющие структуры, которые позволяют передавать группы команд, сформированные с помощью повторений и ветвлений. Типичный пример — цикл ЛОГО:

```
ПОВТОРИ 360 [ ВПЕРЁД 1 НАПРАВО 1]
```

Этот цикл рисует «окружность», составленную из 360 отрезков, а условная команда

```
ЕСЛИ :A [ НАПРАВО 90 ] [ НАЛЕВО 90 ]
```

заставляет Черепашку повернуть под прямым углом направо или налево в зависимости от текущего значения переменной :A.

В ЛОГО просто и естественно записываются и используются процедуры. Поэтому процедура как фундаментальное понятие информатики вводится на самых первых уроках. Действительно, цикл

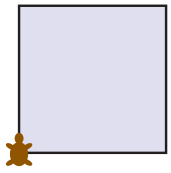
```
ПОВТОРИ 4 [ВПЕРЁД 10 НАПРАВО 90]
```

легко приводит к описанию процедуры КВАДРАТ:

```
ЭТО КВАДРАТ
  ПОВТОРИ 4 [ВПЕРЁД 10 НАПРАВО 90]
КОНЕЦ
```

Здесь заголовок описания — его первая строка — включает режим запоминания (а не выполнения!) процедуры, названной КВАДРАТ, а слово КОНЕЦ, завершающее описание, возвращает программу из режима запоминания в режим выполнения. После этого квадрат можно рисовать простым вызовом процедуры, т. е. упоминанием её имени:

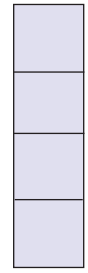
```
КВАДРАТ
```



Например, с помощью такой процедуры можно выложить стенку из квадратных кирпичей — написать процедуру СТЕНА, которая вызывает КВАДРАТ:

```
ЭТО СТЕНА
ПОВТОРИ 4
 [
  КВАДРАТ
  ПЕРОПОДНЯТЬ
  ВПЕРЁД 10
  ПЕРООПУСТИТЬ
 ]
КОНЕЦ
```

СТЕНА

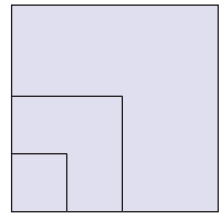


Столь же естественно в описание процедуры вводятся параметры: если тот же квадрат приходится рисовать с разными длинами сторон, то с этой целью полезно описать процедуру с параметром (величина :*СТОРОНА* — параметр процедуры КВАДРАТИК):

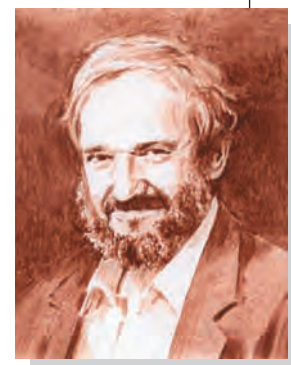
```
ЭТО КВАДРАТИК :СТОРОНА
  ПОВТОРИ 4
    [
      ВПЕРЁД :СТОРОНА НАПРАВО 90
    ]
КОНЕЦ
```

Теперь при вызове КВАДРАТИК с разными значениями фактического параметра можно с помощью одной и той же процедуры рисовать квадраты разных размеров:

```
КВАДРАТИК 10
КВАДРАТИК 20
КВАДРАТИК 30
```



Коль скоро одна процедура способна вызывать другую, образуя цепочку вложенных процедур, то можно представить и такой частный случай,



Сеймур Пайперт.

когда имена вызывающей и вызываемой процедур совпадают, т. е. процедура вызывает сама себя. Процедуру, вызывающую (непосредственно или через цепочку вложенных вызовов) процедуру с тем же именем, называют *рекурсивной,* а соответствующий вызов — *рекурсивным вызовом,* или, короче, *рекурсией.* В ЛОГО такие вызовы разрешены (существуют языки, в которых рекурсивные вызовы не допускаются в принципе). Рекурсивная процедура

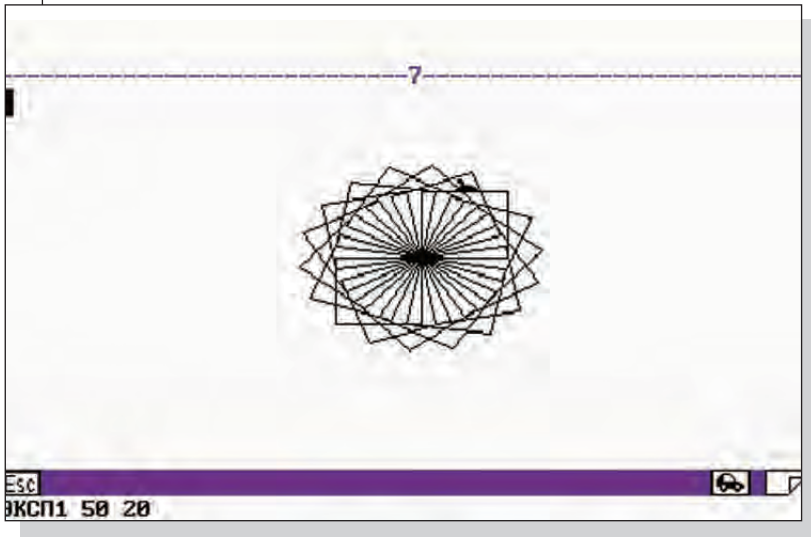
```
ЭТО КВАДРАТИК :СТОРОНА
  ПОВТОРИ 4
    [
      ВПЕРЁД :СТОРОНА НАПРАВО 90
    ]
  КВАДРАТИК :СТОРОНА
КОНЕЦ
```

заставит Черепашку безостановочно двигаться по одному и тому же пути (описывать квадрат).

Более интересны рекурсивные вызовы процедур, в которых значение параметра вызова (фактического параметра) меняется от одного рекурсивного вызова к другому. Вот пример такой процедуры

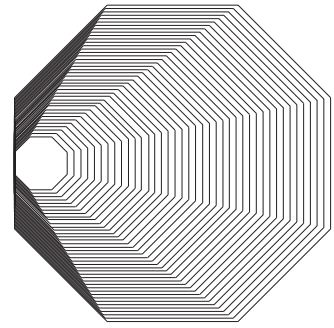
```
ЭТО УЗОР :A
  ПОВТОРИ 8
    [ВПЕРЁД :A НАПРАВО 45]
  УЗОР :A + 2
КОНЕЦ
```

Спираль с шагом 50 и углом 20.



и построенный ею рисунок:

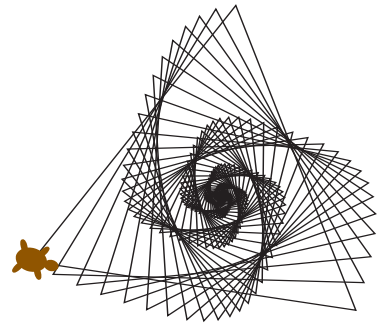
УЗОР 10

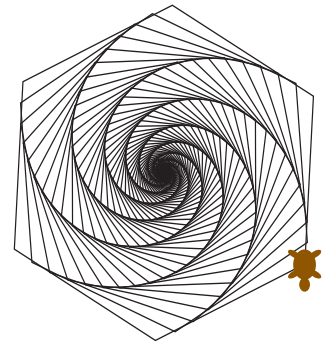


Рекурсивная процедура рисования спирали имеет два параметра — *ШАГ* и *УГОЛ*:

```
ЭТО СПИРАЛЬ :ШАГ :УГОЛ
  ВПЕРЁД :ШАГ НАПРАВО :УГОЛ
  СПИРАЛЬ (:ШАГ + 5) :УГОЛ
КОНЕЦ
```

Примеры нескольких спиралей при разных значениях угла:



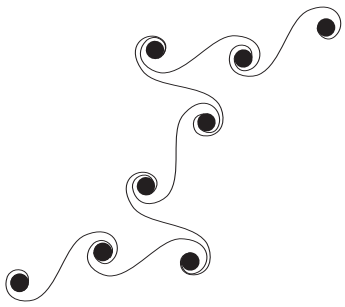


Другой пример не сложнее, но он более привлекателен. Здесь при каж-

дом рекурсивном вызове меняется не длина перемещения, а угол поворота:

```
ЭТО СПИРАЛЬ :ШАГ :УГОЛ
   ВПЕРЁД :ШАГ НАПРАВО :УГОЛ
   СПИРАЛЬ :ШАГ (:УГОЛ + 7)
КОНЕЦ
```

Результат выполнения этой процедуры:



За годы развития Черепашка Пайперта и управляющий ею язык ЛОГО обросли многими новыми возможностями. Сначала появились средства управления цветом. Позднее Черепашка «научилась» воспроизводить мелодии. Ещё интереснее стало работать на ЛОГО, когда появилась возможность в одной программе управлять одновременно несколькими исполнителями.

Язык ЛОГО не только отображал в своём инструментарии средства для обслуживания новых возможностей Черепашки, но и развивался автономно, позволяя обучать программируемый исполнитель новым «умениям».

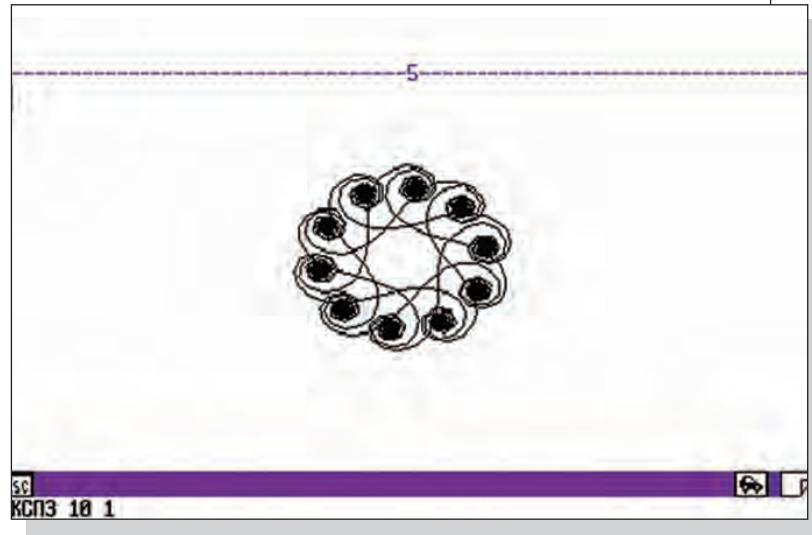
Так, благодаря вычислительным возможностям при работе с числовой информацией и богатым средствам обработки текстовой информации язык ЛОГО допускает работу не только с графикой, но и другими способами представления информации. ЛОГО предлагает эффективные методы обработки таких структур данных, как списки. Важно, что идея рекурсивности проводится в структурах данных столь же последовательно, как и в управляющих структурах. Вот, например, как запрограммировано в ЛОГО последовательное «урезание» слова «ПОБЕДА»:

```
ЭТО УРЕЗАНИЕ :СЛОВО
   ПОКАЖИ БЕЗПЕРВОГО :СЛОВО
   УРЕЗАНИЕ БЕЗПЕРВОГО :СЛОВО
КОНЕЦ
```

На экран будет выведен такой столбец слов:

```
ОБЕДА
БЕДА
ЕДА
ДА
А
```

В ЛОГО удачно сочетаются возможности «учебного» и производственного языка. Интересную интеграцию различных областей применения этого языка продемонстрировал созданный профессором М. Вивé научный центр во французском городе Ле-Мане. Здесь одновременно проводили обучение программированию в начальной школе; подготовку специализирующихся в информатике студентов университета; переподготовку квалифицированных рабочих на промышленных предприятиях региона, где широко использовались станки с числовым программным управлением; исследования психологов, изучавших влияние информационных технологий на мышление и деятельность людей. ЛОГО оказался эффективным инструментом для пользователей с разными уровнями развития и разных сфер деятельности.



Спираль с шагом 10 и углом 1.

ЛОГО рос и развивался не только за счёт увеличения функций языка и Черепашки. Его распространение в мире связано в первую очередь с многообразием лексических версий. Практически в каждой развитой стране используются свои национальные версии этого языка. Хотя, например, в основе французского языка заложена та же латиница, что и в английском (в лексике которого создавались первые версии ЛОГО), тем не менее они предпочитают рассказывать юным французским школьникам о Черепашке, понимающей французские ключевые слова ЛОГО.

Та же картина и в странах, где в национальных языках используется кириллица. Несмотря на близость славянских языков, ЛОГО существует в русской, украинской и болгарской версиях. Более того, при изучении языка ЛОГО в русских школах часто включают такое задание — запрограммировать действия Черепашки при помощи команд на английском языке. И дети легко делают это:

```
ЭТО FORWARD :ШАГ          ЭТО LEFT :УГОЛ
  ВПЕРЁД :ШАГ               НАЛЕВО :УГОЛ
КОНЕЦ                     КОНЕЦ
```

Необходимо ли такое разнообразие лексических версий языка программирования? На этот вопрос последовал бы, вероятно, отрицательный ответ, если бы речь шла о «профессиональном», производственном языке, таком, как C, FORTRAN, Ada, или о языке для изучения информатики старшеклассниками или студентами (таков Pascal). Но малышей, которым приходится осваивать азы программирования — дисциплины и без того насыщенной непростыми фундаментальными понятиями, очень важно освободить от дополнительных методических препятствий, каким, безусловно, является лексика чужого, иностранного языка.

ИСТОРИЯ ЯЗЫКОВ ПРОГРАММИРОВАНИЯ

Самым распространённым способом дрессуры домашних животных является обучение их ограниченному набору команд, например «сидеть», «лежать» и т. д. Однако это не всегда применимо к кошачьим, которые мало того что непослушные, но и не понимают команды. Некоторые прозорливые хозяева легко управляются со своенравными питомцами, разговаривая на их языке, шипя и мяукая, и, поверьте, достигают практически полного взаимопонимания.

Центральное устройство в компьютере — процессор также обладает своим собственным языком. И сделать так, чтобы компьютер стал «понимать» программиста, весьма сложная задача.

Команды для процессора надо как-то закодировать, например представить в числовой форме: 1 может означать сложение, 2 — умножение, 3 — деление и т. д. Помимо команд процессору предоставляются и данные, необходимые для выполнения тех или иных операций. Пусть у некоторого вымышленного процессора для простоты такой формат записи команд:

код команды	первый операнд	второй операнд	ячейка для результата

А данные для команды, т. е. *операнды,* записываются так:

00 xx — число xx;
01 xx — ячейка памяти с номером xx.

Тогда программа нахождения среднего арифметического двух чисел, расположенных в ячейках 1 и 2, выглядит так:

01 03	0101 0103	0102 0002	0103 0103

Или, если записать её в строчку,

01 01 01 01 02 01 03 03 01 03 00 02 01 03.

Даже для такой простой программы *машинный код* представляет собой загадочный шифр, для разгадывания которого надо обладать способностями Шерлока Холмса. В настоящих, невымышленных, компьютерах машинный код намного сложнее. Программировать на таком коде и проверять программу было очень трудно и неудобно, поскольку для всех операций приходилось либо помнить их коды и форматы, либо постоянно пользоваться специальными таблицами. Малейшая неаккуратность при кодировании, ошибка при записи, перепутанные цифры приводили к непредсказуемым последствиям. А найти такую ошибку оказывалось нелегко, поскольку перед программистами был не алгоритм, описанный более или менее понятным языком, а набор цифр.

Программисты искали способы облегчить себе жизнь, создать инструмент, который позволил бы им писать программы, не тратя время на мелкие детали, не отвлекаясь на само кодирование. Они хотели поручить рутинную работу тому, кто умеет выполнять её лучше всех, — компьютеру. А самим заняться тем, что свойственно человеку, — творчеством, составлением алгоритмов.

ПЕРЕХОД К МНЕМОНИЧЕСКОЙ ЗАПИСИ ПРОГРАММ. ЯЗЫК АССЕМБЛЕР

Норма развития — от простого к сложному. В программировании первым шагом было создание системы символьного обозначения команд. Команды представлялись сокращениями обычных слов, обозначающих операции. Несмотря на простоту идеи, её реализация позволила сильно улучшить восприятие существующих программ, уменьшить количество ошибок, облегчить работу программистов.

Вот пример программы, записанной с помощью символьных обозначений:

```
СЛЖ     ПЧ1, ПЧ2     > ПЧ3
ДЕЛ     ПЧ3, 2       > ПЧ3
```

Это означает СЛоЖить ячейки Памяти №1 и №2, результат поместить в ячейку Памяти №3. Потом ячейку Памяти №3 разДЕЛить на 2 и результат снова поместить в ячейку Памяти №3.

Такая символьная система записи команд называется *мнемонической* (от *греч.* «мнеме» — «память»), т. е. удобной для запоминания.

Но одной только мнемонической записи для эффективного создания программ мало. Программистам приходилось постоянно «изобретать велосипед» — писать код для одного и того же действия. Части кода программ, решавшие ту или иную подзадачу, конечно, можно было сохранить для последующего использования, но ценность такого сохранённого кода невелика, поскольку его нельзя сразу же использовать при создании новых программ.

Части машинного кода были «привязаны» не только к определённому процессору, компьютеру, но и к программе, в рамках которой они написаны. В машинном коде явно заданы адреса ячеек памяти для хранения результатов вычисления и адреса, по которым передаётся управление при наступлении тех или иных событий. А в новой программе эти ячейки памяти могут уже использоваться совсем по-другому. Старый машинный код «почувствовал» бы себя так же, как и шофёр, если бы все улицы города внезапно переименовали. Путевой лист привёл бы совсем в другую часть города.

Можно было сохранить идею, алгоритм, общий принцип решения. А сам машинный код приходилось писать (или исправлять) каждый раз заново.

Для решения задачи придумали систему именования адресов ячеек памяти и участков кода. Каждой ячейке памяти присваивалось своё уникальное имя, и дальше в программе использовался уже не номер ячейки памяти, а соответствующее ей имя:

```
ИМЯ "X"                          №1
ИМЯ "Y"                          №2
ИМЯ "РЕЗУЛЬТАТ"                  №3

СЛЖ   X, Y          >   РЕЗУЛЬТАТ
ДЕЛ   РЕЗУЛЬТАТ, 2  >   РЕЗУЛЬТАТ
```

Таким образом, стало возможным выделение частей кода для решения отдельных подзадач в качестве единого целого. Эти небольшие программы называются *подпрограммами* или *процедурами*. У процедуры есть строго определённый набор начальных данных, которые ей необходимы для работы, и набор выходных данных, которые, собственно, и являются результатом её работы.

Для удобства процедуры объединяют в *библиотеки процедур*. Обычно процедуры в рамках одной библиотеки применяются для решения одина-

ковых или похожих задач, обладают сходными правилами и предназначены для совместного использования. Может существовать библиотека процедур для решения алгебраических уравнений, библиотека для обработки звука и видео и даже библиотека для работы с другими библиотеками процедур.

Но недостаточно только придумать красивую систему записи программы. Такую программу компьютер выполнить не сможет, ему необходимы машинные коды. Процесс перевода мнемонической записи в код легко описать в виде простых пошаговых инструкций. Однако если её будет выполнять человек, то вместо преимуществ новой системы возникнет большое количество ошибок, описок и времени это займёт немало. Выполнение подобной скучной, чисто механической, кропотливой и просто неинтересной, но в то же время ответственной работы (по таблицам вместо одного набора значков-операторов подставлять другой набор значков-кодов операций) необходимо передать компьютеру.

В 1952 г. американка Грейс М. Хоппер создала первый в мире мнемонический язык программирования Ассемблер (от *англ.* assemble — «собирать», «компоновать»). Он включал в себя мнемоническую систему команд (список команд), библиотеки процедур и специальную программу для перевода текста программы в машинный код. Такая процедура получения машинного кода называется *компиляцией* (от *англ.* compile — «составлять», «собирать»), а программа, её осуществляющая, — *компилятором,* который также придуман Грейс М. Хоппер.

Это был первый шаг — создание языка программирования как специального языка, понятного и человеку, и компьютеру. До тех пор существовал только один язык — язык компьютеров, а человек как более разумное существо просто «разговаривал» на нём.

Второй шаг — создание библиотеки процедур, т. е. повторное использование кода. До этого, если немного преувеличить, каждому программисту приходилось каждый раз «придумывать» свою алгебру, геометрию, химию, вместо того чтобы воспользоваться готовым учебником и, оттолкнувшись от базовых знаний, изобрести что-то новое. Библиотеки процедур позволили программистам быстрее и качественнее создавать новые сложные программы.

Эти два направления послужили основой для развития новых языков и систем программирования.

ФОРМУЛЫ, ФОРМУЛЫ, ФОРМУЛЫ... FORTRAN

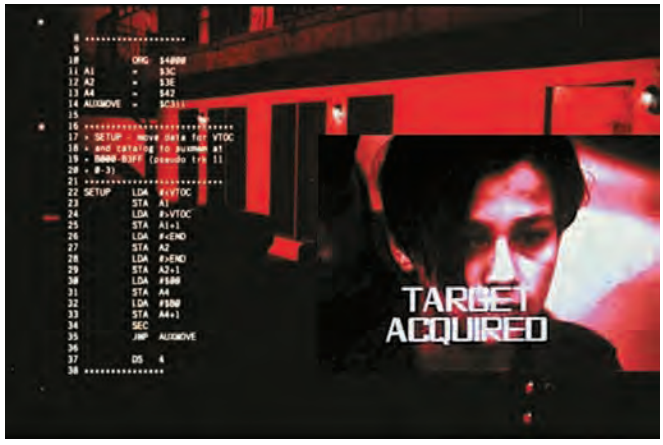
Язык Ассемблер был хорош для программистов, разрабатывающих прикладные программы, операционные системы. Но, несмотря на его мнемоничность, он не подходил для основных пользователей компьютеров того времени — учёных, проводивших сложные математические расчёты.

Язык Ассемблер из-за своей близости к машинному коду сам по себе труден для изучения. Кроме того, он предназначался для определённого процессора, а значит, человек, работающий на нескольких компьютерах, должен был знать несколько иногда существенно разных языков. И наконец, язык Ассемблер представлял собой несколько «очеловеченную» запись машинных кодов.

Для программирования на этом языке необходимо представить все действия в виде простейших кирпичиков, сложные формулы разбить на последовательности отдельных операций. Программист же оперирует в уме более общими категориями: вычислить формулу, напечатать число на экране, повторить операцию десять раз и т. п.

Для решения этих проблем был предложен язык программирования, при создании которого применялся новый способ — разработка языка по предварительной спецификации, описанию. Сотрудники фирмы IBM подготовили документы, подробно описывающие новый язык программирования и требования к нему. В отличие от языка Ассемблер он

Грейс Хоппер.



В фильмах «Терминатор» и «Терминатор-2» робот-убийца запрограммирован на Ассемблере.

разрабатывался не на основе машинного кода, не как язык, понятный компьютеру, а как язык, удобный для человека.

В 1956 г. группой программистов под руководством Джона Бэкуса был создан первый компилятор для нового, более «человечного» языка программирования. Он получил название FORTRAN (от *англ. for*mula *tran*slation — «трансляция», «перевод формул»).

Новый язык программирования стал универсальным, одну и ту же программу можно было запустить на разных компьютерах (естественно, предварительно скомпилировав её). Он позволял записывать формулы в понятном для человека виде. Вычисление среднего арифметического, например, выглядит следующим образом:

```
RES = (A+B)/2
```

Такая запись уже близка к обычной математической, и человеку, пишущему программу для компьютера, не надо учиться хотя бы записывать формулы. Помимо логичного представления вычисления в виде единой формулы FORTRAN предоставил ещё и богатую библиотеку математических функций. Например, для вычисления квадратного корня достаточно вызвать специальную процедуру «сорт»:

```
RES = SQRT(169)
```

Эта процедура может быть включена в сложную формулу как один из операндов:

```
X1 = (-B - SQRT(B*B - 4*A*C)) / (2*A)
```

На языке Ассемблер вычисление корня квадратного уравнения заняло бы много строчек машинного кода, да и операцию вычисления квадратного корня пришлось бы программировать самому или, по крайней мере, взять из библиотеки процедур. Трудно придумать алгоритм вычисления квадратного корня с использованием только четырёх арифметических операций.

Джон Бэкус.

ПРОЦЕДУРНОЕ ПРОГРАММИРОВАНИЕ

Термин «библиотека» подпрограмм или процедур появился ещё при создании языка Ассемблер. Подпрограммы представляли собой выделенные части кода, оформленные как единое целое. Программист должен был определять, каким образом передаются данные в подпрограмму (например, через ячейку оперативной памяти), каким образом подпрограмма запоминает, кто именно её вызвал и как она передаёт управление обратно. Эти *соглашения о вызове подпрограмм* были разными

в различных библиотеках, более того, они могли быть разными для различных подпрограмм в одной библиотеке.

Для организации эффективной работы программистов требовалось автоматизировать, стандартизировать процесс написания и использования подпрограмм. Это осуществилось в языке FORTRAN. То есть подпрограммы, библиотеки подпрограмм уже не являлись чем-то внешним, пристройкой к языку, а были созданы прямо в нём.

Перед выполнением процедуры ей необходимо передать данные, с которыми она будет работать, т. е. *входные данные*. Результат процедуры называется *выходными данными*. Если это процедура вычисления корней квадратного уравнения, то для неё входные данные — три коэффициента при степенях неизвестного, а выходные данные — два искомых корня (хотя, возможно, корень будет всего один, корней не будет или любое число будет удовлетворять уравнению; тогда на выходе процедуры необходимо иметь ещё один параметр — *количество* искомых корней).

Процедура может и не иметь входных или выходных данных, а иногда — и тех и других. Процедура получения текущего времени не имеет входных данных — она берёт время «изнутри» компьютера. У процедуры печати числа нет явных выходных данных — результатом работы является изображение на экране или на принтере. А процедура форматирования всех жёстких дисков системы не имеет ни входных, ни выходных данных. Они ей не нужны, она и так знает, что ей делать, и ей нечего передать наружу — всё будет и так заметно.

При описании процедуры на языке FORTRAN ей даётся имя, по которому её можно будет вызвать. В этой же строчке описываются входные и выходные данные. Каждому элементу данных присваивается имя, под которым оно будет использоваться внутри процедуры:

```
SUBROUTINE SQ (A,B,C,X1,X2)
```

В данном примере описана процедура с именем SQ, у которой есть три входных параметра A, B, C и два выходных — X1, X2. Что является входными и выходными параметрами, определяет программист, пишущий процедуру, и описывает их в комментариях, которые присутствуют в коде программы, но игнорируются компилятором.

```
SUBROUTINE SQ (A,B,C,X1,X2)

C Процедура вычисления корней
C квадратного уравнения
C Входные данные: A,B,C –
C коэффициенты при неизвестном
C Выходные данные: X1,X2 – корни
C квадратного уравнения
C Внимание: работает, только если
C уравнение имеет решение!

D = SQRT(B*B - 4.*A*C)
X1 = (-B-D)/(2.*A)
X2 = (-B+D)/(2.*A)
RETURN
END
```

При вызове необходимо только правильно указать параметры процедуры, и получается решение уравнения:

```
CALL SQ (1,-3,2,RES1,RES2)
PRINT RES1, RES2
```

Машина напечатает ожидаемые решения уравнения $x^2-3x+2=0$:

1 2

Чёткое выделение процедур привело к появлению новой модели программирования.

Большая задача разбивается на множество более мелких простых подзадач. Общий код программы не перегружен деталями, и автор программы всегда может решить, надо ли ему разбираться в алгоритме программы в целом (вероятно, эта теория неприменима для описания полёта ракеты?) или отдельного фрагмента (так вот из-за чего она взорвалась: я не проверил, положительный ли дискриминант при решении квадратного уравнения!).

Такой метод программирования называется *процедурным программированием*. Основой его является разделение большой программы на меньшие кирпичики — процедуры. При использовании технологии процедурного программирования программа может создаваться двумя способами — сверху вниз и снизу вверх.

В первом случае сначала разрабатывается общий алгоритм решения программы; определяется, какие будут нужны вспомогательные функции (процедуры); пишутся специальные пустые «процедуры-заглушки», которые только имитируют работу настоящих процедур; отлаживается общий алгоритм работы, затем пишутся вспомогательные функции.

Во втором случае сначала создаётся набор вспомогательных процедур, они отлаживаются при помощи специальных небольших программ. Например, для процедуры решения квадратного уравнения это может быть программа, которая вводит с клавиатуры коэффициенты уравнения и печатает на экран решения. Когда программист убеждается, что его процедуры работают правильно, он начинает из них, как из кубиков, собирать большую программу.

СТРУКТУРНОЕ ПРОГРАММИРОВАНИЕ. ЯЗЫК ALGOL

За создание следующего языка программирования решили взяться учёные — специалисты в области кибернетики, теории алгоритмов, тех областей науки, которые в англоязычной литературе обычно объединяют под названием «Computer Science», т. е. «наука о компьютерах».

В 1960 г. на основании отчётов научных конференций был создан обобщающий документ, описывающий новый язык — Algol-60. Его наименование произошло от английского algorithmic language — «алгоритмический язык».

Для переменных стали чётко определяться участки программы, в пределах которых эти переменные могут использоваться. Такие участки программы называются *областями видимости,* а про переменную говорят, что она *видна* или *не видна* в данной области программы. Переменная, объявленная внутри функции, процедуры или специального логического

блока операторов, будет видна только внутри данной конструкции.

Например, если в функции *A* есть переменная *tmp* и эта функция вызывает функцию *B*, в которой тоже есть переменная *tmp*, то операции над *tmp* в функции *B* никак не изменят значение *tmp* в функции *A*. Переменная *tmp* из функции *A* не видна в функции *B*.

Стало обязательным указание типа переменных перед их использованием — *объявление*. С введением такого ограничения ошибки, подобные приведённой ниже, оказались невозможными:

```
REAL I1
.... (несколько страниц кода)
I1 = 3./2.
```

Здесь программист, использующий переменную `I1` в операции деления, предполагал, что по описанию стандарта FORTRAN переменная, начинающаяся с буквы `I`, имеет целый тип (целые переменные начинаются с *I*, *J*, *K*, *L*, *M*, *N*, остальные — вещественные). Но он не обратил внимания, что тип переменной был указан необязательным (и, в общем, редко используемым оператором). Как результат значением `I1` является не ожидаемое 1, а 1,5.

Помимо «общеоздоровительного эффекта», связанного с более продуманным использованием переменных, введение областей видимости позволило добавить в язык вызов функции из самой себя. Например, вычисление факториала числа

$$N! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (N-1) \cdot N$$

удобнее всего производить именно с помощью вложенного вызова функции:

```
алг цел факториал (цел N)
  дано N
  надо
нач
| если N<= 1
| | то знач := 1
| | иначе знач := N * факториал (N - 1)
| всё
кон
```

Вызов таких вложенных функций называется рекурсией, а функции, которые вызывают сами себя, — рекурсивными.

В языке Algol были чётко описаны структурные управляющие операторы: операторы выбора и операторы повторения (цикла). Операторы выбора позволяют выполнять те или иные участки кода в зависимости от вычисляемого условия. Операторы цикла предназначены для повторения определённого участка кода заданное количество раз или до тех пор, пока не выполнится определённое условие.

Управляющие операторы, конечно же, присутствовали и в языках Ассемблер и FORTRAN, но они не были основным элементом при написании программ, не обладали удобством использования.

Через восемь лет, в 1968 г., вышел новый, переработанный и дополненный вариант языка Algol-60, который получил вполне предсказуемое имя Algol-68. С этим стандартом связана и опубликованная в 1969 г. статья Эдсгера В. Дейкстры о структурном программировании.

Он доказал, что для записи любого алгоритма достаточно основных конструкций структурного программирования:

Эдсгер Дейкстра.

- последовательность операторов;
- альтернатива (выбор);
- повторение (цикл).

Эта статья подводит итог почти десятилетним исследованиям по структурному программированию. В ней утверждается, что при программировании можно обойтись без оператора перехода на определённую метку в программе (GOTO), при активном использовании которого программы становились непонятными, их логика напоминала «блюдо спагетти». Термин «структурное программирование» отражает тенденции в построении программ с чётким описанием их структуры.

Время шло, и компьютеры, и программы для них становились всё более сложными. Появилось понятие программного продукта, т. е. программы, которая создавалась большим коллективом программистов и была рассчитана на долгосрочное использование в отличие от сиюминутного расчёта уравнения на языке FORTRAN. Программа начинала жить своей собственной жизнью, и очень часто случалось так, что изменения и дополнения в неё вносил человек, не имеющий никакого отношения к её созданию. В связи с этим на первое место стали выходить надёжность программы, понятность её кода, упорядоченность, структурность.

К сожалению, новая версия языка Algol-68 так и не смогла не только завоевать популярность у прикладных программистов, но даже пробиться в ряды коммерческих продуктов. Язык Algol остался научным исследованием, «академическим» языком, который оказал огромное влияние на развитие языков программирования в целом. Он стал фактически стандартом для описания компьютерных алгоритмов в научных исследованиях и породил целое семейство алголоподобных языков (к которым, собственно, и относятся почти все современные процедурные языки, такие, как Pascal, Modula, C).

СТРУКТУРЫ ДАННЫХ. ЯЗЫК PASCAL

В 1970 г. известный швейцарский учёный профессор Никлаус Вирт создал собственный язык программирования, в котором постарался избавиться от всех недостатков языка Algol. Новый язык получил название Pascal.

Этот язык стал первым широко распространённым языком, реализующим в себе концепции структурного программирования, проработанные в 60-х гг. Почему же язык Pascal, созданный одним человеком, смог достичь того, чего не достиг Algol, над которым несколько лет трудилась целая группа видных учёных?

Новый язык вобрал в себя многие черты языка Algol, такие, как строгость описания, богатство управляющих структур, но был более лёгким

для изучения. Это определило возможность его использования для преподавания программирования в университетах, колледжах. Но Pascal избежал участи исключительно «учебного» языка. Он представляет собой достаточно мощный язык программирования, подходящий для создания больших проектов.

Общая структура языка и управляющие конструкции, такие, как циклы, операторы выбора «если — то», напрямую заимствованы из языка Algol. В то же время, поскольку Pascal создавался для обучения программированию, его синтаксис был значительно упрощён. Новый, упрощённый язык позволял практически так же эффективно реализовывать алгоритмы, однако наряду с этим значительно уменьшилось время на изучение языка. Основной его девиз — «Взвешенность, простота, лаконичность».

Но одно лишь упрощение синтаксиса не могло способствовать успешному становлению и развитию нового языка. Каковы же существенные отличия языков Algol и Pascal?

Создатели первого не учли, что компьютер является инструментом не только для сложных расчётов, но и для обработки данных. Способы для выражения данных в языках Algol и FORTRAN — числа, символы (строки символов) и пронумерованные наборы чисел или символов — массивы. Доступ к элементу массива происходит при предъявлении специального жетона — номера этого элемента данных.

Этих трёх типов данных было достаточно на ранних этапах развития компьютеров, когда они в основном *вычисляли* и главным являлось кодирование формул и простых алгоритмов вычисления. Для написания сложных программ этого явно не хватало. Вирт при разработке языка Pascal сделал второй революционный шаг: ввёл новые типы данных и общий механизм для создания производных типов данных, который мог использовать каждый программист в своей программе.

Во-первых, это был ненумерованный набор данных (чисел, символов), представленный как единое целое. В отличие от массива, у элемента такого набора данных нет своего собственного идентификатора, все элементы равноправны. Кроме того, в нём находится по одному элементу с данным значением. Такой набор данных называется *множеством* и в действительности является одним из основных понятий современной математики. Множество применяется в тех ситуациях, когда не важен порядок элементов, даже не важны сами элементы, а важен признак *наличия* элемента данных в наборе.

Например, для коллекционера в общем-то не существенно, какой уникальный номер у марки в его коллекции,

НЕСТРУКТУРИРОВАННЫЙ КОД

Простейшее действие — определение, имеет ли квадратное уравнение корни

```
алг квадратное уравнение
  дано a, b, c
  надо
нач вещ дискриминант
| дискриминант:= b*b-4*a*c
| если дискриминант < 0
| | то вывод "корней нет"
| | иначе вывод "корни есть"
| всё
кон
```

на FORTRANe будет записано так:

```
   DISCR = B*B-4.*A*C
   IF (DISCR) 1,2,2
 1 PRINT 11
11 FORMAT "(Корней нет)"
   GOTO 3
 2 PRINT 22
22 FORMAT "(Корни есть)"
 3 CONTINUE
```

То есть если значение дискриминанта меньше нуля, то происходит переход на оператор с меткой 1, если равно нулю или больше нуля — на оператор с меткой 2. Оператор печати, в свою очередь, ссылается на другой оператор (метки 11 и 22), в котором описан формат выводимого сообщения, в нашем случае слова "корней нет" или "корни есть". Для того чтобы при отрицательном дискриминанте программа после печати соответствующего сообщения не «залезла» на территорию другого варианта, используется оператор перехода GOTO, который просто передаёт управление за пределы нашего блока на оператор с меткой 3.

Никлаус Вирт.

СТРУКТУРИРОВАННЫЙ КОД

```
   DISCR = B*B-4.*A*C
   IF (DISCR) 1,3,2
 1 PRINT 11
11 FORMAT "(Корней нет)"
   GOTO 4
 2 X1 = (-B-D)/(2.*A)
   PRINT 22, X1
22 FORMAT (E8.3, 4X)
 3 X2 = (-B+D)/(2.*A)
   PRINT 33, X2
33 FORMAT (E8.3, 4X)
 4 CONTINUE
```

```
discr:=b*b-4*a*c
if discr < 0 then
  begin
    Print "Корней нет"
  end
else if discr = 0 then
  begin
    x1:=(b)/(2*a)
    Print x1
  end
  else
  begin
    x1:=(b-d)/(2*a)
    x2:=(b+d)/(2*a)
    Print x1,x2
  end
```

По сравнению с программой на языке FORTRAN исчезли операторы перехода на определённую метку. Вместо этого Algol предоставляет блоки операторов, которые выполняются как единое целое. То есть если дискриминант больше нуля, то выполняется блок begin — end, расположенный непосредственно после оператора выбора. Если же дискриминант меньше нуля, то выполнится только следующий блок.

Такая запись, во-первых, улучшает восприятие кода, поскольку не надо отслеживать все переходы по непонятным меткам, человек рассматривает такие блоки как единое целое, он может сначала определить действие в целом (если дискриминант больше или равен нулю, то корни есть, иначе корней нет), а потом уже вникать в то, как устроено данное действие (как именно производится печать текста на экран, каким оператором, какой именно текст печатается).

Во-вторых, при бесконтрольном использовании операторов перехода часто возникала ситуация, когда управление передавалось прямо *внутрь* части кода, выполняющего какое-то действие. Такая методика хотя и позволяла иногда создавать более эффективные программы, вела к огромному количеству ошибок как при самом написании программы, так и при её модификации, когда уже забывалось, что именно и как было реализовано в оригинальной версии.

вряд ли его заинтересуют повторения — ему важен лишь факт наличия или отсутствия у него данного редкого экземпляра. Также ему, возможно, понадобятся следующие операции: *объединение,* которое произойдёт, если два коллекционера решат соединить усилия. Они объединят свои коллекции, а образующиеся повторы раздарят друзьям или поменяют у других коллекционеров на что-то интересное. Скорее всего, ему будет интересно *пересечение,* если он захочет узнать, какие марки есть и у него, и у его друга (потом он в уме вычтет из своей коллекции это пересечение и получит множество марок, которых нет у его коллеги и которыми он может похвастаться).

Во-вторых, в языке Pascal впервые появился универсальный механизм для создания новых сложных структурных типов данных. Он позволял группировать данные, относящиеся к одному объекту, записывать их рядом и потом использовать эти записи как единое целое. Причём группироваться могли не только числа или текстовые строки, а любые другие типы данных. Например, если требуется хранить информацию об ученике, необходимо объединить в одну запись данные о его имени, фамилии, росте. Сначала нужно создать тип данных для хранения имени и фамилии (name_t). Это будет массив из 32 символов:

```
type name_t = array[1..32] of char
```

Потом описывается запись из имени, фамилии и роста ученика. Рост ученика выразим обычным числом

```
type pupil_t = record
  firstname : name_t;
  surname : name_t;
  height : real
end
```

И наконец, для описания класса будет использована запись из массива описаний учеников и номера класса (номер класса может быть только целым числом, поэтому для уменьшения вероятности ошибки мы прямо сейчас скажем компилятору, что это целое число, и любую попытку присвоить ему дробное значение необходимо рассматривать как ошибку):

```
type class_t = record
  pupils : array[1..40] of pupil_t;
  number : integer
end
```

Теперь, для того чтобы напечатать список всех учеников класса, рост ко-

торых выше среднего, можно воспользоваться следующим кодом:

```
var class : class_t;
var N : integer;
{ Здесь запись должна быть заполнена
данными }
...
{ Вычисление среднего роста в классе }
for i := 1 to N do
begin
   sum := sum + class.pupils[i].height
end
mean := sum / N;
{ Печать имён и фамилий учеников
с ростом выше среднего }
for i := 1 to N do
begin
   if( class.pupils[i].height > mean )
   then
      begin
         writeln(class.pupils[i].sur-
name,
            class.pupils[i].firstname)
      end
end
```

Создание языка Pascal явилось значительным этапом в развитии языков программирования и определило путь их развития в 70-х гг. XX в.; большинство языков следующего десятилетия можно считать прямыми потомками языков Algol и Pascal.

В то же время у языка Pascal имелся ряд существенных недостатков. Так, в массивах не могла использоваться

МАССИВЫ

Для хранения и обработки информации о росте всех учеников класса можно создать массив `Height`, в котором будут храниться соответствующие числа. Если необходимо получить рост ученика, у которого номер в журнале 13, то в программе это описывается как `Height[13]`. Следующий пример программы на языке Pascal иллюстрирует нахождение среднего роста учеников в классе.

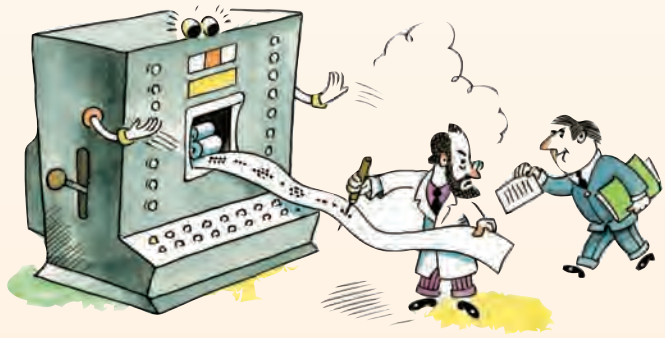
```
{ Пусть Height — массив с ростом учеников,
N — количество учеников в классе }
for i := 1 to N do
begin
   sum := sum + Height[i]
end
mean := sum / N;
```

Если бы не было возможности объединить все данные в массив, то в программе пришлось бы для роста каждого ученика использовать отдельную переменную, и код выглядел бы примерно так:

```
sum := sum + heightVasya;
sum := sum + heightPetya;
sum := sum + heightMasha;
...
```

Да и при изменении количества учеников в классе пришлось бы изменять код программы, а в примере с массивом — только значение переменной *N*. Если данное вычисление используется внутри процедуры и *N* — её параметр, то приведённый выше код с массивом является универсальным и может использоваться в любой программе, в которой необходимо посчитать средний рост школьников в классе (а быть может, и спортсменов в команде или космонавтов на корабле).

ИСТОРИЯ НАПИСАНИЯ ПЕРВОГО КОМПИЛЯТОРА ЯЗЫКА PASCAL



В 1969 г. Вирт поручил одному из своих студентов, Э. Мармье, написать компилятор языка Pascal. Тот начал создавать компилятор на единственном языке программирования, который он знал, — FORTRAN. Это и было ошибкой, поскольку, по признанию самого Мармье, он полностью запутался из-за ограничений при работе со сложными структурами данных.

В 1970 г. после разработки формального описания синтаксиса языка Pascal компилятор создали на самом этом языке. Затем один из разработчиков, Р. Шилд, за две недели вручную перевёл программу компилятора в машинный код. То есть он брал каждую конструкцию языка и записывал её в машинном коде, работая неким «человеком-компилятором». Таким образом, Pascal стал первым языком высокого уровня, реализованным на самом себе.

Для переноса на другие машины использовалась специально разработанная P-машина, исполнявшая так называемый P-код и испытанный временем метод интерпретации. P-код представлял собой машинный код для вымышленного компьютера. Компилятор перевели вручную на этот вымышленный код. На новой же машине создавался только интерпретатор для данного P-кода, что оказалось намного более простой задачей, чем стояла перед Шилдом.

Как только интерпретатор P-кода был сделан, с его помощью стало возможным запустить компилятор, написанный на P-коде, и скомпилировать его ещё раз уже в машинный код для данного компьютера.

Н. Вирт и Э. Дейкстра.

Таким образом, Никлауса Вирта и его язык Pascal можно считать родоначальниками идеи широкого использования виртуальных машин, на которой базируются хорошо известная и широко используемая технология Java и совсем молодая Microsoft.NET.

верхняя «открытая», неизвестная на момент компиляции граница. Это видно из примера, когда для описания класса был использован массив из 40 элементов в надежде на то, что в классе не больше 40 учеников. Но что будет, если найдётся такой гигантский класс?

В языке Pascal не предусмотрен тип данных для работы с символьными строками, несмотря на их широкую распространённость. Каждый раз их приходилось создавать заново (как при определении типа `name_t`, в котором опять-таки предполагалось, что длина имени и фамилии школьника меньше 32 символов).

Pascal не учитывал написания кода программы в разных файлах и последующего объединения при компиляции, поэтому, если создавалась большая программа, объём файла с её текстом мог составлять сотни килобайтов, что делало практически невозможным работу команды программистов. Даже одному программисту — автору — уже было затруднительно найти в ней нужный фрагмент.

В языке Pascal ограничены возможности ввода и вывода информации, не разработана стандартная библиотека вспомогательных функций (например, сравнение строк, которое каждый раз приходилось делать программисту, символ за символом; получение системного времени, работа с файлами операционной системы и др.).

ЯЗЫК C

Один из самых популярных сегодня языков с незамысловатым названием C (читается как буква английского алфавита «с» — «си») создан в 1972 г. Деннисом Ритчи. Язык C был разработан для программирования в новой малоизвестной на тот момент операционной системе UNIX (и это не преувеличение: UNIX тогда использовалась только на нескольких машинах внутри корпорации Bell Telephone Laboratories). Исторически он явился последователем практически неизвестного языка B. Но в то же время он считается наследником языков Algol и Pascal, поскольку вобрал в себя многие их чер-

ты; как и они, это процедурный язык для структурного программирования. Одна из отличительных особенностей C — то, что при его создании не ставились сложные концептуальные цели, такие, как получение универсального языка для записи вычислений (FORTRAN), академического языка для процедурного программирования (Algol), дидактического, учебного языка (Pascal). Язык C был разработан профессиональными программистами как удобный для них язык.

Поскольку язык C — результат труда не комитетов и учёных, а программистов, пишущих реальные программы, в нём были учтены практически все моменты, которые мешали нормальному использованию других языков при создании сложных систем. Снова после языка FORTRAN появилась возможность раздельной компиляции отдельных файлов и сборки их в единую исполняемую программу уже после компиляции (компилятор производил «полуфабрикаты», а специальная программа-сборщик «собирала их на одной сковородке и поджаривала до состояния готовности»). Вместе с языком была создана стандартная библиотека основных функций, облегчающих работу программиста, снимающих с него заботу о написании тривиальных действий и позволяющих сосредоточиться над программой. Причём библиотека основных функций вошла в описание языка, т. е. являлась его неделимой частью, и каждый программист мог быть уверен, что на любой платформе, где есть компилятор C, есть и данная библиотека.

По сравнению с языком Pascal язык C обладал большей гибкостью при выборе разработчиками средств создания программ. Для эффективности стало возможным пожертвовать строгой концепцией структурного программирования и написать программу, используя нетривиальные и временами опасные расширения.

В языке C не только появились новые конструкции, но и прошли доработку старые. С точки зрения управляющих структур ничего кардинального внесено не было, но новые операторы позволили более компактно записать существующие программы, их «смысл» стал ближе машинному коду.

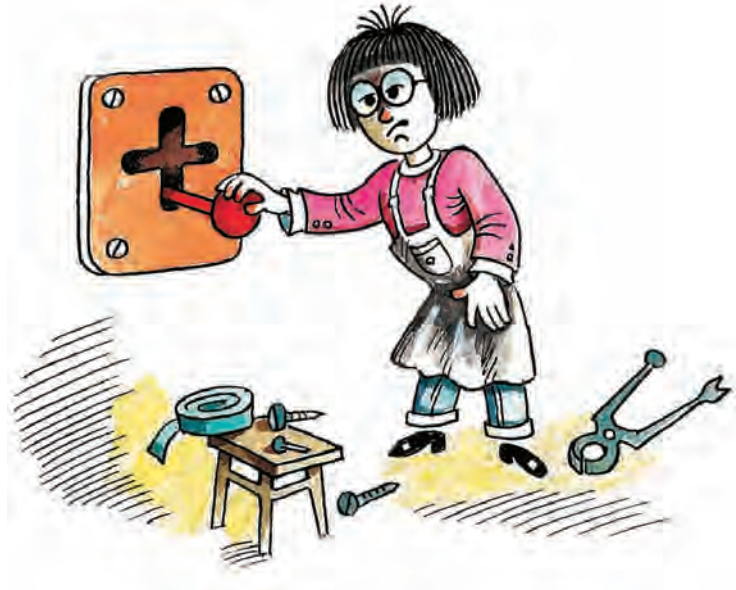
Появился оператор выбора из нескольких альтернатив — switch. Если раньше для выбора того или иного действия в зависимости от значения переменной приходилось использовать цепочку «если — то»:

```
if mark = 3
then
  writeln('Сойдёт ')
else
  if mark = 4
  then
    writeln('Хорошо ')
  else
    if mark = 5
    then
      writeln('Просто замечательно')
```

то теперь в C стало возможным использование одного оператора switch:

```
switch( mark )
{
  case 3:
    printf('Сойдёт');
    break;
  case 4:
    printf('Хорошо');
    break;
  case 5:
    printf('Просто замечательно!');
    break;
}
```

Деннис Ритчи.



МАССИВЫ И УКАЗАТЕЛИ

Если есть массив a и указатель *p*

```
int a[5]; /* Объявление массива целых чисел из 5 элементов */
int *p; /* Объявление указателя на целое число */,
```

то *p* = *a* поместит в p адрес первого элемента массива a. Чтобы получить доступ к нему, используют выражение с указателем в виде *p. При увеличении указателя на 1 он станет указывать на следующий элемент, т. е. запись *(*p* +1) эквивалентна *a*[1]. Указатели можно и вычитать. Если есть два указателя: один — на первый элемент массива, а другой — на какой-то произвольный элемент, то индекс этого элемента получают простым вычитанием указателя начала массива из указателя элемента. То есть, если pa — указатель на начало массива, а *pe* — указатель на произвольный элемент, его индекс *i* равен *pe*-*pa*. Это очевидно, если вспомнить, что указатель на элемент массива является суммой указателя на начало массива и индекса элемента, т. е. *pe* = *pa*+*i*.

Использование указателей позволяет иногда элегантно и необычно решить хорошо известную задачу. Например, копирование одной символьной строки, оканчивающейся символом с нулевым кодом, в другую. Простейшее решение на языке Pascal может выглядеть так:

```
var str : array [0...255] of char;
var copy : array [0...255] of char;
var i : integer;

for i := 0 to 255
begin
   copy[i] := str[i];
   if str[i] = '\0' then goto finish;
end;
finish: ;
```

Оператор for обеспечивает цикл, в начале которого переменной *i* присваивается значение 0, после каждой итерации её значение увеличивается на 1, а условием продолжения цикла является *i*<255. Внутри цикла *i*-й символ копируется из строки str в строку copy. Потом i-й символ строки str сравнивается с нулевым символом ('\0') и, если равенство выполняется, производится выход из цикла с помощью оператора goto.

Если использовать указатели и предположить, что строка обязательно заканчивается символом с нулевым кодом, то копирование строк на языке C может быть намного короче:

```
char *pstr, *pcopy;
pstr = str; pcopy = copy;
...
while(*pcopy++ = *pstr++)
;
```

Используются два указателя на строки pstr и pcopy. Дальше в соответствии с приоритетами операторов в языке C загадочная операция присваивания обрабатывается таким образом:

1. Вычисление правого операнда присваивания.

a) берётся значение указателя pstr, используется для применения оператора *;

b) указатель pstr увеличивается на 1 (т. е. происходит переход к следующему элементу массива).

2. Вычисление левого операнда присваивания.

a) берётся значение указателя pcopy, используется для применения оператора *;

b) указатель pcopy увеличивается на 1 (т. е. происходит переход к следующему элементу массива).

3. Производится само присваивание, т. е. элементу новой строки, полученному на шаге 2, присваивается значение элемента, полученное на шаге 1.

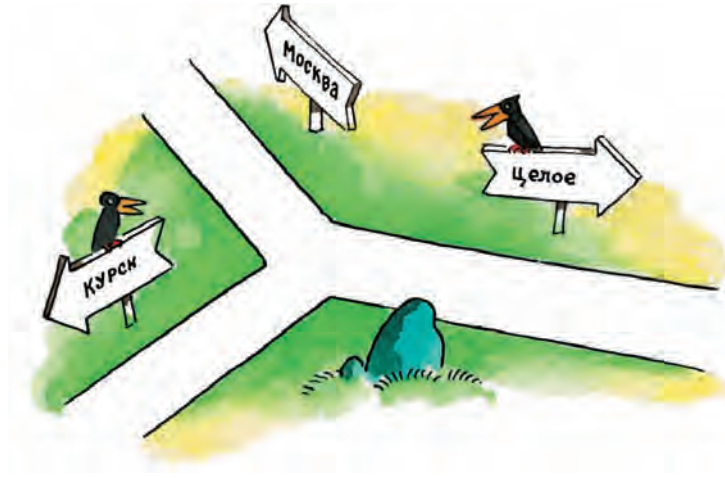
4. Значением всего выражения в целом является значение обрабатываемого элемента массива.

В результате проведения операции были достигнуты все цели. Во-первых, в элемент новой строки попало значение элемента копируемой строки, во-вторых, указатели для обеих строк перешли на следующие элементы, и, наконец, в качестве значения выражения мы получили значение элемента, которое можно сравнить с нулём, чтобы определить допустимый конец строки.

Снова появилась возможность использования глобальных переменных в программе. Переменные, которые используются во всех функциях программы, даже в библиотечных, потенциально опасны, так как могут привести к трудноуловимым ошибкам. Такие переменные противоречат основным академическим идеям о «правильном» структурном программировании. Хотя, с другой стороны, если все функции программы работают с одним и тем же набором переменных, то их можно объявить глобальными, и тогда эти переменные не нужно будет передавать как параметры во все функции, что значительно ускорит работу программы.

В языке С снова появилась возможность работы с указателями на физические ячейки памяти. Конечно, кажется, что это очень опасно. Что будет, если программа запишет произвольные данные в ячейку с адресом 21324, никто не знает. Там могут храниться код операционной системы (хотя правильная операционная система не даст себя «испортить» и просто прекратит работу «взбесившейся» программы) или данные самой программы, и тогда она вроде бы продолжит работу, но результаты будут непредсказуемыми.

Зачем же тогда введение указателей на адреса памяти? Всё дело в том, что при должной квалификации программиста работа с данными программы через их адреса, *адресная арифметика*, позволяет создавать практически такой же эффективный код, как на языке Ассемблер, но со всеми удобствами структурных языков (функции,



управляющие конструкции, составные типы данных и пр.).

Указатели на объекты данных — ссылки — появились ещё в языке Pascal, но их можно было запоминать

ЯЗЫК С И ОПЕРАЦИОННАЯ СИСТЕМА UNIX

В 1973 г. операционную систему UNIX переписали на языке С, изобретённого специально для UNIX. UNIX стал одной из первых операционных систем, написанной на языке программирования высокого уровня. Хотя версия системы на С имела на 20-40 % больший объём и работала медленнее по сравнению с предыдущим вариантом на Ассемблере, преимущества использования языка высокого уровня — возможность легко изучать и изменять программы — перевешивали эти недостатки.

Простота языка С и наличие компилятора, который можно легко адаптировать практически под любую аппаратуру, сделали UNIX первой мобильной ОС, т. е. операционной системой, работающей на ЭВМ с различными архитектурами. Это произошло в 1979 г. с выходом седьмой версии UNIX.

Накопленный опыт переноса UNIX на разные ЭВМ привёл к доработкам языка С: в язык были добавлены контроль типов и библиотека ввода/вывода stdio (произносится «сту-дай-о»).

С 1976 г. создатели UNIX из лабораторий Белла бесплатно распространяли исходные тексты операционной системы в высших учебных заведениях для изучения и использования в процессе обучения. Это повлекло за собой широкое распространение Юникс и как следствие языка С. В первой половине 80-х гг. XX в. Си практически вытеснил все остальные языки программирования.

В 1978 г. вышла книга Брайана Кернигана и Денниса Ритчи «Язык программирования С», которая в течение десяти лет оставалась неформальным стандартом на язык С до появления в 1989 г. международного стандарта на язык С.

PLANKALKUL — ПЕРВЫЙ ЯЗЫК ПРОГРАММИРОВАНИЯ

В последние дни Второй мировой войны немецкий инженер и изобретатель Конрад Цузе оказался в маленьком селении Хинтерштейн в Альпах. Впервые за несколько лет попав в спокойную обстановку (как он писал, «без бомбёжек, телефонных звонков и визитёров») и не имея возможности продолжать работу над своим компьютером Z4, он решил завершить начатое ещё в 1943 г. изучение описаний различных вычислительных задач. Немецкое слово rechnen означает «вычислять», поэтому Цузе использовал для этих описаний термин Rechnenplan — «план вычислений» (слово «программа» было придумано несколько позднее). Разработанную систему для записи «планов вычислений» Цузе назвал Plankalkul. Считается, что это первый в истории язык программирования, более чем на десять лет опередивший FORTRAN и Algol-60.

Однако сам Цузе не рассматривал Plankalkul как язык программирования. Он хотел создать универсальную систему обозначений, дающую возможность описывать любое преобразование информации. Plankalkul можно было использовать для записи алгоритмов — и, в самом деле, Цузе написал на нём множество алгоритмов решения самых разных задач. С неизменным успехом он применял Plankalkul и для описания работы аппаратуры компьютера. Тем более удивительно, насколько Цузе удалось предугадать многие свойства появившихся гораздо позднее языков программирования, а некоторые его идеи реализованы совсем недавно! Что же представлял собой язык Plankalkul?

Прежде всего, в языке принята очень интересная система типов данных. «Вычисления начинаются с бита», как считал Цузе, поэтому простейшим типом у него является бит (состояние да/нет), обозначавшийся *S*0. Биты принимают значения *L* и 0, так что последовательности битов записываются как *L*0*L*0, 0*LL*0 и т. д. Из битов строятся скалярные типы — двоичные целые числа с фиксированной и плавающей запятой. Составные объекты (сейчас их назвали бы структурами данных) также образуются из битов. Например, можно определить массив произвольной размерности: $n \times S0$ (одномерный булевский массив длиной n), $m \times n \times S0$ (двумерный булевский массив $m \times n$) и т. д.

Кроме массивов в языке Plankalkul есть записи, объединяющие данные разных типов (их элементами могут быть и другие записи). Хотя записи и появились в специализированном языке Cobol в 1959 г., затем использовались в языке PL/I и Pascal.

Зато ещё одна идея Цузе долго не имела аналогов в других языках. В текст своих программ он включал выражения, задающие условия, которые должны быть истинными при выполнении тех или иных операторов. Например, пусть объект (массив из четырёх битов) используется для записи десятичных цифр. Но так как цифр всего десять, а объект может принимать 16 разных значений, на них можно наложить ограничение: быть не больше десяти. Только спустя 50 лет аналогичная идея была реализована программистом из Калифорнии Бертраном Мейером, автором языка программирования EIFFEL.

Столь же уникально предложение языка Plankalkul, позволяющее выбирать из множества объектов лишь те, которые обладают заданным свойством. Конечно, это просто сделать с помощью оператора цикла или процедуры, но специальные конструкции есть не во всех современных языках программирования.

Plankalkul достаточно богат выразительными возможностями. Помимо операторов присваивания в нём имеются условные операторы, операторы повторения (аналог современных циклов), возможность обращения к подпрограммам и т. д. Несомненно, он об-

только в других переменных и позже использовать для доступа к необходимым объектам. В языке С указатели на объекты, переменные и функции можно не только запоминать, но также складывать и вычитать.

Язык С предоставляет возможность очень компактного, эффективного решения поставленной задачи, но следствием этого является большая вероятность труднообнаружимых ошибок. Для их исключения необходима высокая степень самоконтроля и мастерства программиста, что, впрочем, не так уж неожиданно, поскольку язык С создавался профессиональными программистами для себя, а в своих-то силах они были уверены.

Программист решает сам, какой код ему нужен: переносимый или эффективный, понятный и надёжный или более производительный? Общая логика программы может учитывать все требования классического структурного программирования, а отдельные функции могут использовать указатели и прочие машинные «фокусы». Программист определяет, что и как ему делать, а единый инструмент С помогает ему в решении этой задачи, не диктуя при этом строгие ограничения, как Pascal или Algol.

Б. Мейер с коллегами в гостях у А. Ершова. 1977 г.

Конрад Цузе.

ладает многими свойствами, которые обычны в современных языках программирования. Зато в нём отсутствуют такие привычные нам возможности, как, например, безусловные переходы или работа со ссылками, кроме того, принятая форма записи страдает избыточностью. Перечень недостатков можно продолжить, но все они не перечёркивают главного: Plankalkul явился первым в истории языком программирования высокого уровня. Хотя он не предназначался для написания компьютерных программ, когда через четверть века алгоритмы Цузе перевели на другие языки, оказалось, что программы работают! Поэтому не будет большим преувеличением сказать, что Цузе стал первым в истории программистом, написавшим множество действительно сложных программ: сортировка данных, анализ синтаксической правильности логических формул, выполнение арифметических действий (в том числе извлечение квадратного корня) над целыми числами и числами с плавающей запятой и др. Особенно интересны его программы для игры в шахматы — фактически первый опыт программирования в области искусственного интеллекта (задолго до формирования этой науки). Кстати, ещё в 1938 г. Цузе предсказывал, что через 50 лет машина начнёт выигрывать у человека в шахматы. Удивительно, но он ошибся всего на несколько лет: чемпион мира Гарри Каспаров проиграл матч суперкомпьютеру Deep Blue в 1997 г.! Впрочем, это Каспаров, а рядовые любители уже лет за десять до этого не могли соревноваться с шахматными программами — так что, может быть, Цузе оказался прав в своём прогнозе.

К сожалению, язык не был реализован ни на одном компьютере, а его полное описание Конрад Цузе опубликовал только в 1972 г. Поэтому многие считают, что Plankalkul не повлиял на развитие современных языков программирования, однако это совсем не так.

Гарри Каспаров играет матч с компьютером Deep Blue.

ОБЪЕКТНО-ОРИЕНТИРОВАННЫЕ ЯЗЫКИ ПРОГРАММИРОВАНИЯ

ОБЪЕКТНО-ОРИЕНТИРОВАННЫЙ ПОДХОД

В 50-х гг. XX в. при написании программ основное внимание уделялось алгоритмам, процедурам, действиям, которые необходимо было предпринять для решения задачи. Подобный процедурный подход был оправдан в те годы, когда сложность программных систем ещё не превысила определённого уровня.

По мере развития компьютеров всё чаще их начинали применять для решения задач, сложность которых обычно обусловлена не алгоритмом, а большим количеством участвующих в процессе объектов и очень сложными отношениями между ними. Алгоритмы, использующиеся для автоматизации деятельности банка, вроде бы не особенно сложны. Они уж точно не включают в себя решения уравнений

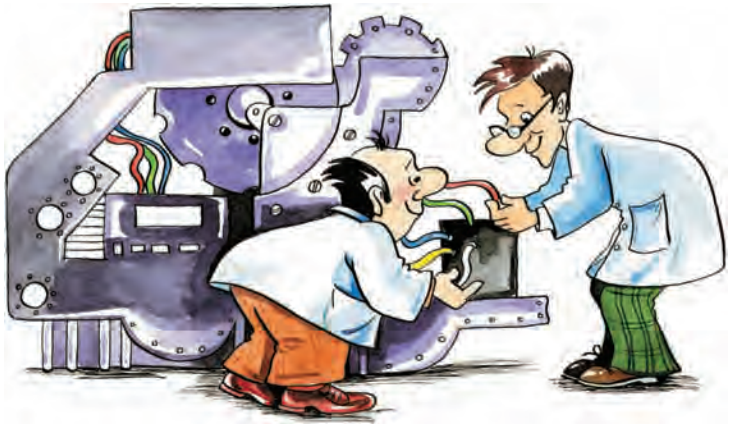
Алан Кей.

аэродинамики или ядерной физики. Но это не означает, что получающаяся программа или скорее даже программный комплекс будут простыми — в повседневной работе банка задействовано большое количество разнообразных объектов: операционисты, брокеры, клерки, клиенты, банкоматы, сети для связи с другими банками, государственные и иные аудиторские комиссии, биржи и т. п. И добиться надёжной работы банковской системы ничуть не проще, чем получить программу, решающую некие уравнения, моделирующие работу ядерного реактора.

Появившиеся в 60—70-х гг. XX в. структурные языки с реализованными в них средствами управления и типами данных облегчили разработку сложных программных комплексов. Но всё же они не давали приемлемого решения проблемы, поскольку пытались внести дополнения и улучшения в уже существующую модель программирования, а не представить новую, которая бы лучше подходила для решения всё возрастающего круга задач.

Процедурные языки программирования исторически развивались как логическое средство выполнения задач, которые стояли перед компьютерами на заре их развития — т. е. вычислений и пакетной обработки данных. (Это отражено и в том, как они назывались в то время — электронные вычислительные машины).

Но для моделирования процессоров нашего мира такая интерпретация (т. е. основанная на чётких последовательных утверждениях) подходит не в полной мере. Гораздо логичнее рассматривать систему как набор объектов, которые «живут своей жизнью» и тем или иным способом реагируют на события (сообщения) внешнего мира. Иногда, правда, они могут сами послать сообщение, чтобы «попросить» другой объект сделать что-то такое, что они не в состоянии.



И именно эта идея, т. е. разделение моделируемой системы на четко выраженные объекты, принимающие и посылающие сообщения, знающие, как делать то, для чего они предназначены, и как запросить другие объекты сделать что-то выходящее за эти рамки, привела к созданию объектноориентированного анализа и разработки, к созданию объектноориентированных языков программирования.

Понятие программных объектов было впервые введено Оле-Джоаном Далем, Бьорном Мюрхогом и Кристеном Ныгардом из Норвежского вычислительного центра в Осло в языке Simula-67, который создан по мотивам языка Algol-60. Simula-67 являлся языком моделирования и описания сложных систем.

Однако по-настоящему широкое внедрение этой идеи произошло при разработке языка SmallTalk в 1970 г. Аланом Кейем в Исследовательском центре фирмы Xerox в Пало-Альто. SmallTalk состоит лишь из объектноориентированных конструкций. Подобный подход позволяет сделать описания систем компактными и универсальными, но при использовании только такого механизма решение задачи не всегда получается эффективным.

Объект состоит из структур данных и алгоритмов. Каждому объекту известно, как выполнять операции над его собственными данными, но для остальных объектов он может представлять собой «чёрный ящик». Разные объекты будут использовать различные алгоритмы для осуществления операций, определяемых одним и тем же ключевым словом. Так, объект, чьи данные состоят из комплексных чисел, и объект, чьи данные — целые числа, будут пользоваться различными мето-

дами для выполнения арифметических операций.

Формально описание объекта и его свойств называют *классом*. Все объекты, принадлежащие к одному и тому же классу, имеют одинаковую структуру и свойства, но могут отличаться значениями своих атрибутов: все джинсы Levi's 501-й модели сшиты из денима, имеют две штанины, пять карманов и застёгиваются на болты, но могут быть разных размеров, поношенными или абсолютно новыми.

В настоящее время наибольшей популярностью обладают следующие объектно-ориентированные языки программирования: C++, Python, Java, C#. Объектно-ориентированная модель программирования характеризуется следующими понятиями и методами.

Абстрагирование. При построении сложной модели необходимо выделить самые важные детали предметов, явлений и отвлечься от несущественных (хотя эти несущественные детали могут, наоборот, стать существенными при построении других моделей). Такой процесс называется абстрагированием. Если для примера взять систему автоматизации работы библиотеки, то основными действующими лицами этой системы (объектами) окажутся:

Читатель
-Фамилия -Имя -Паспорт № -Контактный телефон -Список книг

Книга
-Автор -Название -Внутренний № -ISBN -Количество страниц -Цвет обложки -Количество экземпляров -Местонахождение

Стеллаж
-Этаж -Номер комнаты -Где находится

- книга;
- стеллаж, где хранятся книги;
- библиотекарь, который может эти книги выдать читателю;
- читатель.

Необходимо абстрагироваться от несущественных деталей и выделить только значимые свойства каждого объекта. Цвет волос и куртки читателя, возраст библиотекаря, материал, из которого сделан стеллаж, скорее всего, не существенны, а вот цвет обложки книги и её толщина, наверное, существенны, потому что читатель может попросить «толстый учебник

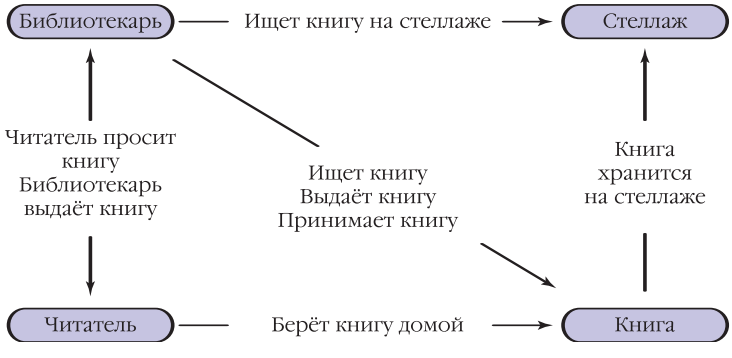
ИНТЕРФЕЙС И РЕАЛИЗАЦИЯ

Интерфейс:

```
class Book
{
   String Author();     /* Функция, возвращающая имя
                        автора */
   String Name();       /* Функция, возвращающая название
                         книги */
   String Code();       /* Функция, возвращающая код
                        книги */
   String ISBN();       /* Функция, возвращающая ISBN
                        книги */
   Integer NoPages();   /*Функция, возвращающая количест-
                        во страниц */
   Color CoverColor();  /* Функция, возвращающая цвет
                        обложки */
   Integer NoCopies();  /* Функция, возвращающая
                        количество экземпляров книги
                        данного вида в библиотеке */
   Location Location(); /* Функция, возвращающая
                        местонахождение данной книги
                        в библиотеке */
}
```

Реализация:

```
/* Функция получения имени автора книги */
String Book::Author()
{
   /* Установить соединение с базой данных */
   DB db = GetDatabaseConnection();
   /* Сделать запрос к базе данных */
   RS rs = db.Execute( "SELECT Author FROM Books WHERE
         Book_Id = %1", m_Id );
   /* Вернуть полученное имя автора */
   return rs("Author");
}
```



немецкого языка, с коричневой обложкой, который написан тремя авторами».

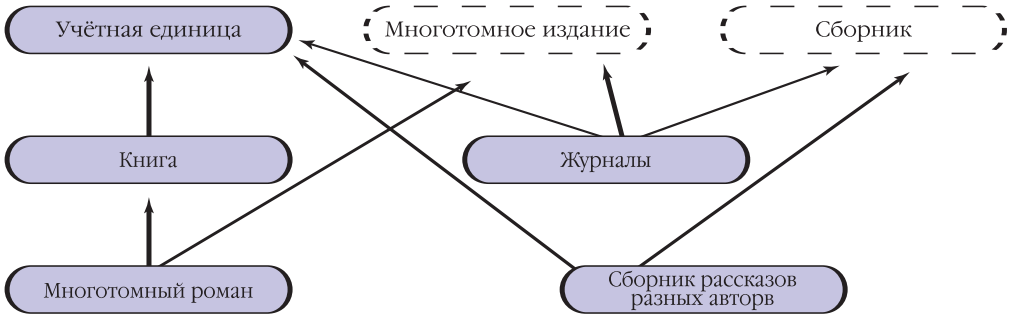
К свойствам книги относятся её название, авторы, внутренний библиотечный учётный номер и специальный международный учётный номер ISBN, количество страниц, тип и цвет переплёта. Читатель же характеризуется именем, фамилией, паспортными данными, контактным телефоном.

Инкапсуляция. Хотя объект «книга» и описывается как список её свойств, это не означает, что она представляет собой просто некоторую структуру данных, набор строк (подобно конструкции языка C `struct { ... }` или `record` — как в Pascal).

Пользователю объект «книга» предоставляет список запросов, т. е. свойства, которые она реализует, и то, в каком виде реализованы эти свойства. Например, запрос «автор» выводит символьную строку с полным именем автора, а запрос «количество экземпляров» — число. Это и называют *интерфейсом* объекта.

Для того чтобы обращаться к объекту, необходимо знать только его интерфейс, внутреннее же его устройство «спрятано» при помощи *инкапсуляции,* т. е. сокрытия реализации отдельных деталей. (Для других объектов, общающихся с «книгой», её внутреннее устройство совершенно неважно.) Инкапсуляция позволяет, в частности, полностью изменить реализацию объекта без каких-либо изменений в остальной программе. Можно легко и быстро совершенствовать работу программы, добавлять новые возможности. Например, чтобы хранить описание книг в базе данных при использовании объектно-ориентированного подхода, необходимо переписать только реализацию объекта «книга», а если бы применялся обычный процедурный подход, то, скорее всего, нужно было бы переписать бóльшую часть программы.

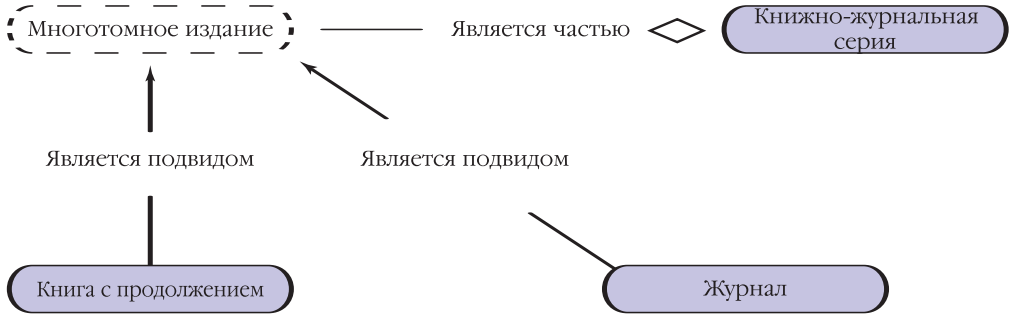
Иерархия. Использование абстрагирования и инкапсуляции было возможным ещё в структурных языках «старшего» поколения, таких, как Pascal или C. Но практически во всех случаях, кроме самых простых, чис-



ло абстракций в системе получается очень большим.

В библиотеке «книга» в действительности является родовым понятием. Это может быть книга одного автора, сборник рассказов, журнал, подшивка газет. Книги могут состоять из нескольких томов, иметь продолжение. Если каждый из таких объектов описать как отдельный независимый класс, то число абстракций в системе будет очень большим. Более того, придётся отдельно учитывать выданные читателю журналы, отдельно — книги и т. д. А если создать всего один объект «учётная единица», то этот класс будет чересчур сложным. То есть потеряется одно из основных предназначений объектно-ориентированного подхода: уменьшение сложности путём разложения целого на фрагменты.

В этом случае используется *иерархическая* структура абстракции — *подвид*. Определим абстракцию «учётная



ИСПОЛЬЗОВАНИЕ ОБЪЕКТОВ-НАСЛЕДНИКОВ КАК РОДИТЕЛЬСКИХ ОБЪЕКТОВ

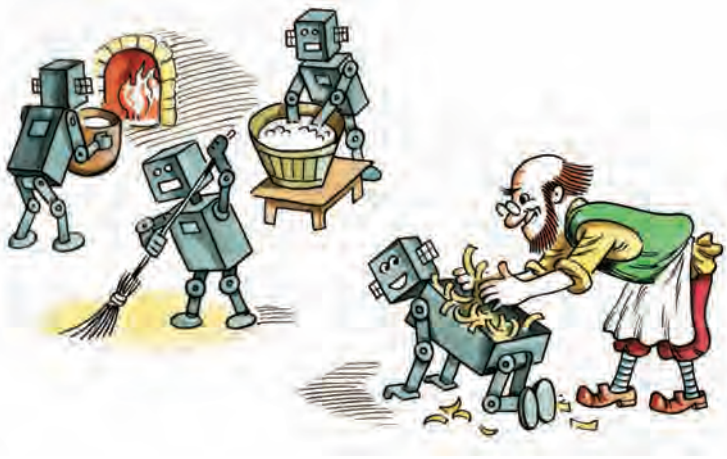
```
/* Функция, которая возвращает список "просроченных" читателем книг */
BiblioItems Reader::CheckDue()
{
   /* Получить список всех книг, находящихся на руках у читателя */
   BiblioItems bi = GetItems();
   /* Создать объект, в который будет помещён список книг,
   для которых срок возврата прошёл */
   BiblioItems due;

   /* Выполнить цикл для всех книг, находящихся на руках
   у читателя */
   for( bi.First(); ! bi.Last(); bi.GetNext() )
   {
      /* Если дата возврата книги меньше, чем сегодняшняя дата, значит, эту книгу читатель не вернул
      вовремя и её необходимо добавить к списку "просроченных" книг */
      if( bi.Current().GetReturnDate() < Today() )
      due.Add( bi.Current() );
   }

   return due;
}
```

В этой функции внутри списка книг, находящихся на руках у читателя (`bi`), могут встречаться объекты типа «книга», «журнал», «сборник» и пр. Но поскольку в базовом для них классе «учётная единица» объявлена функция для получения даты возврата `GetReturnDate()`, то эту функцию можно вызывать и для объектов-наследников.

В противном случае пришлось бы хранить несколько списков, а именно список взятых читателем книг, список взятых читателем журналов и т. п.



единица» как базовый класс, описывающий объекты, которые хранятся в библиотеке и могут быть выданы читателю. «Книга» или «журнал» будет подвидом учётной единицы.

В классе «журнал» необходимо ввести дополнительные свойства, такие, как год и месяц выхода, список авторов и статей. Но «книга» тоже может представлять собой сборник произведений различных авторов.

Если создать отдельный объект для «книги» — сборника рассказов, то получится, что один и тот же код будет использован в совершенно разных местах, окажется затруднена его реализация, поддержка. Поэтому лучше установить ещё одно иерархическое соотношение: «журнал» и «сборник рассказов» являются подвидом абстрактного «сборника».

Для отчётности и поиска иногда бывает полезным рассматривать годовые подшивки журналов и многотомные романы как единое целое. Тогда появляется ещё один вид иерархии: отношение «является частью». Многотомное издание является частью общей серии. Иногда говорят, что отношение «является подвидом» — это иерархия классов, а отношение «является частью» — это иерархия объектов. Классы, которые являются подвидами более общего класса, — это *наследники* или *потомки,* а более общий класс — это *класс-прародитель* или просто *родительский* класс.

С отношением «является подвидом» связано ещё одно очень важное свойство объектно-ориентированного программирования.

Полиморфизм. Для объекта «читатель» необходимо хранить список выданных на руки книг, журналов, газет. Из созданной структуры видно, что хранить надо список абстрактных «учётных единиц». Поскольку классы «книга», «журнал» являются подвидами учётных единиц, то объекты, описывающие реальные книги, имеют те же свойства и допускают те же операции, что и абстрактные «учётные единицы». Для них можно установить ISBN, внутренний номер, найти местоположение на стеллаже.

Для каждого из учётных объектов необходимо получить полное название.

Для книги это будет автор и название произведения, для журнала — название журнала, номер и год выпуска. Определение различных функций (например, GetMagazineDescription, GetBookDescription) для различных классов нарушит иерархическую структуру, сведёт практически на нет так тщательно выстроенную иерархию.

После функции (например, GetDescription( )) в базовом классе «учётная единица» и переопределения её в остальных классах в классе «книга» описание будет состоять из имени автора и названия произведения, для многотомного романа, наверное, добавится ещё и порядковый номер, для сборника — имя составителя, для журнала — номер и год выпуска. Если каждый из таких специализированных объектов будет использоваться как представитель родительского класса «учётная единица», то всегда доступной станет эта функция (GetDescription). Причём вызывающий объект необязательно будет знать, к какому подтипу принадлежит используемый объект. Объектно-ориентированный язык программирования сам позаботится обо всём и вызовет соответствующую функцию подкласса.

Такое поведение называют *полиморфизмом:* для различных классов объектов имеется общий набор функций, определённый в их родительском классе. Эти функции могут быть переопределены в классах-наследниках, и для каждого объекта вызывается соответствующая функция независимо

ПОЛИМОРФИЗМ

```
BiblioItems Reader::RemindDue()
{
  /* Получить список всех книг, для которых подошёл срок возврата */
  BiblioItems due = GetDue();
  /* Создать новое электронное письмо и адресовать его читателю */
  MailMessage mail( Reader.EMailAddress );
  /* Добавить в текст письма вежливое напоминание */
  mail.addText( "Напоминаем вам, что подошёл срок вернуть следующие
  книги:" );

  /* Выполнить цикл для всех книг, подлежащих возврату */
  for( due.First(); ! due.Last(); due.GetNext() )
  {
    /* Добавить в текст письма подробное название книги */
    mail.addText( due.Current().GetDescription() );
  }

  /* Отослать письмо читателю */
  mail.Send();
}
```

В этой функции при формировании текста письма просматривается список всех книг, для которых подошёл срок возврата. Для каждого объекта из списка вызывается функция получения подробного описания (имя автора, название, быть может, ISBN) `GetDescription()`. Но в отличие от предыдущего примера здесь эта функция уже по-разному определена в различных подклассах «учётная единица».

Используемая объектно-ориентированная система программирования сама позаботится о том, чтобы для каждого объекта вызвать правильную функцию.

Бьёрн Страуструп.

от того, обращаются к нему как к объекту данного класса или класса-прародителя.

Типизация. Для создания эффективно работающей, правильной программы необходимо как можно раньше проверять все операции на допустимость. Это желательно делать на этапе создания программы, но никак не во время её выполнения.

При создании объектов появляются новые типы данных, описывающие их структуру и допустимые операции. В языках программирования встроен специальный механизм, который отслеживает использование объектов различных типов данных и не позволяет выполнять операции, предусмотренные для одного типа данных, над объектами другого типа.

Например, если система пошлёт электронное письмо не читателю, а «книге», ничего особенно страшного не случится, письмо, скорее всего, просто затеряется в компьютерных недрах. Но вот если в список взятых читателем книг попадёт «стеллаж» или другой «читатель», то это уже будет неприятно. Наиболее чувствительные натуры могут упасть в обморок, получив письмо с просьбой срочно вернуть в библиотеку целый стеллаж книг.

Чтобы такого не происходило, в объектно-ориентированных языках программирования имеется механизм *типизации,* или способ внутренней защиты от использования объектов одного класса вместо других. Этот механизм срабатывает либо на этапе написания программы (тогда уже в обморок от усталости будет падать программист, который пытается найти ошибку), либо в момент выполнения (программа остановится, но ужасное письмо всё-таки не будет послано читателю).

ОБЪЕКТНО-ОРИЕНТИРОВАННЫЕ ЯЗЫКИ ПРОГРАММИРОВАНИЯ

SmallTalk. В версиях SmallTalk-72 и SmallTalk-74 были заложены основы объектно-ориентированного программирования, но не было ещё механизма наследования классов (такой язык — язык с классами, но без наследования — можно назвать объектным языком программирования в противовес объектно-ориентированному).

SmallTalk представлял собой не только язык программирования, но и целую среду разработки с редактором текста, средствами для просмотра иерархии класса, библиотеками графического пользовательского интерфейса. Его версия SmallTalk-80 получила широкое признание и была перенесена на многие компьютерные платформы.

Object Pascal. Язык программирования Object Pascal был создан сотрудниками фирмы Apple Computers совместно с Никлаусом Виртом, автором процедурного языка Pascal.

Язык Object Pascal был включён в 1986 г. в среду разработки MPW (Macintosh Programmer's Workshop) для компьютера Macintosh. Этот язык явился первым объектно-ориентированным языком разработки для персонального компьютера. Он не только предоставлял программисту удобный инструментарий, все необходимые библиотеки, но и позволял воспользоваться всеми преимуществами объектно-ориентированного подхода без потери производительности программы.

В то же время для облегчения языка из него были убраны многие возможности, такие, как множественное наследование (класс «журнал» является наследником нескольких классов: «учётная единица», «многотомное издание» и «сборник»), глобальные функции и переменные класса. Например, общее количество книг в библиотеке можно определить как глобальную переменную класса «учётная единица». При создании нового объекта — закупке книг — она увеличивается, при удалении объекта — списании — уменьшается.

C++. На базе языков C и Simula Бьёрном Страуструпом, сотрудником AT&T Bell Laboratories, был создан язык C with Classes, т. е. «C с классами». Потом на его основе разработали настоящий объектно-ориентированный язык C++, который поддерживает оди-

ночное наследование, перегрузку методов, строгую типизацию, полиморфизм.

Выходило несколько стандартов C++, в каждом из которых язык на основании широкого опыта использования перестраивался, дополнялся необходимыми возможностями. Во второй версии C++ появилось множественное наследование, в третьей был проработан механизм реакции программы на ошибки, непредусмотренные события — механизм *исключения*. Этот механизм тоже строился на объектно-ориентированной основе, исключениями являлись лишь объекты специальных классов, выстроенных в иерархию, представляющую возможные типы ошибок.

Поскольку язык C++ создавался на основе, пожалуй, самого эффективного и популярного на тот момент языка программирования, он и сам получился эффективным по скорости выполнения программ (в отличие от внутренне красивого, но очень медленного SmallTalk). Похожесть синтаксиса языков C и C++ позволила программистам легко перейти на новый язык, постепенно осваивая его новые возможности.

В начале XXI в. язык C++ является одним из самых популярных языков программирования. Компиляторы этого языка и системы программирования (как коммерческие, так и бесплатные, с открытым кодом) существуют практически на всех компьютерных платформах.

Java. Язык Java (произносится «джава») зародился как специализированный язык для разработки программного обеспечения встроенных компьютеров.

В 90-х гг. XX в. сотрудник фирмы Sun Microsystems Патрик Нотон в ходе работы по поддержанию большого количества однотипных программ для различных устройств начал испытывать разочарование. Когда он уже собирался покинуть компанию и перейти на другую работу, ему предложили составить проект решения возникшей проблемы. Данный документ неожиданно нашёл поддержку не только у других инженеров компании, но и у её руководства. Сразу была создана группа по разработке специального объектно-ориентированного языка для встроенных систем.

Название языка C# произносится как «си шарп», # — sharp, английское название музыкального символа «диез», повышающего ноту на полтона. В самом названии скрыта хитрая игра слов. С одной стороны, sharp обозначает «острый», «проницательный», «ловко», «искусно», т. е. положительную характеристику нового языка. С другой стороны, символом «+» обозначается «бемоль», т. е. знак, понижающий ноту на полтона. Это значит, что язык C# выше, лучше языка C++.

За основу нового языка программирования был взят самый популярный на тот момент объектно-ориентированный язык — C++. Но полностью использовать его во встраиваемых системах было невозможно из-за ограничений на доступные машинные резервы, а именно память и мощность процессора. Поэтому его синтаксис был переструктурирован, упрощён и сделан более строгим.

Другой особенностью нового языка стало то, что он компилировался не в машинный код для каждого конкретного устройства, а в специальный

промежуточный код для определённой машины. Во встраиваемых системах существовала специальная среда выполнения, которая обеспечивала функциональность этой машины и позволяла воспринимать полученный код без перекомпиляции.

Для переноса получившейся программы на другую компьютерную платформу (конечно, такую, на которой существует соответствующая среда выполнения) её не надо перекомпилировать, поскольку двоичный код не зависит от процессора и будет выполняться специальным интерпретатором.

После создания языка разработки Java и среды выполнения— виртуальной машины Java предполагали, что её основным применением будут Java-апплеты на страничках Интернета, которые эти странички помогут сделать интерактивными, да и просто оживят. Но, как это часто бывает с творениями рук человеческих, творцы не всегда могут предугадать судьбу своего детища. Так случилось и на этот раз.

Сейчас Java используется в двух основных областях. Во-первых это большие компьютерные системы для управления производством, финансами, персоналом и т. п.

Java позволяет сделать переносимое решение, не только не зависящее от конкретного компьютерного оборудования, но даже и такое решение, отдельные части которого могут выполняться на различных системах с сильно отличающейся архитектурой. Кроме того, выполнение программы не напрямую на процессоре, а на виртуальной машине позволяет значительным образом повысить надёжность системы в целом. В самом деле, фатальная ошибка программы будет отловлена виртуальной машиной. И это не только не помешает выполнению других задач предприятия, но, вполне возможно, что система сможет предпринять такие заранее предусмотренные действия, которые позволят продолжить выполнение «захворавшей» программы.

Со второй же областью, мы думаем, сталкивались все из вас. Это разработка программ для встраиваемых компьютеров и особенно игр для мобильных телефонов. Достаточно в телефоне реализовать поддержку Java, и он сразу превращается не просто в средство для связи, а в маленький мобильный центр развлечений.

C#. Язык C# был создан фирмой Microsoft в 2000—2001 гг. как основной язык программирования для новой платформы Microsoft.NET. Для облегчения процесса перехода на новый язык его синтаксис был большей частью заимствован из самых популярных языков: C++ и Java. Некоторые даже называют его точной копией Java, созданной из коммерческих соображений, но это не совсем так. C# настолько же похож на Java, насколько Java похож на C++ — помимо явного сходства синтаксиса есть и достаточно большие различия.

Например, в языке Java даже самый простой тип, число или символ всегда являются объектом соответствующего класса, что значительно снижает производительность программ. В языке C# число может быть и объектом класса, и обычной переменной, как в C или C++.

ОБЪЕКТНО-ОРИЕНТИРОВАННАЯ ПЛАТФОРМА MICROSOFT.NET

Как и Java, язык C# компилируется в промежуточный код, единый для различных машин. Но в отличие от Java этот промежуточный код псевдо-Ассемблер сам является объектно-ориентированным. Иными словами, при компиляции не теряется информация об объектной структуре программы. Этот промежуточный код, MSIL (MicroSoft Intermediate Language), и является одним из основных «строительных блоков» новой платформы Microsoft.NET.

Такое нестандартное решение позволяет создавать большие объектно-ориентированные системы, в которых разные части написаны на разных языках, и при этом обеспечивается лёгкость их взаимодействия и возмож-

ность наследования. То есть класс-родитель может находиться в модуле, реализованном на C#, а класс-потомок — в модуле на любом другом языке, который поддерживает компиляцию в MSIL, например Eiffel, Python, SmallTalk, Pascal и др. Точно так же не возникнет проблемы и при обращении одного объекта к сервисам другого, несмотря на различие программных средств.

Пока платформа Microsoft.NET ещё очень молода, поэтому сложно давать точные прогнозы. Но вполне вероятно, что именно она сможет наконец решить обычную проблему программирования, а именно невозможность повторного использования уже созданного кода.

С момента появления объектно-ориентированного программирования одной из главных задач было повторное использование уже написанного кода. Тысячам программистов приходится заново писать практически один и тот же код, поскольку прямое использование уже готового невозможно по тем или иным техническим причинам (например, различие в языках программирования или даже в диалектах одного и того же языка).

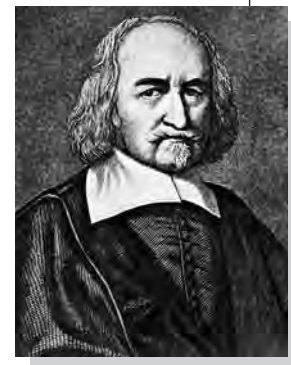
Готовые фрагменты кода бывают оформлены как внешняя системная библиотека, что позволяет использовать части программ, написанные на разных языках программирования, но на одной платформе. Однако при этом теряется объектная ориентированность, ведь библиотека выдаёт «наружу» только набор функций, полностью скрывая своё внутреннее устройство.

В платформе Microsoft.NET появляются специальные библиотеки кода, называемые *сборками*. В этих сборках могут храниться фрагменты программ, отдельные модули, предоставляющие те или иные сервисы. Благодаря использованию MSIL эти сборки позволяют повторно применять программные компоненты в последующих разработках, поскольку объектная ориентированность самого этого промежуточного языка обеспечивает возможность наследования классов, описанных в них.

ЯЗЫКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

На протяжении многих веков человек производил действия над числами и одновременно совершенствовал форму их записи, методы вычислений, придумывал всё более сложные технические устройства для облегчения этих действий. Но лишь немногие задумывались над тем, что вычисления можно понимать более широко. Например, великий английский философ Томас Гоббс (1588—1679) писал в своём сочинении «Elementa philosophiae»: «Не следует думать, что вычисления могут проводиться только с числами». В XIX в. Ада Лавлейс говорила: «Машина может упорядочивать и комбинировать числовые значения так же, как и буквы или другие символы общего характера. В сущности, при выполнении соответствующих условий она могла бы выдавать результаты и в алгебраическом виде». Однако до появления первых компьютеров такие идеи всё-таки были достаточно далеки от практического воплощения. Да и компьютеры сначала использовались в основном для действий с числами, хотя Конрад Цузе уже в 1945 г. написал программу для игры в шахматы. Очень скоро учёные всерьёз задумались над возможностью применения компьютеров для обработки нечисловых данных. Речь шла в первую очередь о задачах из области *искусственного интеллекта* — к ним относятся моделирование процессов, протекающих в сознании человека, автоматическое доказательство теорем, планирование действий робота и некоторые другие. Для их решения характерно использование *символьных вычислений* (когда обрабатываются не числа, а символы).

Ранние языки программирования, такие, как FORTRAN или Algol-60, не были приспособлены для решения



Томас Гоббс.

Конрад Цузе.

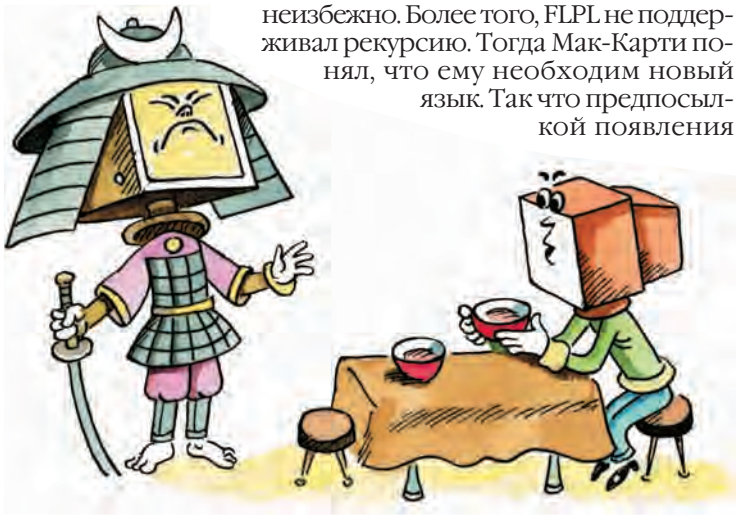
Джон Мак-Карти.

задач символьной обработки. Попытки же встроить в них нужные свойства (например, корпорация IBM разработала в 1958 г. расширение языка FORTRAN — язык FLPL) не были удачными. Поэтому для решения задач искусственного интеллекта стали разрабатываться специализированные языки.

LISP

Первым таким языком стал LISP (от *англ.* List processing — «обработка списков»), появившийся в 1959 г. Его автором был выдающийся американский учёный Джон Мак-Карти, внёсший огромный вклад в развитие информатики. Кстати, термин «искусственный интеллект» впервые использовал именно он. Первоначально Мак-Карти не имел намерения разработать именно язык программирования. Его целью было создание новой модели вычислений, более удобной, чем машины Тьюринга.

В процессе работы выяснилось, что основой такой модели может стать рекурсивная обработка символьных списков. Используя её, Мак-Карти сумел найти весьма элегантный алгоритм решения задачи символьного дифференцирования алгебраических выражений. Однако ему не хотелось утяжелять алгоритм сугубо техническими деталями, такими, как явное введение операторов динамического размещения и удаления списков из памяти, но при использовании языка FLPL это было неизбежно. Более того, FLPL не поддерживал рекурсию. Тогда Мак-Карти понял, что ему необходим новый язык. Так что предпосылкой появления первой версии языка LISP (его ещё называют «чистый LISP») явилась потребность в удобных средствах обработки списков.



Дальнейшее изучение новой модели вычислений привело Мак-Карти к мысли построить универсальную функцию языка LISP, способную вычислять любую функцию этого языка (подобно универсальной машине Тьюринга, имитирующей работу любой другой машины Тьюринга). Такая функция (она получила название EVAL) была разработана и использована в качестве интерпретатора языка LISP 1.5, описание которого было опубликовано в 1965 г. Именно с этого момента начинается широкое распространение языка. Он молниеносно завоёвывает популярность — на новом языке пишут программы почти все исследователи, работающие в области искусственного интеллекта. Язык совершенствуется, появляется множество его версий и диалектов (среди них можно особо отметить интересный язык Scheme, 1975 г.), но одновременно возникает проблема переносимости программ. Для её решения в 1984 г. был выработан единый стандарт языка Common LISP.

LISP относится к классу *функциональных* языков программирования. В отличие от обычных *(императивных)* процедурных языков они не используют ни переменные, ни операторы присваивания. В них отсутствуют операторы цикла, а повторение действий реализуется с помощью рекурсии. Программа в таких языках представляет описание функций, а выполнение программы — процесс применения функций к аргументам. Правда, современный LISP нельзя назвать функциональным языком в полном смысле этого слова (им был только чистый LISP). Он обладает и некоторыми свойствами, присущими процедурным языкам.

В языке LISP используются лишь два типа данных — *атомы* и *списки*. Атомы очень похожи на обычные идентификаторы в традиционных языках; кроме того, атомами считаются и числовые константы. Списками называют заключённые в скобки последовательности элементов — ими

могут быть как атомы, так и другие списки. Например, список L = (A (B) C (D E F) (G H)) состоит из пяти элементов. Первый и третий из них — это атомы A и C, а остальные являются списками. Обратите внимание, что второй элемент, (B), — это не атом, а список, состоящий из одного элемента, атома B.

LISP содержит набор элементарных (или системных) функций, основными среди них являются (CAR L), (CDR L) и (CONS M L). Функция CAR выбирает из списка L его первый элемент («голову»), функция CDR удаляет из списка L его «голову» (оставляя «хвост»). В нашем примере (CAR L) = A, (CDR L) = ((B) C (D E F) (G H)). Эти две функции «разбирают» список на части, а вот функция CONS создаёт новый список, помещая элемент M в «голову» списка L: (CONS M (B C D)) = = (M B C D).

Указанными тремя функциями набор элементарных функций, разумеется, не исчерпывается. Кроме того, язык предоставляет удобные средства, с помощью которых из элементарных функций можно строить более сложные.

В языке LISP принята единая форма записи данных и программ. Например, выражение (F X Y) может восприниматься либо как список из трёх элементов, либо как обращение к функции F с аргументами X и Y. Благодаря этому программа пользователя способна в процессе работы не только создавать новые функции и программы, но и немедленно выполнять их. Эта возможность языка является его уникальной особенностью (ведь в программе на императивном языке все операторы должны быть выписаны до начала её работы и в процессе работы не могут изменяться) и очень мощным оружием в руках программистов.

PROLOG

Альтернативным подходом к решению задач искусственного интеллекта стала концепция *логического программирования,* реализованная в языке Prolog (Programming in logic). Его разработали в начале 70-х гг. XX в. Ален Колмерье и Филипп Руссель из Марсельского университета и Роберт Ковальски из Эдинбургского университета. Они развивали свои идеи почти десять лет, но те не получали особой известности. Всё изменилось в 1981 г., когда появился так называемый «японский вызов» — масштабный проект правительства этой страны, связанный с созданием ЭВМ пятого поколения. Резкое повышение интеллектуального уровня будущих компьютеров и возможность общения с ними на языке, максимально приближённом к естественному, были важнейшими целями проекта. Язык Prolog стал основой разработок, что привлекло к нему всеобщее внимание и вызвало бурный рост его популярности во всём мире.

LISP более 20 лет безраздельно господствовал в области искусственного интеллекта. Но и сегодня, несмотря на появление многочисленных конкурентов, он по-прежнему является самым распространённым языком для решения задач в этой области.

Что же такое логическое программирование? В императивных и функциональных языках программист знает, что должна выполнить программа. Его задача — дать указание компьютеру, как именно надо выполнить вычисления. Логическое программирование коренным образом отличается и от императивного, и от функционального. Программа на языке Prolog не содержит указаний, как вычислять результат. Она лишь описывает требуемый вид результата. Получение его — это дело заложенных в язык механизмов логического вывода (основанных на так называемом *методе резолюций*). Программа на языке Prolog состоит из операторов. Операторы делятся на *факты* и *правила*. Факт — это некое безусловное утверждение, которое предполагается истинным. Приведём примеры фактов, записанных в виде операторов языка:

```
слон (джумбо).
лягушка (флик).
сильнее (джумбо, флик).
```

Это означает, что объект `джумбо` является слоном, объект `флик` — лягушкой, а также что `джумбо` сильнее `флика`.

Более сложен другой тип операторов. На языке Prolog правила записываются в таком виде: Y: — X, Z... Эта запись определяет некоторое условие следования и читается так: «справедливо Y, если справедливы X и Z, и...».

Prolog используется для решения самых разных задач. Наиболее важное из его применений — разработка экспертных систем.

Следовательно, запятая в правой части правила понимается как логическое И (конъюнкция). Правило

```
сильнее (джумбо, флик) :– слон (джумбо), лягушка (флик).
```

означает, что объект `джумбо` сильнее объекта `флик`, если `джумбо` — это `слон` и `флик` — `лягушка`. Иными словами, правила используются для описания логических отношений между фактами.

Цель работы программиста на языке Prolog — создать совокупность фактов и правил, полностью описывающих некоторую ситуацию или предметную область. После этого Prolog-системе можно задавать (например, вводя с клавиатуры) вопросы в виде предложений, называемых *целью*. Prolog-система должна проверить цель и либо подтвердить, либо опровергнуть её.

Принципиальная особенность программы на языке Prolog — то, что результатом её работы может стать получение новых фактов, неизвестных системе (а может быть, неизвестных и самому программисту!). Именно это свойство языка обеспечивает ему внимание исследователей и разработчиков систем искусственного интеллекта.

Большие надежды связывают с языком Prolog (и вообще с логическим программированием) создатели систем управления базами данных (СУБД) и систем обработки текстов на естественных языках, а также другие специалисты. Многие даже полагают, что именно Prolog в будущем станет основным языком при разработке более простого и более надёжного программного обеспечения. А вот в том, что именно с языка Prolog необходимо начинать обучение основам программирования, некоторые исследователи и педагоги убеждены уже сегодня.

Карл Хьюитт.

В то же время существует мнение, что Prolog до сих пор всего лишь эксперимент, хотя и грандиозный. Дело в том, что при использовании языка возникает немало проблем, которые пока ещё не имеют удовлетворительного решения. Основная из этих проблем — повышение эффективности выполнения программ. Да и в целом методы разработки программ на языках логического программирования ещё до конца не сформировались.

PLANNER

LISP и Prolog — два самых популярных языка символьной обработки. Однако интересные идеи предлагались и создателями других языков, не получивших широкой известности. Один из таких языков — редко вспоминаемый сейчас Planner. Своё название (которое переводится как «планировщик») он получил благодаря тому, что был в основном предназначен для создания систем планирования действий роботов. Planner был разработан Карлом Хьюиттом из Массачусетсского технологического ин-

ститута в 1971 г. и стал серьёзным этапом в развитии идей обработки символьной информации.

Planner обладает всеми возможностями языка LISP, т. е. фактически LISP является его подмножеством. Но набор встроенных функций в языке Planner намного шире, а сами они обычно гораздо мощнее соответствующих функций языка LISP. К примеру, функции ELEM и REST обобщают функции CAR и CDR, позволяя выделять и отбрасывать из списка произвольные элементы: [ELEM 1 L] = = A, [REST 2 L] = (C (D E F) (G H)). Кроме того, введение в запись нескольких типов скобок (в языке LISP допускаются только круглые скобки, что весьма затрудняет чтение программ) сделало программы гораздо понятнее. Но самое главное, язык был обогащён дополнительными возможностями.

Первая из них — поиск и анализ данных по образцу. Подобный механизм был успешно опробован в языке Snobol и широко используется сейчас — например, когда в Norton Commander задаётся поиск всех файлов вида *.doc, то выполняется именно это действие.

Вторая дополнительная возможность позволяет весьма эффективно применять Planner при создании систем искусственного интеллекта: можно строить и изменять описание среды, в которой решается задача. Описание среды называется *базой данных* и содержит отдельные утверждения (факты), истинные в данной среде. Кроме фактов, содержащихся в базе данных, в программе можно задавать логические отношения между используемыми в задаче понятиями, а также описания действий, которые разрешено производить. Такие описания называются *теоремами*. Каждая теорема имеет список условий (предпосылок), которые должны быть выполнены, прежде чем станет возможным её применение. В результате применения теоремы некоторые факты из базы данных перестают быть истинными и должны быть из неё удалены, а другие, напротив, становятся истинными и должны её пополнить. Списки фактов обоих видов также содержатся в описаниях теорем.

Самым важным нововведением языка Planner стал *режим возвратов*. Он позволяет во время работы программы отказываться от принятых ранее решений, если оказывается, что они не приведут к цели. Для искусственного интеллекта вообще характерно использование перебора вариантов во время решения задач. Встроенный в язык Planner режим возвратов (его ещё называют механизмом поиска с возвратом, или, по-английски, backtracking) избавляет программиста от необходимости самому реализовывать такой перебор.

Работы в области искусственного интеллекта ведутся уже около 50 лет. Первоначальный энтузиазм, ожидание быстрого и решительного успеха в создании универсального искусственного разума давно уступили место кропотливой работе по автоматизации решения отдельных задач, традиционно предполагающих участие мыслительных способностей человека. Роль языков программирования в построении интеллектуальных систем крайне велика. Современные языки искусственного интеллекта позволяют успешно решать многие проблемы, однако ограниченность их возможностей очевидна. Так что очень трудно сказать, какие языки будут использоваться в будущем — функциональные, похожие на LISP, или логические, похожие на Prolog. А может быть, появятся новые языки, основанные на совершенно новых идеях, обладающие такими свойствами, о которых сегодня никто не подозревает?

ПРОГРАММИСТЫ И ПРОГРАММИРОВАНИЕ

АВГУСТА АДА БАЙРОН, ЛЕДИ ЛАВЛЕЙС

Дочь, птенчик, Ада милая! На мать
Похожа ль ты, единственно родная?
В день той разлуки мне могла сиять
В твоих глазах надежда голубая...

Дж. Байрон

Джордж Гордон Байрон.

В июле 1980 г. в Министерстве обороны США был разработан язык программирования, получивший название Ada. Учебники по этому языку были переведены и в России. Название одной из книг — «Язык программирования Ада» — порождало негативные ассоциации. Однако к тёмным силам зла и ада этот язык не имеет никакого отношения. Он был назван в честь первого в мире программиста — Августы Ады Лавлейс.

Августа Ада Байрон родилась 10 декабря 1815 г. Спустя месяц её мать, Анна Изабелла Милбэнк (1792—1860), бросает своего мужа, Джорджа Ноэла Гордона Байрона (1788—1824), великого английского поэта-романтика, и возвращается в отчий дом. Семейный разлад Байрона и его жены находит громкий отклик в высшем свете и становится поводом для злобной травли поэта — он оказывается перед необходимостью покинуть родину. Отправляясь в Италию, с тем чтобы уже никогда не вернуться в Лондон, Байрон даже не предполагал, что оставляет в колыбели будущую легенду кибернетики.

В доме, где росла и воспитывалась Ада Байрон, царил закон: никакого упоминания об опальном поэте. Все его книги были исключены из семей-

ной библиотеки, и даже имя Августа Ада сократилось до Ады (Августой девочка была названа в честь сводной сестры Байрона).

Августа Ада была похожа на отца лицом, но пристрастия унаследовала материнские. Анна Изабелла Байрон, в лучшие дни своей семейной жизни получившая от мужа прозвище Королева Параллелограммов, увлекалась математикой. Аналитические способности она хотела развить и в дочери в противовес романтическим склонностям, которые вполне могли достаться девочке от отца.

Девочка превосходно играла на нескольких музыкальных инструментах и владела иностранными языками, но самым сильным её увлечением была математика. В этом Аду поддерживали и мать, и учёные друзья Анны Изабеллы — семья профессора де Морган и супруги Соммервиль (Мэри Соммервиль знаменита тем, что перевела с французского «Трактат о небесной механике» астронома Пьера Лапласа). В 13 лет Ада уже рисовала чертежи летательных аппаратов. Профессор Огастес де Морган был настолько высокого мнения о способностях своей ученицы, что даже сравнивал её с Марией Аньези (1718—1799), выдающейся итальянской женщиной-математиком. Одновременно девочка тайком писала стихи, стыдясь этого как какой-нибудь наследственной болезни. Свои поэтические наклонности она реализовала гораздо позднее.

Семейная жизнь Августы Ады сложилась счастливее, чем у её родителей. В июле 1835 г. она вышла замуж за Уильяма, 18-го лорда Кинга, ставшего впоследствии первым графом Лавлейсом. Он с одобрением относился к научным занятиям жены и во всём помогал ей. Супруги вели светский образ жизни, регулярно устраивая вечера и приёмы, на которых бывал «весь Лондон».

Один из завсегдатаев этих вечеров редактор популярного журнала «Экзаменатор» Олбани Фонбланк составил такой портрет хозяйки дома: «Она была ни на кого не похожа и обладала талантом не только поэтическим, но математическим и метафизическим... Наряду с совершенно мужской

Августа Ада Байрон.

способностью к пониманию, проявлявшейся в умении решительно и быстро схватывать суть дела в целом, леди Лавлейс обладала всеми прелестями утончённого женского характера. Её манера, её вкус, её образование, особенно музыкальное, в котором она достигла совершенства, были женственными в наиболее прекрасном смысле этого слова, и поверхностный наблюдатель никогда не угадал бы, сколько внутренней силы и знания скрыто под её женской грацией. В той же степени, в какой она не терпела легкомыслия и банальности, она получала удовольствие от истинно интеллектуального общества и потому энергично искала знакомства со всеми, кто был известен в науке, искусстве и литературе».

Однажды осенью на обеде у Мэри Соммервиль Ада впервые услышала об аналитической машине Чарлза Бэббиджа, профессора кафедры математики Кембриджского университета (см. статью «Чарлз Бэббидж»).

Огастес де Морган.

Чарлз Бэббидж.

Вскоре она была представлена знаменитому учёному. Вот как описывает в своих мемуарах миссис София де Морган первое посещение юной Адой мастерской Бэббиджа: «Пока часть гостей в изумлении глядела на это удивительное устройство, с таким чувством, как, говорят, дикари первый раз видят зеркальце или слышат выстрел из ружья, мисс Байрон, совсем ещё юная, смогла понять работу машины и оценила большое достоинство изобретения».

Общая увлечённость наукой связала Аду и Бэббиджа на долгие годы плодотворного сотрудничества.

В 1840 г. Бэббидж побывал с визитом в Турине, куда его пригласили для чтения лекций об аналитической машине. В Италии к нему отнеслись с большим пониманием, чем на родине, лекции имели шумный успех. Один из слушателей, молодой инженер Луиджи Менабреа, составил и опубликовал конспект этих лекций. Менабреа был глубоким мыслителем, он закончил свой очерк удивительными словами, которые стоило бы услышать тем, кто наделял и наделяет машины сверхчеловеческими способностями: «Машина — не мыслящее существо, это просто автомат, выполняющий заложенное в него».

«Спустя некоторое время после появления этой статьи, — писал Бэббидж в своих “Страницах о жизни философа”, — графиня Лавлейс сообщила мне, что она перевела очерк Менабреа. Я спросил, почему она не написала самостоятельной статьи по этому вопросу, с которым была так хорошо знакома. На это леди Лавлейс отвечала, что такая мысль не пришла ей в голову. Тогда я предложил, чтобы она добавила некоторые комментарии к очерку Менабреа. Эта идея была немедленно принята». План комментариев разрабатывался совместно с Бэббиджем, который коротко упомянул об этом в «Страницах...» фразой: «Мы обсуждали вместе различные иллюстрации, которые могли быть использованы; я предложил несколько, но выбор она сделала совершенно самостоятельно». В 1843 г. Ада Лавлейс перевела статью на английский, снабдив подробными комментариями (по объёму они превосходили основной текст). Кроме того, она привела ряд примеров практического использования машин и, выражаясь современным языком, составила программу вычисления чисел Бернулли по довольно сложному алгоритму. Именно Адой были предложены термины «рабочая ячейка» и «цикл».

Ознакомившись с её трудом, Бэббидж договорился с редактором солидного научного журнала «Учёные записки Тейлора» о публикации перевода статьи Менабреа и комментариев Ады к нему. В этой работе наилучшим образом освещены алгоритмические основы аналитической машины.

Ада Лавлейс перевела замыслы Бэббиджа на математический и «технологический» языки. Вот несколько

фрагментов: «Машину нельзя наделить разумом, она только реализует предложенные представления. Эти представления зафиксированы на перфокартах, они передаются различным механизмам, выполняющим последовательность действий...», «Весь интеллектуальный труд ограничен подготовкой необходимых для вычисления выражений...», «Машину можно рассматривать как настоящую фабрику чисел».

В то время как статья Менабреа касается в большей степени технической стороны дела, комментарии леди Лавлейс посвящены в основном математическим вопросам. Поэтому сама статья, в отличие от комментариев, представляет сейчас лишь исторический интерес, поскольку сегодняшние вычислительные машины построены на иных технических принципах, тогда как дополнения к ней заложили основы современного программирования.

Ада Лавлейс была человеком разносторонней одарённости, глубоких знаний и интересов — от математики и вычислительных машин до лошадей и музыки. Она предположила, что со временем аналитическая машина будет сочинять музыкальные произведения, рисовать картины и использоваться в практической и научной деятельности. Сейчас возможно оценить её правоту и точность прогнозов.

А вот другое предложение, которое сделала Ада Бэббиджу, чуть не погубило его научную карьеру. Леди Лавлейс была уверена, что машина уже может решать практические задачи, а именно прогнозировать беспроигрышные ставки на бегах. Однако то ли с машиной что-то было не в порядке, то ли с природой, но лошади упорно отказывались бегать по придуманной для них системе. Августа Ада проиграла не только свои деньги, но и деньги мужа, а от полного разорения семью, как это ни печально звучит, спасла только скоропостижная смерть Ады от рака.

Фамильный склеп Байронов в Ноттингемшире.

Она скончалась 27 ноября 1852 г. в возрасте 36 лет, как и Джордж Гордон Байрон. Согласно завещанию, она была похоронена рядом с отцом в фамильном склепе Байронов в Ноттингемпшире при Хакнеллской церкви.

Судьба отца, от влияния которого так хотела уберечь Аду мать, повторилась в судьбе его единственной законной дочери. Они были похожи: идеалисты с горящими глазами, готовые умереть за свободу чужой далёкой страны или пожертвовать всем ради изобретения, которого никто не принимает.

Во многих энциклопедиях Августа Ада Кинг Лавлейс фигурирует как английский математик. В качестве основного её научного труда указываются перевод статьи Менабреа «Элементы аналитической машины Бэббиджа» и комментарии к ней.

ПРОГРАММИСТЫ

Это одна из самых молодых и очень популярных профессий. *Программист* — человек, составляющий программы для ЭВМ.

В 60-х гг. XX в. в создании программ участвовали три человека: учёный, ставивший задачу и определявший необходимые данные для её решения; математик-алгоритмист, придумывавший алгоритм решения задачи и описывавший его на особом языке; программист, кодировавший

на машинном языке заданный алгоритм, пытаясь вместить большую программу в маленькую память медленно работающего компьютера.

Каждая программа являлась достижением человеческой мысли. За её эффективность отвечали автор алгоритма и программист. Математик должен был знать допустимый размер матрицы действительных чисел, при котором возможно производить обработку данных. Кодирование также требовало внимания — не меньшего, чем у авиадиспетчера.

С появлением языков программирования высокого уровня задачи программиста начали расширяться: теперь он сам придумывает алгоритмы и кодирует их на подходящем языке. Процесс создания эффективного алгоритма сохранил творческое начало. А вот кодирование становится рутинной операцией, которую можно перепоручить компьютеру.

Когда количество ЭВМ возросло, поменялся и круг задач, что вызвало острую нехватку программистов. В наши дни творческий подход и промежуточный результат удовлетворяют только запросы учёных, а финансистам и менеджерам необходимы готовые решения в минимальные сроки и с гарантированными показателями. Поэтому разработаны специальные технологии создания программных продуктов. Например, сначала система описывалась полностью, проект излагался в деталях, и лишь затем начиналась реализация системы. А вот документация по системе уж точно пишется в последнюю очередь. Правда, так возникают только сопровождающие программные продукты: редакторы, операционные системы и пр. Труд программиста, как правило, составляет 10—15 % общих трудозатрат по созданию и продвижению на рынок программной системы.

Среди менеджеров, художников, дизайнеров, технических писателей, тестеров и остального персонала, работающего над программной системой, одни программисты — истинные творцы. Наиболее удачные программные продукты выполнены и придуманы программистами для самих себя и уже во вторую очередь они полезны другим. Но этими творческими личностями надо как-то руководить.

Создание бригад главного программиста — одна из попыток организовать совместный труд нескольких программистов. В коллектив из 5—7 человек назначался более опытный главный руководитель-програм-

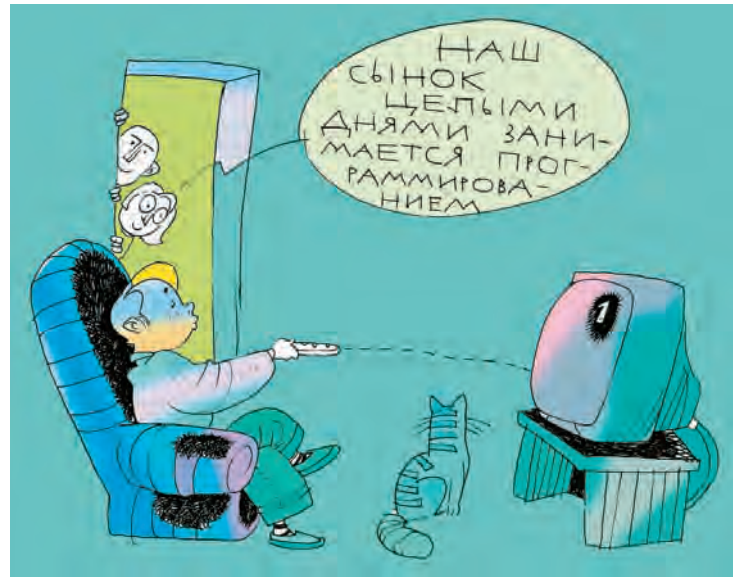
В одной из серий известной компьютерной игры «Larry» герой попадает в учреждение, где работают программисты. Они сидят за дисплеями и трудятся не поднимая головы. Между рядами столов ходят накачанные менеджеры с кнутами в руках и бьют по спине любого, кто оторвёт взгляд от экрана. Вероятно, в этой игре воплотилась тайная мечта менеджера программистской компании.

мист. Он определял ежедневные задания и проводил собрания, на которых программисты представляли фрагменты своих программ, их обсуждали, критиковали и утверждали. Результат работы такой группы действительно был гарантированным, но мало эффективным и дорогим. Порой один программист выполнял работу, предназначенную для всего коллектива.

Коллектив программистов похож на экипаж космического корабля. При создании важной программной системы решающим фактором эффективности его работы будет внутренняя психологическая атмосфера среди сотрудников.

Профессия программиста требует от человека определённых черт характера. Программисты, как правило, трудятся в одиночку, каждый имеет свой стиль работы. Подчас для реализации общего проекта эти стили совместить тяжело. Здесь всё зависит от профессионализма руководителя: правильно ли он распределил работу, удачно ли при этом подобрал программистов. Ведь просчёты в архитектуре системы могут привести к фатальным последствиям для всего проекта.

Однако не следует думать, что сегодня миллионы программистов испытывают муки творчества. Современные системы программирования позволяют в большинстве случаев осуществлять *параметрическую* настройку системы. Так, при создании базы данных может быть предусмотрено только представление отчётов



и заполнение полей информацией, при этом программист при выполнении этой работы не напишет ни строчки кода. Его по старинке называют программистом баз данных. Таких «программистов» сейчас большинство во всех областях деятельности, где используется компьютер. Однако, например, дома никто не называет себя программистом, когда «программирует» видеомагнитофон, независимо от того, что не все домашние могут справиться с этой операцией. И вряд ли назовёт себя хирургом медсестра, обрабатывающая неглубокую рану тому, кто не в состоянии сделать это сам.

РЕКУРСИЯ

Наиболее распространённый приём в программировании — обращение к ранее созданным и описанным программам, или, другими словами, вызовы подпрограмм (процедур или функций). Это естественно: если выполнение некоторой группы действий или вычисление одинаково определяемых значений должно осуществляться в программе неоднократно, то целесообразно один раз составить описание этих действий (вычислений), дать подпрограмме имя, а затем вызывать её столько раз, сколько это будет необходимо.

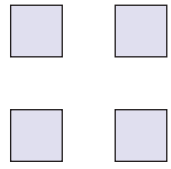
Действительно, если написать (на языке ЛОГО) процедуру КВАДРАТИК

```
ЭТО КВАДРАТИК
   ПОВТОРИ 4 [ВПЕРЁД 5 ВПРАВО 90]
КОНЕЦ
```

оставляющую на экране след в виде маленького квадрата со стороной 5,

Параметром процедуры называют переменную величину, используемую в теле процедуры, если её значение задаётся при вызове процедуры. Формальный параметр (или параметр описания) — это имя переменной, а фактический параметр (или параметр вызова) — это значение.

то с помощью другой процедуры — КВАДРАТЫ, которая вызывает при своём выполнении первую процедуру, можно изобразить, например, такую картинку:

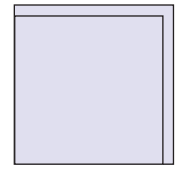


```
ЭТО КВАДРАТЫ
КВАДРАТИК
ПОВТОРИ 3
   [
     ПЕРОПОДНЯТЬ
     ВПЕРЁД 10
     ПЕРООПУСТИТЬ
     КВАДРАТИК
     ВПРАВО 90
   ]
КОНЕЦ
```

А вот пример *вызывающей программы* КВАДРАТИК_ПОШИРЕ; при этом роль *вызываемой программы* играет процедура *УГОЛ*:

```
ЭТО КВАДРАТИК_ПОШИРЕ
   ПОВТОРИ 4 [ВПЕРЁД 5 ВПРАВО 90]
   УГОЛ
КОНЕЦ
```

```
ЭТО УГОЛ
   ПОВТОРИ 4 [ВПЕРЁД 6 ВПРАВО 90]
КОНЕЦ
```



Процедура *УГОЛ* удивительно похожа на КВАДРАТИК и отличается только значением шага в команде ВПЕРЁД (из-за этого стороны квадратика в процедуре *УГОЛ* получаются на один шаг больше). Если же процедуру КВАДРАТ описать как процедуру с параметром, то исчезнет и это отличие:

```
ЭТО КВАДРАТ :K
   ПОВТОРИ 4 [ВПЕРЁД :K ВПРАВО 90]
КОНЕЦ
```

(В языке ЛОГО принято имя переменной величины предварять двоеточием без пробела.)

Теперь вызов программы КВАДРАТ 5 будет работать точно так же, как процедура КВАДРАТИК, а вызов КВАДРАТ 6 — как процедура *УГОЛ*.

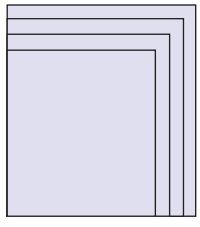
В примере, где программа КВАДРАТИК_ПОШИРЕ обращается к программе *УГОЛ*, нет ничего необычного: одна процедура вызывает другую, отличающуюся от первой. А что произойдёт, если заменить вызывающую и вызываемую программы на процедуру с параметрами КВАДРАТ *:K*

```
ЭТО КВАДРАТ :K
   ПОВТОРИ 4 [ВПЕРЁД :K ВПРАВО 90]
   КВАДРАТ :K + 1
КОНЕЦ
```

где значение параметра *K* = 5? Не парадокс ли: программа КВАДРАТ будет вызывать программу КВАДРАТ?

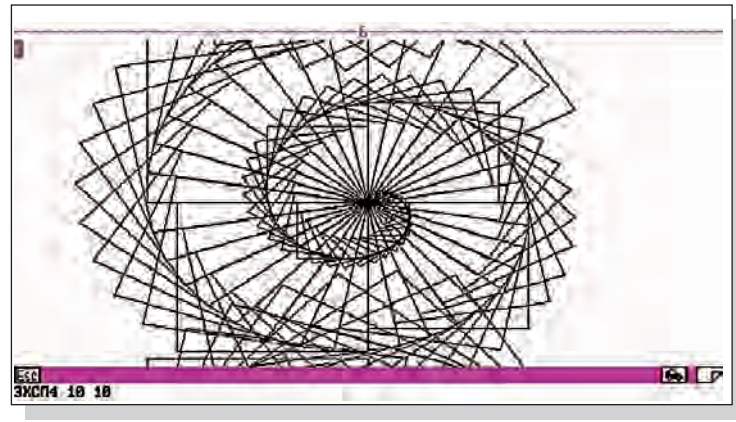
Парадокса нет. Ситуация, в которой процедура *A* вызывает процедуру *A* (процедуру с тем же именем, но, возможно, с другим значением фактического параметра), часто встречается в программировании и имеет специальное название — *«рекурсия»*. Следовательно, выше был приведён пример рекурсивной процедуры КВАДРАТ.

Рассмотрим, как работает данная процедура при вызове КВАДРАТ 5. Первая строка в теле процедуры прорисует квадрат со стороной 5, а затем управление перейдёт к команде во второй строке тела. Следующее действие — обращение к процедуре КВАДРАТ. Правда, предварительно вычисляется фактический параметр: предыдущее значение 5 увеличивается на 1 и становится равным 6. Процедура КВАДРАТ прочерчивает поверх уже имеющегося квадрата ещё один, чуть пошире. Снова изменяется значение параметра, и вновь — вызов процедуры КВАДРАТ, на сей раз изображается ещё более широкий квадрат. Теперь уже, очевидно, можно сказать «и так далее».



Такую рекурсию называют бесконечной: однажды начавшись, она никогда не завершится... если не сломается компьютер или программист не прервёт выполнение программы. И всё же если смотреть не с теоретической, а с практической точки зрения, то, конечно, лучше бы называть эту программу почти бесконечной. Дело даже не в том, что процесс вычерчивания квадратов перестанет быть заметным с какого-то достаточно большого значения параметра *К*: построение рисунка продолжится за пределами экрана, а в том, что бесконечную рекурсию возможно реализовать только в компьютерах с бесконечной памятью. Объём же памяти у современных компьютеров хотя и очень большой, но, во всяком случае, конечен.

Обращаясь к любой процедуре, нельзя забывать о том, что после её выполнения необходимо вернуть управление в ту самую точку программы, откуда был сделан вызов. Такой



КУКАРАЧА

Этот язык настолько прост, что в нём даже отсутствует понятие переменной. Тем не менее рекурсии применяются Кукарачей весьма активно.

Исполнитель действует в среде, которая составлена из клеточных полей, уложенных в несколько слоёв-этажей. Он перемещается по ячейкам командами ВЛЕВО, ВПРАВО, ВВЕРХ и ВНИЗ (в пределах одного этажа), ВПЕРЁД и НАЗАД (с этажа на этаж) при условии не покидать среду. Кукарача может передвигать кубики, стоящие на его пути (в том числе несколько кубиков), в направлении выполняемой команды. На кубиках поставлены буквы (или любые вводимые с клавиатуры символы). Некоторые из кубиков опрокинуты: на верхней грани вместо буквы изображён вопросительный знак. Кукарача может выяснить, какая буква стоит на расположенном в соседней ячейке кубике, толкнув его. После такого толчка ответ на вопрос о букве даёт условная команда

ЕСЛИ <*условие*> ТО <*команда-1*> ИНАЧЕ <*команда-2*>

где <*условие*> — это буква, разыскиваемая исполнителем. Например, по командам

```
ВПРАВО
ЕСЛИ А ТО ВВЕРХ ИНАЧЕ ВНИЗ
```

Кукарача направляется вверх (во вторую ячейку первой строки), если на кубике с вопросительным знаком стоит буква «А», и вниз (во вторую ячейку третьей строки) — в противном случае.

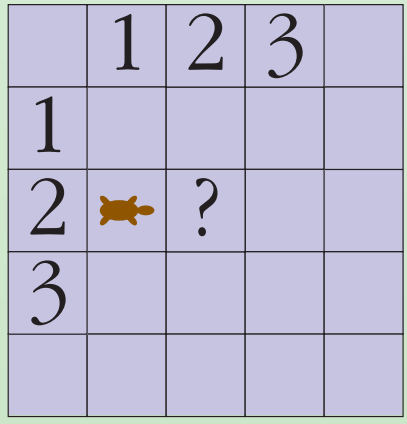
Условия могут включать и отрицания. Например, НЕ А означает «любая буква, кроме А».

Ограничитель управляющей структуры в этом языке не нужен, поскольку на ветвях ТО и ИНАЧЕ разрешается записывать не более одной команды.

Цикл с фиксированным числом повторений записывается так:

ПОВТОРИ *n* <*команда*>

где целое *n* — это число повторений.



процесс «регистрации» входов в вызываемую процедуру и запоминания точек входа не зависит от того, является ли процедура рекурсивной или нет. Та часть памяти, которая автоматически отводится компьютером для запоминания точек возврата, называется *стеком отложенных (открываемых) процедур.* В нём содержится перечень процедур (или, в случае рекурсивных вызовов, экземпляров рекурсивных процедур), которые открыты выполняемой программой. Стек устроен так, что в нём самой первой доступной является последняя из перечисленных процедур. Чем длиннее цепочка вложенных вызванных процедур, тем больше места занимает стек открываемых процедур. В обычных вызовах места в стеке достаточно при выполнении самых сложных (не рекурсивных!) программ. Но любая бесконечная рекурсия рано или поздно переполнит стек. Вот почему, хотя она и может выполняться на современном компьютере достаточно долго, этот процесс не вечен.

Однако как ни увлекательна теоретически бесконечная рекурсия, естественно поставить вопрос: а возможна ли рекурсия конечная? Ответить на него легко. В примере с рекурсивной процедурой непосредственно перед рекурсивным вызовом вставим сокращённую условную команду:

```
ЭТО КВАДРАТ :K
   ПОВТОРИ 4 [ВПЕРЁД :K ВПРАВО 90]
   ЕСЛИ K < 100 ТО КВАДРАТ :K + 1
КОНЕЦ
```

Такое «усовершенствование» программы останавливает цепочку рекурсивных вызовов после прочерчивания квадрата со стороной 100 или больше. Конечная рекурсивная процедура очень напоминает цикл с предусловием. В этом качестве её нередко используют программисты.

Механизм рекурсии считается непростым. Видимо, именно этим можно объяснить тот факт, что не в каждом языке программирования разрешается её использование. В ЛОГО рекурсия представлена так, что производит впечатление основной управляющей структуры. Описание Pascal подаёт её

как одну из возможных управляющих структур, а в очень распространённом в своё время языке программирования Basic рекурсии не допускались вовсе. Из этих примеров ясно, что место рекурсии в том или ином языке никак нельзя связывать со «сложностью» этого языка.

Несколько следующих понятий и примеров приводятся в рамках языка управления исполнителем Кукарача. Например, программа

```
ЭТО ПРОГУЛКА
    ПОВТОРИ 9 ВПРАВО
    ПОВТОРИ 9 ВНИЗ
    ПОВТОРИ 9 ВЛЕВО
    ПОВТОРИ 9 ВВЕРХ
ПРОГУЛКА
КОНЕЦ
```

заставит исполнителя долго-долго обходить поле размером 10 × 10. Правда, остановить выполнение такой программы иначе как прерыванием трудно. Допускаемый языком цикл с предусловием (цикл ПОКА) использует условие так же, как и в условных командах. Например, по командам

```
ПОКА ПУСТО ВНИЗ
ВПРАВО
```

Кукарача спускается по столбцу вниз до тех пор, пока он остаётся на полях, свободных от кубиков (условие ПУСТО), и переходит к команде ВПРАВО, следующей за циклом, как только условие нарушится. Это произойдёт в момент, когда Кукарача коснётся ближайшего кубика.

Та же задача — определение ближайшего непустого поля— решается и с помощью конечной рекурсии:

```
ЭТО СПУСК
    ВНИЗ
    ЕСЛИ ПУСТО СПУСК
КОНЕЦ
```

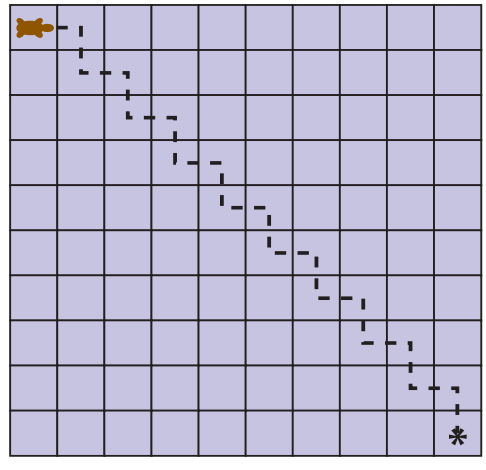
В следующей задаче — спуске Кукарачи по лесенке, в конце которой стоит кубик с символом *, обозначающей окончание программы, — отношения между вызывающей и вызываемой программами складываются иначе.

Кукарача решает задачу с нерекурсивной процедурой.

Вот вызов программы, в которой «выясняются» эти отношения. Здесь и процедура ЛЕСЕНКА, и процедура СТУПЕНЬ не обращаются к одноимённой процедуре.

```
ЛЕСЕНКА
ЭТО ЛЕСЕНКА
    СТУПЕНЬ
КОНЕЦ
ЭТО СТУПЕНЬ
    ВНИЗ
    ВПРАВО
    ЕСЛИ НЕ * ЛЕСЕНКА
КОНЕЦ
```

Казалось бы, обе они не подходят под ранее приводившееся упрощённое определение рекурсивной программы. Но определение рекурсии можно естественным образом расширить.



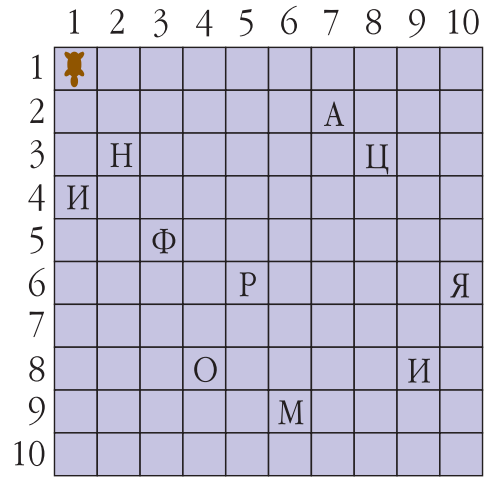
В этой задаче Кукарача использует три рекурсии.

Если программа *A* вызывает программу *B*, а программа *B* (возможно, через цепочку других вызываемых программ) вновь обращается к *A*, то эти две процедуры (*A* и *B*) образуют *косвенную* рекурсию (в отличие от *прямой,* или *непосредственной,* рекурсии). При таком определении непосредственная рекурсия рассматривается как частный случай косвенной рекурсии.

Во всех предшествующих примерах рекурсий, как прямой, так и косвенной, рекурсивный вызов всегда оказывался последней командой процедуры. В таком случае рекурсия работает в основном как один из способов управления повторяющимися процессами. Если же в описании вызываемой процедуры вслед за рекурсивным вызовом следуют другие команды, то возникают дополнительные, неожиданные на первый взгляд эффекты.

Вот задача, в которой будет задействована многоэтажная структура среды. По полю 10 × 10 (на первом этаже), за исключением первой и последней строк, беспорядочно разбросаны десять кубиков, по одному в каждом столбце. Кукарача, расположенный исходно в клетке (1,1,1), должен навести порядок на поле — собрать все кубики в первой строке (которая, таким образом, окажется заполненной).

Главной частью этой задачи является процедура, которая устанавливает на первой строке кубик из произвольного положения в столбце. Спуститься из первой строки до кубика нетрудно: достаточно использовать



косвенную рекурсию (или цикл ПОКА) с условием, определяющим прикосновение Кукарачи к кубику. Существенно сложнее вернуться с кубиком в исходную строку. Ведь предстоит подняться на столько же шагов вверх, сколько было сделано спусков вниз. Но у Кукарачи не может быть счётчика, так как нет переменных! Как же отсчитывать и запоминать шаги исполнителя? Здесь и приходит на помощь приём, называемый *отложенной рекурсией.*

Процедуры СТОЛБЕЦ и ОБХОД толкают кубик и возвращают его в первую строку:

```
ЭТО СТОЛБЕЦ
   ВНИЗ
   ЕСЛИ ПУСТО ТО СТОЛБЕЦ ИНАЧЕ ОБХОД
   ВВЕРХ
КОНЕЦ

ЭТО ОБХОД
   ВПЕРЁД ВНИЗ ВНИЗ НАЗАД ВВЕРХ
КОНЕЦ
```

В программе ОБХОД использованы команды Кукарачи, которые перемещают исполнителя на следующий этаж, ВПЕРЁД, и опускают его на предыдущий этаж, НАЗАД. Обратите внимание, что в языке ЛОГО такие команды перемещают исполнителя в горизонтальной плоскости, а в языке управления Кукарачей — по вертикали.

ОБХОД — это обычная процедура, с помощью которой Кукарача обходит по нижнему этажу исследуемую

ячейку. СТОЛБЕЦ — косвенно-рекурсивная процедура, которая вызывается при каждом «пустом» (без толчка) шаге вниз. Её особенность — последняя команда ВВЕРХ. Дело в том, что начальная команда процедуры СТОЛБЕЦ ВНИЗ и условная команда ЕСЛИ повторяются при каждом вызове процедуры СТОЛБЕЦ, а выполнение команды ВВЕРХ при этом откладывается до тех пор, пока не будут завершены команды на ветви ИНАЧЕ условной команды.

Как выполняются команды процедуры «столбец» для кубика, стоящего, например, в третьей строке?

1. Запуск процедуры СТОЛБЕЦ. Выполняется команда ВНИЗ. Кукарача спускается на вторую строчку.

2. Поскольку в предыдущем движении Кукарача не встретил кубика, то условие ПУСТО истинно, происходит второй вызов рекурсивной процедуры. Выполнение команды ВВЕРХ, составляющей часть первого вызова, отложено — она запоминается в стеке отложенных команд.

3. При втором вызове вновь выполняется команда ВНИЗ. Кукарача оказывается в третьей строке.

4. Здесь он встречает кубик. Условие ПУСТО нарушено. Поэтому исполнитель обходит кубик по соседнему полю и встаёт в позицию, из которой ему удобно толкать кубик вверх — на пятую строку.

Выполняется команда ВВЕРХ внутри процедуры «обход» . Кукарача поднимается на четвёртую строку и толкает кубик на третью.

5. Теперь выполняется команда ВВЕРХ, относящаяся к вызову процедуры СТОЛБЕЦ. Кукарача поднимается на третью строку, а кубик — на вторую. Второй рекурсивный вызов завершён.

6. Но не завершён ещё первый рекурсивный вызов, поскольку осталась невыполненной команда ВВЕРХ. Интерпретатор извлекает её из стека отложенных команд и передаёт исполнителю. Кукарача встаёт на вторую строку, толкая кубик в первую. Завершён первый (последний в стеке) рекурсивный вызов, а с ним и выполнение процедуры СТОЛБЕЦ.

Описанные этапы условно отображаются схемой:

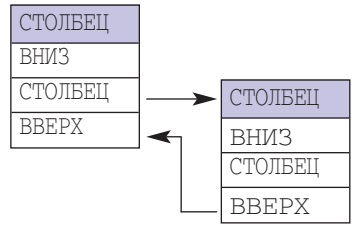
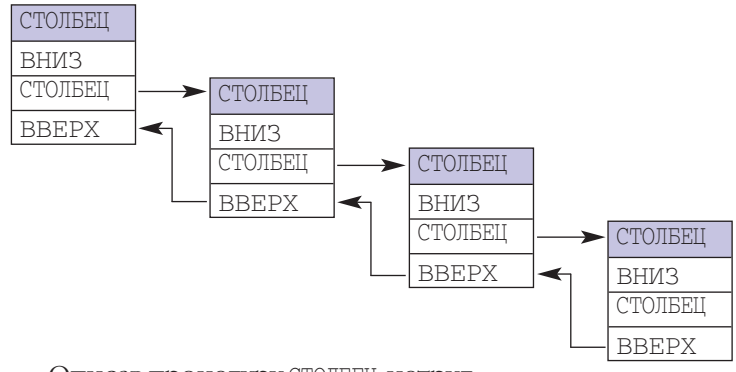
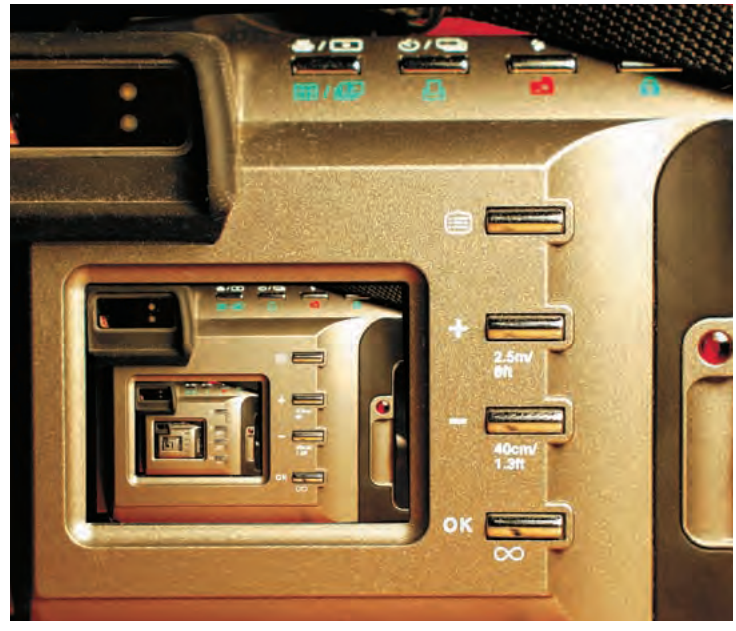


Схема для кубика пятой строки имеет вид:



Описав процедуру СТОЛБЕЦ, нетрудно включить её в процедуру ПОРЯДОК, которая решает нашу задачу:

```
ЭТО ПОРЯДОК
   ПОВТОРИ 9 {СТОЛБЕЦ ВПЕРЁД ВВЕРХ
   ВПРАВО НАЗАД}
   СТОЛБЕЦ
КОНЕЦ
```



ВЫЧИСЛЕНИЕ ФАКТОРИАЛА ЧИСЛА *n*

Примером рекурсивной функции может служить программа вычисления факториала числа *n*, обозначаемого *n!* (*n* с восклицательным знаком). Факториал от целого числа n представляет собой произведение всех целых чисел от 1 до *n* включительно и вычисляется по формуле

$$Ф(n) = 1 \cdot 2 \cdot 3 \ldots (n-2) \cdot (n-1) \cdot n$$

Принято считать, что 0! = 1.

Программа-функция вычисления факториала приведена ниже на языке КуМир:

```
алг цел факториал (цел n)
  дано n
  надо
нач
| если n <= 1
| | то знач :=1
| | иначе знач := n * факториал (n-1)
| всё
кон
```

Эта функция возвращает числовое значение *n!* вызвавшей её программе. При этом в начале работы программа не знает ещё, что такое факториал (*n*), и, следовательно, откладывает его вычисление, по крайней мере, до следующего рекурсивного вызова. Однако и тогда может случиться, что вычисление факториала окажется отложенным. И только при *n*-м вызове компьютер начнёт вычисления, извлекая из стека отложенных команд сначала вызов факториал (*n*), за ним факториала (*n*-1) и т. д. Здесь мы тоже имеем дело с отложенной рекурсией.

По командам

```
a := 5
b := 6
вывод (факториал (a) + факториал (b)) : 2
```

компьютер выведет на экран числовое значение

поскольку

$$5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120, \text{ а } 6! = 120 \cdot 6 = 720.$$

(Фигурные скобки, употреблённые в этом примере, позволяют считать окаймлённую ими группу команд единой командой в соответствии с соглашениями языка программирования.)

Во всех приведённых выше примерах в качестве рекурсивных программ фигурировали только процедуры. Однако рекурсии с таким же успехом используются и в программировании функций.

Не следует думать, что с рекурсией люди сталкиваются только в информатике. Рекурсивными функциями наполнена современная математика. Да и в повседневной жизни можно найти немало примеров рекурсий. Художники любят создавать рекурсивные картинки: например, на рисунке изображён телезритель, рассматривающий экран, на котором видно зрителя, тоже взирающего на телеэкран. На том экране — снова зритель и т. д.

Пример литературной рекурсии — известный стих-притча про попа, который любил свою собаку:

У попа была собака,
Он её любил.
Она съела кусок мяса —
Он её убил.
Убил и закопал,
И надпись написал:
У попа была собака...

КАК ПИСАТЬ НАДЁЖНЫЕ ПРОГРАММЫ

Любую программу можно рассматривать как описание процесса преобразования информации: есть некая исходная информация, из неё нужно получить результат. В эту схему укладываются не только вычислительные задачи, но и интерактивные программы, включая игры, при этом входной информацией считаются действия пользователя, а результатом — действия компьютера.

В современном мире с помощью машин выполняются всё более сложные задачи. Поэтому растут требования к надёжности программ. Надёжные программы выдают верный результат при допустимых исходных данных и фиксируют ошибочные случаи. Чтобы проверить правильность программ, существуют специальные методы. Они делятся на две группы — *методы доказательства* и *методы тестирования*.

Если программа доказана, то зачем тестировать? Во-первых, иногда само доказательство бывает неверным, во-вторых, в тексте программы могут оказаться опечатки. При доказательстве автор часто опирается на то, что он подразумевал, когда писал программу; компьютер же всегда делает то, что написано. Чтобы убедиться в надёжности программы, тестов должно быть много.

Некоторые заблуждаются, считая, что цель тестирования — возможность убедиться в правильной работе программы. В действительности это поиск ошибок. Разница между этими взглядами на тестирование — в психологическом подходе к составлению тестов и их результатам.

В первом случае тестирующий признаёт, что программа работает, и стремится подтвердить это. Поэтому количество тестов сокращается (чего там ещё проверять, и так всё ясно), они не содержат подвохов и описывают обычно в точности тот же набор ситуаций, который обрабатывается программой. Такое тестирование наиболее характерно для авторов, особенно начинающих, которые при написании программ и их тестировании не допускают мысли о непредусмотренных ситуациях.

ЗАДАЧА О ТРЕУГОЛЬНИКЕ

Даны три натуральных числа. Определить тип треугольника по заданным сторонам (равносторонний, равнобедренный, разносторонний) и углам (остроугольный, прямоугольный, тупоугольный).

Ограничения: $0 < a, b, c < 1000$.

Ввод: 3 4 5

Вывод: треугольник разносторонний, прямоугольный.

Чтобы подготовить хороший набор тестов, надо хотя бы в общих чертах представлять себе возможные решения и ошибки. Исходные данные заведомо корректные, т. е. удовлетворяют заданным в условиях ограничениям.

Подготовка тестов начинается с анализа условий. Легко заметить, что в условиях совсем не говорится об одном крайне важном обстоятельстве: треугольник может вообще не существовать. Это бывает при нарушении неравенства треугольника: каждая сторона должна быть меньше суммы двух других. Таким образом, получается следующий список возможных реакций программы на корректные входные данные:

- не существует;
- равносторонний, остроугольный;
- равнобедренный, остроугольный;
- равнобедренный, тупоугольный;
- разносторонний, остроугольный;
- разносторонний, прямоугольный;
- разносторонний, тупоугольный.

(В списке нет варианта «равнобедренный, прямоугольный», так как он невыполним при сторонах, заданных натуральными числами.)

Очевидно, что при тестировании каждый из перечисленных ответов нужно получить хотя бы по одному разу.

Какие типичные ошибки могут быть допущены при решении этой задачи? Обратим особое внимание на симметрию входных данных. Все три заданных числа равноправны, так как в условиях ничего не сказано о каком-то специальном порядке. В такой ситуации в решении часто возникают ошибки. Неравенство треугольника, например, может проверяться не для всех трёх сторон. Практический вывод: нужны тесты, в которых одни и те же числа будут располагаться в различном порядке.

Необходимо отдельно рассмотреть предельный случай, когда сумма двух сторон равна третьей. Кроме того, нужно проверить предельные значения, лежащие на границе допустимого диапазона.

Часто встречается ошибка переполнения. Она возникает, когда в программе использованы 16-битные числа, поскольку их не хватает для возведения в квадрат (для определения углов треугольника стороны возводят в квадрат). Тесты в последней секции специально подобраны так, чтобы данная ошибка сразу обнаруживалась.

В таблице приведён примерный набор тестов для задачи о треугольнике.

Ввод	**Ожидаемый результат**	**Комментарий**
3 20 100	не существует	Для всех ли сторон
20 100 3	не существует	проверено неравенство
100 3 20	не существует	треугольника?
999 1 1	не существует	Несуществующие
1 1 999	не существует	«равнобедренные»
1 999 1	не существует	треугольники
2 3 5	не существует	Вырожденные
30 50 20	не существует	треугольники
500 200 300	не существует	
1 1 1	равносторонний, остроугольный	Крайние значения
999 999 999	равносторонний, остроугольный	Одновременно тестируем правильный треугольник
500 500 300	равнобедренный, остроугольный	Равнобедренные треугольники с разными углами и разным порядком ввода сторон
400 700 400	равнобедренный, тупоугольный	
32 48 56	разносторонний, остроугольный	Аналогично
30 40 50	разносторонний, прямоугольный	для разносторонних
256 368 600	разносторонний, тупоугольный	треугольников
255 256 257	разносторонний, остроугольный	Проверка переполнения

При втором подходе всё меняется. Цель тестирующего — сломать программу, доказать её ошибочность, найти трудные для обработки случаи и предложить программе именно их. Понятно, что здесь дело не ограничивается простейшими тестами. В ход идут самые каверзные входные данные, самые неожиданные их комбинации. И если программа выдержала подобный натиск, можно надеяться, что она действительно написана правильно.

Все методы тестирования можно разделить на две группы: тестирование по принципу чёрного ящика и по принципу белого ящика.

Чёрным ящиком обычно называют объект, который способен как-то реагировать на внешние воздействия, но его внутренняя структура неизвестна. Например, любая бытовая техника для большинства людей — типичный чёрный ящик. Все знают, что, если нажать на определённую кнопку, звук телевизора станет громче, но мало кто представляет, какие реальные физические процессы при этом происходят.

Тестирование по методу чёрного ящика основано на анализе задачи, которую программа должна решать. Изучая условия задачи, определяют различные возможные ситуации, формируют для каждой из них пример входных данных и обязательно находят правильный результат. Таким образом, полученный набор тестов применим к любой программе, решающей данную задачу.

Тестирование по методу белого ящика основано на тексте конкретной программы, в которой выделяются отдельные блоки. Каждый такой блок можно тестировать как единое целое (чёрный ящик) или разбивать на ещё более мелкие блоки (белый ящик).

Обычно в процессе работы над программой автор тестирует её по методу белого ящика. Когда же работа закончена, программу обязательно должны проверить по методу чёрного ящика, причём очень жёстко и строго.

Самое трудное в тестировании — составить набор тестов. Это творческая, трудноформализуемая операция, однако существуют общие принципы:

- тесты составляются на основе условий задачи, они не должны учитывать особенности конкретной программы;
- если задача разбивается на разные случаи, необходимы тесты, проверяющие их все;
- составляя набор тестов, желательно представить разные способы решения задачи, в том числе готовить тест, который выявляет заведомо ошибочное решение, дав косвенный ответ.
- для каждого типа данных существуют определённые критические значения. Необходимо проверить, как обрабатываются эти значения во входных данных.
- если допустимые значения входных данных ограничены, требуется

протестировать работу программы при предельных и близких к ним значениях. Например, если в условии задачи говорится о целом *N* из диапазона от 1 до 100, надо отдельно рассмотреть работу программы при *N*=1, 2, 99, 100;

• в тесты обязательно нужно включать допустимые в задаче значения (положительные, отрицательные и нулевые), а также запрещённые (нечисловые или нетекстовые).

Для некоторых тестирование — деструктивная деятельность: ломать не строить. На самом деле умение грамотно провести исследование, выявить глубоко скрытые в программе ошибки очень ценный навык, поэтому такой специалист весьма нужный человек в коллективе программистов.

КАК ПИСАТЬ КРАСИВЫЕ ПРОГРАММЫ?

В процессе разработки каждая программа не один раз подвергается исправлениям, и при этом человеку каждый раз приходится читать её, разбираться в структуре, искать места, в которых могут возникнуть затруднения.

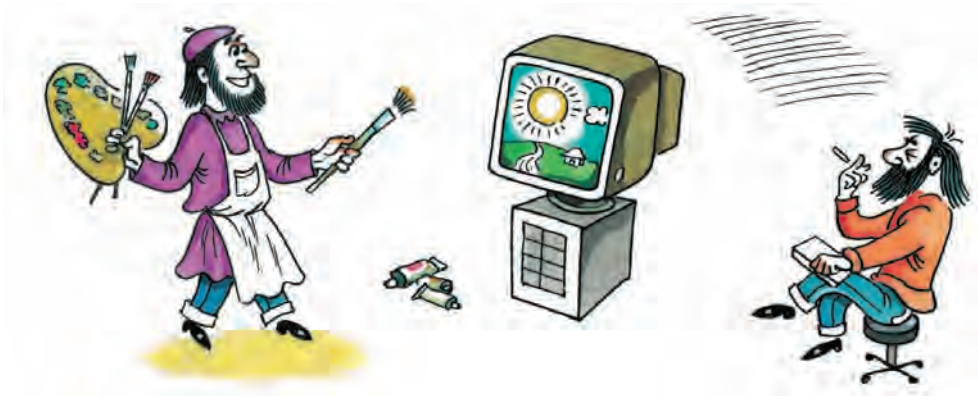
Начинающие программисты часто считают, что им нет необходимости специально заботиться о внятности программы: ведь они сами её придумали и всё в ней знают, зачем тратить время на какое-то оформление? Увы, это заблуждение, хотя и очень распространённое.

Как бы хорошо программист ни понимал свою программу, есть предел, после которого невозможно удержать в памяти все детали. Если работа ведётся с продолжительными перерывами, вспомнить подробности бывает не так-то просто. Довольно частая жизненная ситуация — программу начинает один человек, а доделывает другой.

Опытные программисты всегда уделяют внимание тому, как выглядит их программа. Получится красивой — её будет легко читать, понимать и исправлять, а значит, у неё больше шансов оказаться правильной и полезной.

Правила оформления программ обычно называют стилем программирования. Одни и те же действия почти всегда можно выполнить по-разному. У каждого программиста есть свои привычки и пристрастия, которые и образуют его стиль. Со временем у него обязательно вырабатывается свой неповторимый «почерк», свой способ использования инструментов, по которому его нетрудно опознать, как по подписи на бумаге.

Казалось бы, если стиль столь индивидуален, стоит ли обсуждать его? Обратимся опять к сравнению.



Общеизвестно, что у любого человека почерк уникален, как отпечатки пальцев. Но почерк, в отличие от отпечатков, можно оценить: красивый — некрасивый, разборчивый — неразборчивый. Руку специально готовят к письму: особые упражнения по образцам помогают развить внимание, мышление и воображение.

Это применимо и к стилю программирования, который бывает как хорошим, так и плохим. Ему нужно учить, и лучше, если начинающий программист придерживается некоторого стандарта, как первоклассники — прописей.

Стиль программирования сравним и со стилем речи, где имеются свои нормы (есть даже понятие стилистической ошибки). У каждого человека есть свои любимые слова, своя манера построения фраз, свой способ изложения. Эксперты-литературоведы способны по фрагменту текста определить его автора. Но ведь и речевой стиль тоже можно оценивать! Ведь говорят о хорошем и плохом стиле речи.

И ещё одно необычное сравнение — кулинарное. Стиль программирования подобен приправам. Пищевая ценность блюда определяется основными использованными продуктами. Полученные калории и чувство сытости от приправ не зависят, а вот вкусовые ощущения складываются во многом благодаря именно им. Если повар забыл положить их или, что ещё хуже, использовал неверно, насытиться полученным блюдом будет всё-таки можно, а вот получить удовольствие — вряд ли.

Необходимо всё время помнить, что программу кто-то будет читать, и надо сделать так, чтобы при этом не возникало лишних трудностей.

Вот несколько рекомендаций, которые позволяют понять, что такое хороший стиль.

Кто ясно мыслит, тот ясно излагает!

Прежде чем писать программу, надо чётко уяснить, какая поставлена задача и как её решить. Если в мыслях нет ясности — не помогут никакие приёмы и ухищрения.

Не забывать о комментариях!

Все языки программирования позволяют включать в программу комментарии — пояснительный текст, который нужен только для человека.

Мыслитель. Скульптура Огюста Родена.

Примеры плохих комментариев: «счётчик», «дополнительная переменная», «вещественное число», «строка». Примеры хороших комментариев: «количество найденных совпадений», «самое большое число в списке».

Компьютер не читает комментарии, он просто не замечает их. Комментарии — важнейший элемент стиля, это очень мощный инструмент, позволяющий сделать программу действительно понятной.

• У программы должно быть предисловие — большой вступительный комментарий, который может занимать несколько страниц текста. Он включает следующие разделы:

— имя автора программы (творчество не должно быть анонимным!);

— название программы и её назначение;

— по возможности подробное описание решаемой задачи;

— описание исходных данных и требуемых результатов задачи;

— описание основной идеи решения;

— ссылки на литературу и другие источники, связанные с задачей и её решением;

— дата написания первоначальной версии программы и история последующих изменений.

• Каждое описание переменных, объектов, величин должно сопровождаться комментарием, поясняющим их содержательную суть.

• Для каждой процедуры требуется давать небольшой вступительный комментарий — для чего предназначена данная конкретная процедура.

• Программы порой основаны на довольно сложных алгоритмах, понять которые по тексту не так-то легко. В этом случае хорошие комментарии, поясняющие смысл алгоритма в целом, могут оказаться очень полезными.

• Не доверять комментариям.

Этот совет предназначен не для тех, кто пишет, а для тех, кто будет читать программу.

Вот типичный фрагмент из программы на языке Pascal:

```
k := k + 1; {увеличим  k на 1}
```

Комментарий абсолютно ничего не прибавляет к тексту программы: читатель и так видит, что `k` увеличивается на 1 (а если не видит, то здесь ему не поможет даже самый подробный комментарий). Желательно описать, почему происходит это увеличение. Например, «нашли очередное совпадение» или «переходим к следующему элементу».

Необходимо помнить, что компьютер всегда делает то, что написано в программе, а не в комментариях. Хороший комментарий рассказывает о замысле программиста, а анализ самой программы — о реализации задуманного.

Но, с другой стороны, подробный комментарий в иных случаях помогает заметить, что выполняются не те действия, о которых было заявлено.

Понимание переменных — ключ к пониманию программы.

Любая программа обрабатывает какую-то информацию, представленную в виде переменных. Поэтому необходимо в первую очередь понять смысл переменных, которые в ней используются.

• Все переменные должны быть описаны.

Некоторые языки программирования — КуМир, Pascal, C — заставляют описывать переменные, иначе программа считается синтаксически неверной.

А вот Basic и некоторые другие языки дают программисту больше свободы: использование переменных подразумевает автоматическое сообщение о них, а тип переменной и другие необходимые данные определяются из контекста.

Но это очень обманчивая свобода. Обязательные описания дисциплини-

руют программиста, стимулируют более тщательную проработку структуры данных. И читать такие программы намного легче: информация обо всех переменных собрана в одном месте, и если читатель забыл, что означает какая-то переменная, он знает, где сможет найти эти сведения.

• Надо тщательно выбирать имена переменных.

Описание и комментарий — это очень хорошо, но когда читают программу, они остаются далеко в начале текста, и не очень хочется слишком часто к ним возвращаться. Имя же переменной совсем другое дело.

Имя неотделимо от переменной, оно всегда рядом и соответствует её значению в программе.

Следует избегать однобуквенных имён и конструкций типа «буква с цифрой» — в них легко запутаться. Надо давать близкие имена родственным данным, но нельзя называть похожими именами данные, которые ничего общего не имеют.

Разделяй и властвуй!

Нельзя объять необъятное, поэтому грамотное разбиение программы на процедуры — ключ к успешному решению больших задач. (Здесь под процедурами понимаются также функции, подпрограммы и все остальные воплощения процедур в различных языках программирования.)

• Любая процедура решает свою небольшую задачу, обозначенную в виде начального комментария. В идеале она должна описываться одной короткой фразой. Если формулировка получается длинной и сложной, значит, разбиение проведено не до конца: можно выделить и записать отдельно ещё одну или несколько процедур.

• Объём процедуры должен быть таким, чтобы её было несложно охватить как единое целое, понять сразу всю логику её работы. Обычно процедура без описаний и вступительного комментария должна целиком помещаться на экране. Если экрана не хватает, есть смысл подумать о дальнейшем разбиении.

Текст должен быть удобен для чтения.

Для компьютера чаще всего неважно, как ваша программа разбита на

строки и где в них поставлены пробелы. Компилятор должен отделить лексемы (минимальные элементы, из которых состоит программа: имена, знаки операций, служебные слова) друг от друга, а остальное ему безразлично.

Человек же, наоборот, обращает особое внимание на то, как оформлен текст, ведь от этого зависит его восприятие.

• Надо «выделять» алгоритмическую структуру.

Опытные программисты всегда пишут программу «лесенкой». Каждый раз, когда в ней используется алгоритмическая структура (цикл, ветвление, выбор и т. д.), текст сдвигается вправо. Например:

```
<действия до цикла>
| while <условие цикла> do
| begin
| | <тело цикла>
| end
<действия после цикла>
```

Существует даже строго доказанная теорема, что любую программу можно записать, не применяя goto (если, конечно, в языке программирования есть конструкции if-then-else и while). Опыт профессиональных программистов говорит о том, что использование goto чаще всего оказывается следствием того, что разработчик просто не умеет грамотно строить структуру программы.

Эта запись позволяет сразу увидеть, где начинается и где заканчивается конструкция, какие команды входят в неё, а какие нет.

• Смещение означает подчинение.

Данное правило — пояснительное расширение предыдущего.

Вообще, всегда, когда надо показать, что один фрагмент подчинён другому, подчинённый фрагмент записывают со смещением вправо. Тем самым поясняют, что это не самостоятельное действие, а часть предыдущего.

• Не надо экономить на пробелах.

Они облегчают чтение и понимание программы. Если в одной строке записывается несколько операторов, необходимо поставить пробелы между ними. Во многих случаях желательно использовать пробелы при знаках действий.

Надо избегать волшебных чисел.

Числа в программе не стоит использовать в явном виде. Когда в тексте программы встречается такое число (что-нибудь вроде 179), читателю часто бывает совершенно непонятно, что оно означает и откуда взялось.

Тяжело приходится программисту, когда возникает необходимость изменить подобное число. Допустим, программа обрабатывает массив из 100 элементов. Понятно, что число 100 будет встречаться довольно часто. А если теперь потребуется задать исходным массив из 200 элементов, программисту придётся заменить по всей программе 100 на 200. Мало того, где-то нужно будет менять 99 на 199, 101 на 201, а где-то окажется, что 100 — это температура кипения воды, а вовсе не количество элементов и в замене не нуждается. В общем, выполняется кропотливая и не очень приятная работа.

Числам надо давать имена и работать только с ними. В Pascal для этого существует понятие константы. В языках, где механизм констант не предусмотрен (КуМир, Basic), вводят дополнительную переменную и присваивают ей значение в самом начале программы.

Конечно, такие числа, как 0 или 1, не стоит делать константами — это только усложнит программу.

Шаг за шагом.

Известно, что проще всего разобраться в линейной программе, в которой все действия выполняются последовательно, одно за другим, как записаны. Это не означает, что надо избегать циклов и ветвлений — без них невозможно построить ни один сложный алгоритм. Но лучше, чтобы конструкции были короткими, тогда каждая из них может рассматриваться как отдельное действие в общей линейной последовательности.

• Без goto.

Оператор goto делает программу чуть-чуть короче, нарушая последовательность выполнения. После goto надо читать не следующую строчку, а ту, на которую задан переход.

Однако в участках, куда совершается переход, может таиться опасность. Именно там чаще всего происходят ошибки. Причина очевидна: в одно и то же место программы можно попасть из различных мест с разной предысторией. Уследить за соблюдением всех необходимых условий становится очень трудно, отсюда и сложности.

• Не надо прерывать циклы.

Ещё один способ нарушения последовательности выполнения — до-

срочное прекращение цикла. В некоторых языках есть даже специальные средства для этого: break — в C, выход — в КуМире. Создатели языка Turbo Pascal разработали процедуру «break», аналогичную одноимённому оператору в языке C.

Цикл — самая сложная для понимания алгоритмическая структура. Обычно предполагается, что заголовок цикла полностью описывает условия его продолжения и завершения. Досрочное прекращение нарушает это правило, поэтому им лучше не пользоваться.

• Программа должна заканчиваться естественно. Не надо вставлять в середину программы команды, ведущие к прекращению работы. Особенно опасно включать такие команды в процедуры. Завершение должно быть естественным — процедура обязана вернуть управление в точку вызова.

Единственное исключение из этого правила — обнаружение фатальной ошибки, которая делает дальнейшую работу программы бессмысленной.

Использование логических величин.

В некоторых языках есть специальные обозначения для логических величин (boolean — в Pascal, лог — в КуМире). В C логические значения представляются с помощью числовых величин, но с ними также выполняются все необходимые операции.

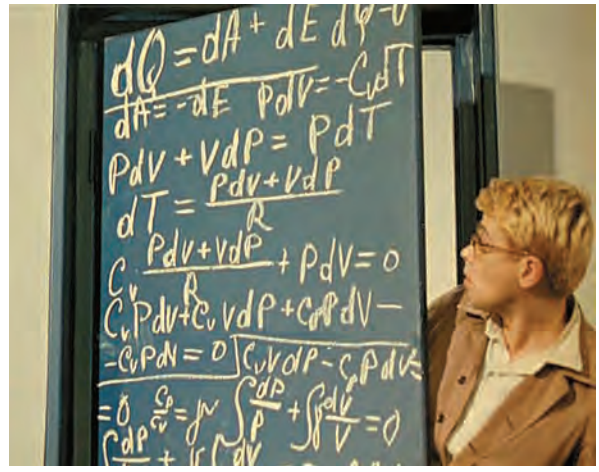
Начинающие программисты часто не понимают сути логических величин и создают громоздкие выражения там, где можно написать коротко и ясно.

Например, в программах на языке Pascal встречается такое сравнение: `if a=true then...` Более простой вариант: `if a then...` Логическая величина — это условие, не нуждающееся ни в каком дополнительном сравнении.

Ещё пример: выражение `if a=true then a:=false else a:=true` преобразуем в `a:=not a`. Как говорится, почувствуйте разницу!

А вот пример неудачного использования логической величины в языке C: `if (a > b) return 1; return 0`. Лучше оформить этот фрагмент так: `return (a > b)`.

Последняя запись подчёркивает, что результат функции — не число,



а логическая величина, условие, которое может быть истинным или ложным.

***

Конечно, здесь представлены далеко не все тонкости стиля, на которые следует обращать внимание. Но если программист будет строго следовать хотя бы только этим правилам, то очень скоро на его программы станет приятно смотреть, их будет легче читать и понимать. А ещё через некоторое время он начнёт сам придумывать правила, формировать свой собственный стиль.

ХРАНЕНИЕ ИНФОРМАЦИИ

КОМПЬЮТЕР И КНИГОПЕЧАТАНИЕ

Первопечатник
Иван Фёдоров

Рубикон между двумя эпохами мировой культуры и научно-технического прогресса перейдён. Человечество вступило в эпоху глобальных компьютерных технологий.

Первый опыт в книгопечатании был предпринят ещё в 1041—1048 гг. в Китае Би Шэном. Тем не менее традиционно изобретателем книгопечатания считают Иоганна Гутенберга (около 1394—1468), придумавшего первый печатный станок в Европе. Печатный набор, предложенный им, позволил делать любое число идентичных отпечатков текста, составленного из элементов — *литер*. Их можно было заменять на другие — для печати новой страницы. Гутенберг также сконструировал прибор для изготовления литер. В полый металлический стержень со съёмной нижней крышкой из мягкого металла, на котором предварительно выбивался рисунок буквы, заливался специальный сплав. Застывший сплав — готовая литера на торце имела зеркальный рисунок буквы. Иоганн Гутенберг разработал рецепты типографского сплава и типографской краски. Своими изобретениями он проложил дорогу стандартизации шрифтов и удешевил процесс производства книг.

Однако Гутенберг полиграфическим способом изготовлял лишь текст. Отпечатанные рисунки и орнаменты впервые появились у немецкого печатника Петера Шёффера в 1457 г. на страницах «Псалтыри». Гравированные на металле иллюстрации и текст с наборной формы впервые отпечатал на одном листе флорентийский типограф Николо ди Лоренцо в 1477 г.

В 1564 г. Иван Фёдоров (около 1510—1583) в Москве издал первую русскую печатную книгу «Апостол». В последующие годы модернизация печатного станка сводилась к механизации отдельных процессов, первоначально выполнявшихся вручную, а также к последовательной замене деревянных частей металлическими.

Распространение технологии книгопечатания — технологии тиражирования информации позволило не только издавать в большом количестве дешёвые книги, но и организовать выпуск периодических изданий — журналов и газет.

Последовали и другие открытия, в корне изменившие технологии хранения и передачи информации в обществе: фотография, кино, телеграфия, телефония, радио и телевидение.

Достоинства книгопечатания очевидны: операции тиражирования и копирования автоматизированы и могут идти без участия человека. Но обработку информации, за редкими исключениями, производит человек, и она почти не автоматизирована.

На исходе второго тысячелетия стало ясно, что в информационном обеспечении жизни общества начинаются кризисные явления.

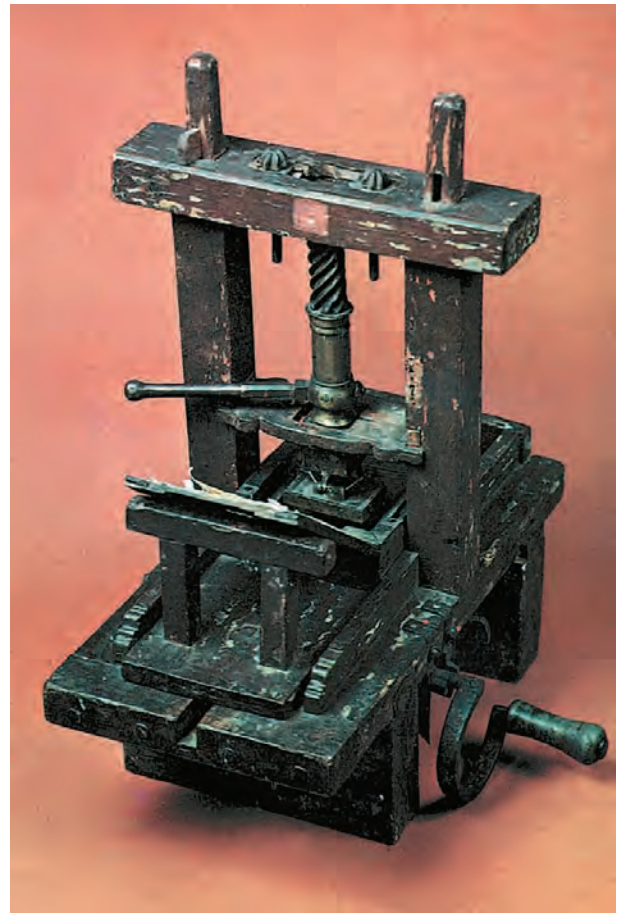
• Накопленную информацию трудно перерабатывать. В эпоху книгопечатания объём циркулирующей в обществе информации растёт лавинообразно. При этом обработка информации по-прежнему остаётся «ручной работой» и осуществляется исключительно человеком. Миллионы людей заняты только систематизацией и поиском нужной информации. Поиск становится всё более и более дорогим, долгим и ненадёжным.

Необходимую технологию или устройство, изобретённые ранее, оказывается дешевле изобрести заново, чем найти в море имеющейся информации.

• Накопленная информация разнородна. Разные виды информации — тексты, изображения, фильмы, звук — представлены в разном виде, требуют совершенно разных устройств для записи, воспроизведения и копирования, разных условий хранения и методик обработки человеком.

• Накопленную информацию трудно сохранять. Не отвечают новым потребностям и старые технологии хранения информации.

Носители информации — книги, журналы, фотографии, киноленты, магнитные ленты — со временем стареют или изнашиваются при эксплуатаци. Операции копирования напечатанных текстов, фотоматериалов, аудиозаписей вносят небольшие искажения, но после нескольких копирований эти искажения становятся значительными. Как ни совершенствуй записывающую и воспроизводящую аппаратуру, при аналоговом её представлении искажения информации избежать нельзя.



Печатный станок Ивана Фёдорова.

Естественным выходом является решение трёх поставленных проблем. Использование компьютера как универсального инструмента обработки информации позволяет автоматизировать этот процесс и во многих областях осуществлять его без участия человека. А перевод информации при хранении и передаче из аналогового в цифровую форму даёт возможность не только решить проблему долгосрочного хранения, но и сделать более удобным поиск нужной информации.

БУМАГА

Старинный способ изготовления бумаги.

Свои первые тексты человек выводил прутом на песке и углем на стене пещеры, выцарапывал на табличках из сырой глины и высекал на каменных плитах... Но все они были либо недолговечны, либо хрупки, либо слишком неудобны в пользовании (говорят, что в Древнем Китае для перевозки некоторых сочинений требовалось несколько телег, ведь написаны они были на бамбуковых дощечках). Подходящий для письма материал первыми обнаружили жители Древнего Египта. Это был похожий на нашу осоку папирус, в изобилии произраставший в долине Нила и даже возделывавшийся как культурное растение. Из него изготовляли верёвки, циновки, обувь, лодки и паруса к ним. Из его стеблей египтяне вырезали длинные полосы, смазывали их клеем, клали под пресс и высушивали на солнце. Получались удобные для письма тонкие, гибкие и прочные листы, которые склеивали и сворачивали в длинные свитки. Так что не случайно слово «бумага» в европейских языках напоминает нам о папирусе: papier — во французском и немецком, paper — в английском.

Из Египта приблизительно в VII в. до н. э. папирусные свитки проникли в Грецию, а уже оттуда — в Рим. Папирус был достаточно редок, поэтому в других странах находили иной материал для письма. В Китае это был шёлк, на Руси в быту употребляли берёсту — на ней писали письма, деловые документы: расписки, завещания и т. д. (знаменитые новгородские берестяные грамоты донесли до наших дней многие подробности повседневной жизни предков). В Пергамском царстве стали использовать особым образом выделанные телячьи и бараньи шкуры — пергамен. К сожалению, он был слишком дорог и, кроме того, весьма подвержен действию влаги. А в повседневной жизни с античных времён применяли маленькие вощёные дощечки, на которых острым стержнем писали какие-либо заметки или расчёты (другой конец стержня делался в виде плоской лопаточки, чтобы разравнивать воск).

Во II в. н. э. в Китае изобрели способ производства бумаги из растительного сырья: бамбука, тростника, древесной коры. Позднее секрет её изготовления узнали арабы (применявшие пеньковое и льняное тряпьё), а приблизительно в X столетии — и европейцы. Качество бумаги было столь высоким, что рукописи того времени дошли до наших дней в прекрасном состоянии: она осталась такой же белоснежной, сохранила свою гибкость. Она оказалась гораздо дешевле пергамена и благодаря этому быстро вы-

Знаменитый норвежский учёный и путешественник Тур Хейердал (1914—2002) совершил плавание через Атлантику на папирусной лодке «Ра».

теснила его. Однако ещё долго считалась предметом роскоши — пользовались ею лишь в самых важных случаях. Только спустя несколько веков бумага стала действительно массовым продуктом, и это способствовало бурному росту книгопечатания, развитию газетного дела и т. п.

В течение тысячи лет бумага верой и правдой служила человеку. Совершенствовались только технологии производства да способы нанесения текста на бумагу: гусиное, а потом стальное перо, обмакиваемые в чернила, авторучка, шариковая ручка, фломастер... Параллельно развивались технические средства — печатный станок, пишущая машинка, а в последние десятилетия подключённые к компьютеру печатающие устройства — принтеры.

С появлением компьютеров впервые заговорили о необходимости внедрения безбумажных технологий. Ведь писать, исправлять и хранить тексты, обмениваться ими можно и без переноса на бумагу. И хотя совсем без неё по-прежнему обойтись нельзя, оказалось, что человечество очень быстро утрачивает навыки письма! Так что профессия графолога, определяющего характер человека по его почерку, судя по всему, скоро уйдёт в прошлое. Люди почти перестали писать друг другу письма на бумаге, журналисты пересылают в редакцию репортажи по электронной почте, а поэты и писатели забыли, что такое рукопись. Свои труды они хранят теперь на дискетах в виде файлов формата doc. Нетрудно представить, насколько беднее были бы мы сегодня, не имея возможности увидеть рукописи Пушкина и Льва Толстого, если бы от десятилетней работы Михаила Булгакова над рукописью «Мастера и Маргариты» осталась лишь последняя версия файла...

О том, что когда-то в школах существовали уроки чистописания, уже давно никто не вспоминает. Для современного ребёнка становится проблемой разборчиво написать несколько фраз. Не случайно многие учёные бьют тревогу: доказано, что человеческий мозг развивается особенно интенсивно именно во время письма!

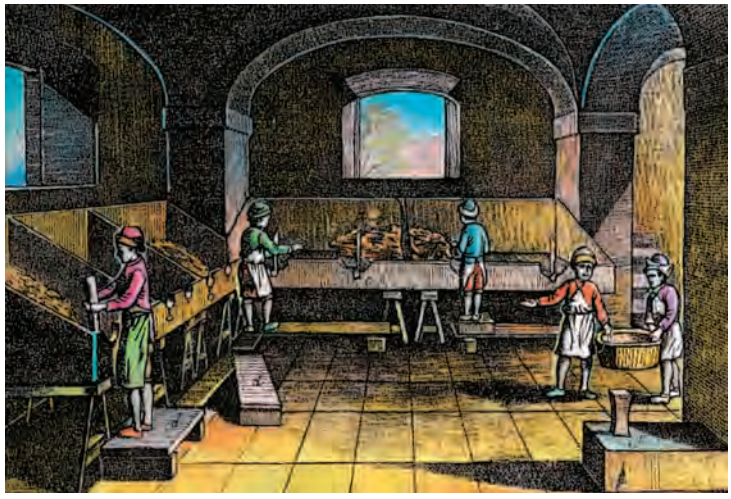
БИБЛИОТЕКА ДЛЯ УЧЁНЫХ

Не раз в прошлом тысячелетии учёные обращались к способам хранения и поиска нужной информации в книгах и журналах. Так, Чарлз Дарвин, будучи состоятельным человеком и живо интересуясь новыми научными открытиями, имел весьма своеобразную библиотеку. Всю литературу он приобретал в двух экземплярах. Из полученных книг и журналов вырезал страницы лишь по интересовавшим его вопросам, а остальное уничтожал. Нужные листки он систематизировал и хранил.

Такой подход требует непрерывного балансирования, ведь существует опасность выбросить показавшийся ненужным фрагмент, который впоследствии будет востребован. Кроме того, подобный метод организации личной библиотеки не слишком помогает при поиске нужной книги или статьи.

В XX в. неоднократно пытались систематизировать информацию для удобства пользования. Поиск информации в библиотеке подчас занимал не только несколько часов или дней, но месяцы и даже годы. Наиболее удобно искать в энциклопедиях, где статьи упорядочены по алфавиту. Однако даже Большая советская энциклопедия, насчитывающая более 70 увесистых томов, не содержит и сотой доли специализированной информации. Реферативные журналы являлись хорошим подспорьем для учёных в поиске научной информации. Полноценный же поиск был немыслим без перевода текстов в цифровое представление.

Академик Иван Васильевич Обреимов в брошюре «О цифровом кодировании информации» предлагал создать цитатную машину, которая позволит научному работнику создать библиотеку, систематизируя и переводя в цифровое представление (на специальные карточки) часть информации, хранящейся в библиотеке. Скорость поиска с помощью такой частично механической машины могла достигать 225 карточек (цитат) в час. Очевидно, что подобные скорость и удобство несравнимы с современными компьютерными библиотеками и энциклопедиями, когда вся Большая советская энциклопедия помещается на трёх CD ROMах.



Подготовка материала при изготовлении бумаги.

Однако бороться с прогрессом бессмысленно, да и не надо. Гораздо правильнее использовать его достижения — так решили шведские учёные и инженеры, объявившие о создании уникального набора для письма, состоящего из особенных бумаги и авторучки.

Внешне цифровая авторучка ничем не отличается от обычной. Но кроме баллончика с чернилами внутри неё находятся встроенная миниатюрная цифровая камера, микропроцессор, память и передающее устройство. При этом весит авторучка менее 50 г! Чтобы начать работу, не надо нажимать никаких кнопок, достаточно просто снять колпачок. При движении пера по бумаге камера делает в секунду около 100 снимков, которые обрабатываются процессором, а получаемая информация пересылается в память мобильного телефона или компьютера. Одновременно с появлением на бумаге рукописный текст воспроизводится и на экране монитора, а при необходимости его можно преобразовать в печатный.

Писать такой ручкой надо на особой бумаге. Впрочем, внешне «цифровая бумага» отличается от обычной лишь своим оттенком, он называется «белая ночь». На каждый лист с использованием специальной технологии наносится множество микроскопических точек, отстоящих друг от друга на 0,3 мм. Когда перо авторучки движется по листу бумаги, чернила закрашивают эти точки. Обрабатывающий снимки процессор однозначно определяет их координаты (точки расположены на листе так, что для этого достаточно их минимального количества) и таким образом формирует цифровое изображение.

Использование этой технологии позволяет посылать на дисплей мобильного телефона или по электронной почте не только печатные сообщения, но и рукописные тексты. Её можно применять, например, для пересылки заполненных от руки анкет, заявлений, рецептов врача (может быть, скоро слово «подпись» снова обретёт свой первоначальный смысл, утраченный в словосочетании «электронная подпись»?). Несомненно, будет предложено и множество других применений. Так что, вполне вероятно, эпоха рукописей продлится и в информационном веке.

лПДЙТПЧЛЙ ФЕЛУФВ

Название этой статьи — не заклинание, а всего лишь слова «кодировки текста», написанные в другой его кодировке.

Если при выводе (на экран или на печать) хранимого в компьютере текста использовать разные кодировки, то и результат вывода будет разным. Число возможных кодировок очень велико. Чтобы текст был читаем на нескольких компьютерах, должна существовать кодировка, поддерживаемая на каждом из этих компьютеров. Так возникает задача выбора стандартных кодировок.

СТАНДАРТНЫЕ КОДИРОВКИ ТЕКСТА

Самой первой стандартной кодировкой стала кодировка ASCII (American Standard Code for Information Interchange). Её начали создавать в конце 50-х гг. XX в., а в 1968 г. она была утверждена Национальным институтом стандартизации США (ANSI).

ASCII использует для кодирования символов числа от 0 до 127. В таблице приведены коды текстовых символов ASCII от 32 до 126. Остальные числа (0—31 и 127) служат для кодировки управляющих (нетекстовых) символов (например, символ звукового сигнала или символ конца строки, переводящий вывод следующих за ним символов в новую строку наподобие клавиши «возврат каретки» у пишущих машинок).

Код	Символ
32—47	! " # $ % & ' ( ) * + , - . /
48—63	0 1 2 3 4 5 6 7 8 9 : ; < = > ?
64—79	@ A B C D E F G H I J K L M N O
80—95	P Q R S T U V W X Y Z [\ ] ^ _
96—111	` a b c d e f g h i j k l m n o
112—126	p q r s t u v w x y z { | } ~

Двоичное представление чисел от 0 до 127 содержит не более 7 цифр.

Это значит, что в ASCII код символа занимает 7 бит компьютерной памяти. (Наименьшая адресуемая единица информации в компьютере — байт содержит 8 бит.)

Было признано удобным выделить под размещение кода символа 1 байт памяти. Легко подсчитать, что в одном байте можно разместить 256 чисел (0—255). Это значит что в одном байте кроме символов ASCII можно закодировать ещё 128 символов.

Кодом символа называют число, преобразующееся при выводе в этот символ. Таблицу соответствия символов языка и набора чисел называют таблицей кодировки текста или просто кодировкой.

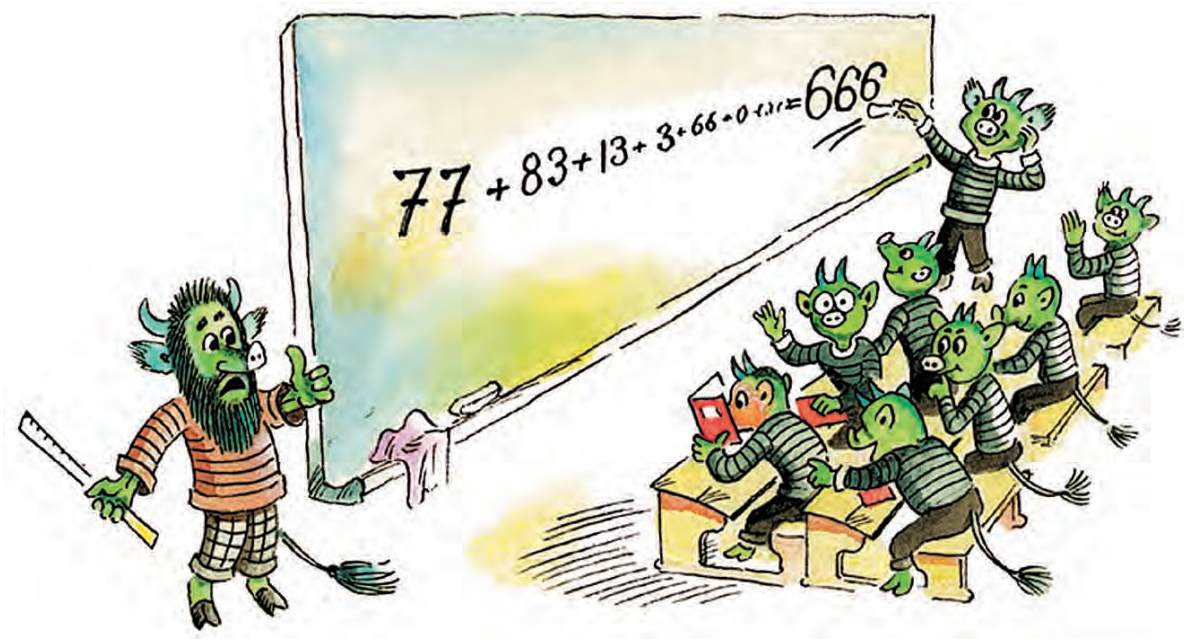
Например, символы русского, греческого или какого-либо другого алфавита, математические символы (знаки корня, интеграла, суммы и др.), символы псевдографики (вертикальная, горизонтальная и другие линии, используемые для изображения таблиц) и т. д.

Во вновь создаваемых 8-битных кодировках текста бралась за основу кодировка ASCII, а для кодирования дополнительных символов использовались числа от 128 до 255, а иногда также от 0 до 31 (вместо кодов управляющих символов ASCII).

РУССКИЕ КОДИРОВКИ

В России наибольшее распространение получили три русскоязычные кодировки: КОИ-8, альтернативная (которую часто называют кодировкой DOS), а также Windows.

КОИ-8 (кодировка для обмена информацией 8 бит) является стандартной русской кодовой таблицей на компьютерах, работающих под управлением операционной системы UNIX. Кроме того, КОИ-8 фактически стала



стандартом для представления русскоязычных текстов в глобальной сети Интернет.

Альтернативная кодировка широко применялась в операционной сети MS DOS. Её название произошло из описания Государственным стандартом (ГОСТ) в 80-х гг. двух кодировок — «основной» и «альтернативной». «Основная» кодировка вышла из употребления из-за своего неудобства.

Кодировка Windows используется в операционной системе MS Windows. От КОИ-8 и альтернативной кодировки она отличается не только кодировкой символов, но и набором кодируемых символов. Например, в этом наборе могут отсутствовать символы псевдографики, зато много разнообразных символов кавычек.

ПРОБЛЕМЫ 8-БИТНЫХ КОДИРОВОК

Программа, работающая с текстом, даже если она поддерживает все три русскоязычные кодировки, способна правильно изобразить текст на экране или на принтере, только если знает, в какой именно кодировке был введён этот текст.

Некоторые программы умеют сами вычислять кодировку текста по частоте вхождения некоторых букв в начальной части текста. Однако этот метод может не сработать для нестандартных текстов (например, если начальная часть текста написана с использованием только латинских букв, как это часто бывает с текстами программ). Тогда пользователю приходится перебирать всевозможные кодировки, т. е. подбирать кодировку вручную.

В случае если неизвестен язык, на котором написан текст, задача определения использованной при вводе текста кодировки оказывается достаточно сложной при огромном количестве 8-битных национальных кодировок.

Решение этой проблемы в последнее время возлагается на стандартную международную кодировку, включающую в себя коды всех национальных символов.

МЕЖДУНАРОДНЫЕ КОДИРОВКИ

Для создания международной кодировки текста был создан UCS (Universal Character Set — универсальный набор символов), куда вошли символы всех

языков мира, а также некоторые другие знаки, например фонетический алфавит IPA, используемый для записи произношения слов.

В настоящее время существует две международные кодировки для набора символов UCS: Unicode и ISO/IEC 10646. В них для представления в компьютерном виде одного символа используется 2 байт памяти. Это позволяет закодировать 65 536 символов, что больше объёма UCS.

Кодировки Unicode и ISO/IEC 10646 полностью идентичны и в наборе кодируемых символов, и в их кодах. Отличие состоит только в практическом применении. Так, Unicode включает в себя множество алгоритмов и спецификаций для переноса текстов между программными приложениями и операционными системами.

Будущее — за международными кодировками. Unicode всё шире используется в таких популярных операционных системах, как Windows и UNIX. Причём конкурентоспособность этих операционных систем сильно зависит от степени использования данной кодировки, поскольку это заметно упрощает процесс обмена информацией.

ТЕКСТОВЫЕ РЕДАКТОРЫ

КОМПЬЮТЕРЫ В КНИГОПЕЧАТАНИИ

Появление возможности ввода текстов в компьютер произвело такую же революцию в письменности, как и изобретение книгопечатания. Хотя ввод текстов в ЭВМ схож с набором книги в типографии — набрав один раз, можно напечатать сколько угодно копий, — он имеет ряд преимуществ. При обнаружении ошибки не надо коверкать изображение текста — раздвигать символы для вставки недостающих или, наоборот, растягивать их на всю строку после удаления лишних. Перемещения будут сделаны в памяти компьютера и на печати не отразятся.

Известный колумбийский писатель, лауреат Нобелевской премии по литературе, автор незабываемого романа «Сто лет одиночества» Габриэль Гарсиа Маркес ещё в далёких 80-х гг. XX в. был в числе первых, кто создавал свои произведения на компьютере. Сейчас набор текстов является, пожалуй, одной из самых широко используемых возможностей компьютера.

ЭВОЛЮЦИЯ РЕДАКТОРОВ

Для ввода текста в память компьютера и изменения (редактирования) введённого текста используются специальные компьютерные программы, называемые *редакторами*.

Первые редакторы были командными: компьютеру отдавались команды, что надо сделать с текстом. Человек мог видеть текст на мониторе компьютера и после выполнения команды вывести текст на печать.

Например, типичный сеанс работы с текстом в редакторе TECO, разработанном в 1975 г., выглядел примерно так:

```
! DISPLAY CURRENT LINE'S ASCII CODE
, 20 PER DISPLAY-LINE.!
! D F KOENIG, 1989-11-24. !
@^U.L[^[^A Line length = ^A^[

MN0U.0^[QN<Q.0A:=^[^A
^A^[((%.0/10)*10)
Q.0"E^[((Q.0/20)*20)-Q.0"E^[^A
^A^[ ^[|^[^A  ^A^[ ^['^['^[>^[^A
^A^[[
```

Каждый символ в приведённом фрагменте обозначает вызов какой-либо команды редактора, например вставки некоторого символа в определённую позицию заданной строки текста.

Габриэль Гарсиа Маркес.

У программистов существовала игра: записать свою фамилию в качестве команд редактору и точно угадать, что будет с редактируемым текстом, если к нему применить этот набор команд.

Фото с экрана программы редактора текста «МикроМир» для DOS.

Запись последовательности команд имеет устрашающий вид бессмысленного набора символов. Командные редакторы, такие, как TECO, называются также строчными (*англ.* line editors). Их главное неудобство заключается в том, что при редактировании необходимо знать и указать в команде номер строки и номер символа, которые надо изменить.

Следующее поколение редакторов — экранные редакторы (screen editors), или WYSIWYG-редакторы (от *англ.* what you see is what you get — «что видите, то и имеете»). При их использовании текст выводится на экран в таком виде, какой он будет иметь на печати. Место изменения текста указывается специальным значком — курсором, который можно «гонять» по тексту вверх, вниз, вправо и влево. Курсор бывает в виде тонкой вертикальной/горизонтальной чёрточки, находящейся слева/внизу от текущего символа, либо в виде прямоугольника, расположенного поверх символа, так что изображение самого символа становится негативным.

РЕДАКТОРЫ — СИСТЕМЫ ДЛЯ ВИЗУАЛИЗАЦИИ И ИЗМЕНЕНИЯ ТЕКСТА

Все современные редакторы при вводе и изменении текста поддерживают его изображение на экране монитора. Основная работа редактора — показать введённый текст; при нажатии пользователем какой-либо клавиши определить, как это должно изменить текст; откорректировать изображение текста на экране.

Кроме того, современные редакторы выполняют много других функций: проверяют набираемый текст, форматируют его, выводят на печать, отправляют по электронной почте и др.

ЗАЧЕМ НУЖНЫ МЕНЮ

Каждую команду редактора можно вызвать нажатием какой-либо определённой клавиши. Чем больше в редакторе разнообразных функций (команд), тем больше различных клавиш требуется для их вызова.

На запоминание комбинации клавиш придётся потратить немало сил и времени, поэтому для удобства используют меню. Меню в последнее время становятся всё более стандартными. В верхней строке экрана редактора располагаются несколько ключевых слов: «Файл» (папка), «Правка», «Вид», «Формат» и некоторые другие. Каждому слову соответствует своё меню, которое скрыто от пользователя. Оно высвечивается только при вызове: если стрелку-указатель «мыши» подвести к подходящему слову и нажать кнопку «мыши». В меню перечислены команды, связанные с ключевым словом.

РАБОТА С ФАЙЛАМИ

Слово «файл» происходит от английского file, что в информатике понимается как «папка». Папка-файл может хранить в себе текст, картинку, компьютерную программу или что-либо ещё.

Работа с текстом начинается с команды «Открыть». После того как папка открыта, можно увидеть текст, который в ней содержится, и редактировать его. Сохранять изменённый текст разрешается как в исходном файле в исходном формате записи на диск (команда «Сохранить»), так и в другом файле и/или формате (команда «Сохранить как...»). Это делается для того, чтобы внесённые изменения не пропали при неожиданном отключении напряжения в сети или сбое в компьютере, поскольку изменения текста во время сеанса редактирования не вносятся сразу в файл с текстом, а хранятся в промежуточной памяти.

При выходе из редактируемого текста редактор обычно спрашивает, сохранить ли изменения в файле. Можно закрыть файл без сохранения изменений, если текст испорчен и хочется вернуть его в первоначальный вид.

СОВМЕСТИМОСТЬ РЕДАКТОРОВ

Освоение пользователем редактора — длительный, трудоёмкий процесс. При смене редактора приходится заново осваивать новые понятия, новые команды, новые функции клавиш, новые пункты меню. Такая перспектива может даже привести к отказу от нового редактора. Подобная сложность применения невыгодна производителю,

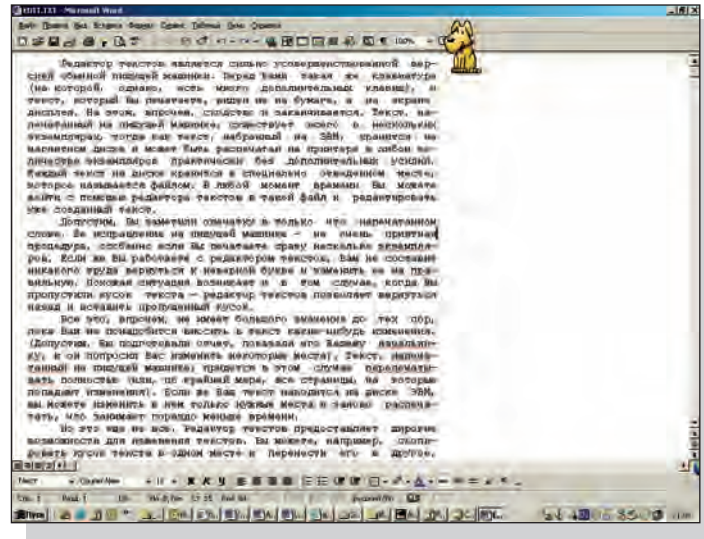


Фото с экрана редактора Microsoft Word.

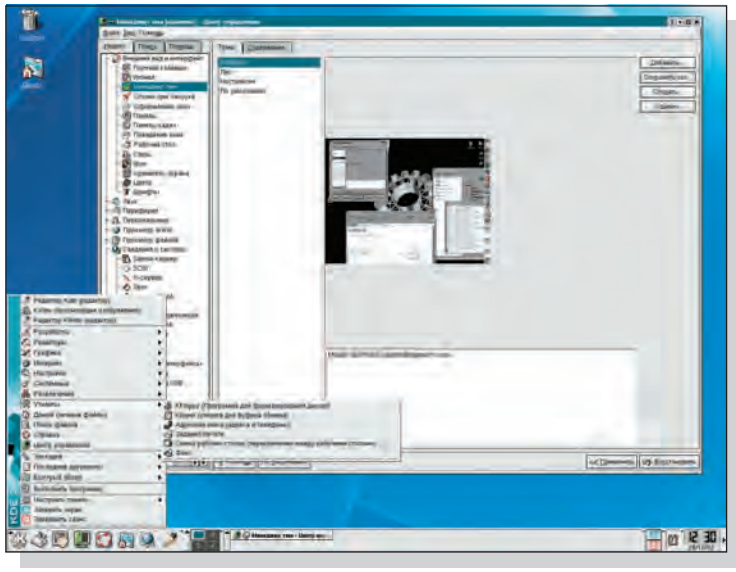


Фото с экрана офисной системы X-Windows Linux.

поэтому пользовательский интерфейс стандартизируют.

Например, появилось понятие: «программа совместима с Microsoft Office». Оно означает, что программа специально разработана так, чтобы выглядеть похожей на приложения из комплекта Microsoft Office, которые уже знакомы многим пользователям.

Все редакторы текстов предназначены для одной и той же работы — создания текста. Поэтому набор команд и функций у них схож. Отличаются же они друг от друга специальными возможностями (например, проверкой орфографии или наличием встроенных электронных таблиц). Так что выбор редактора определяется задачами пользователя и его личными вкусами.

МАКЕТИРОВАНИЕ

Неотъемлемыми командами текстовых редакторов являются ввод и удаление символа. Кроме этого, в редакторах есть команды, облегчающие набор (убыстряющие перемещения по тексту, позволяющие копировать фрагменты текста).

Огромное количество книг, журналов, газет привычно для современного человека, но только в XVI в., когда люди придумали машину, способную печатать тексты на бумаге, появилась реальная возможность изготовлять много экземпляров одной и той же книги. Процесс подготовки к печати был сложен и трудоёмок. Изображение, которое нужно было напечатать, вытравливалось (с поверхности убирался верхний слой в тех местах, где рисунка не будет), а каждая литера (буква в зеркальном изображении, буква-«наоборот») помещалась строго на своё место, чтобы получился стройный ряд слов, знаков препинания. Из этого строя и состояли строки будущей книги. В таком тексте очень трудно было находить ошибки набора — опечатки, ведь текст набирался как отражение будущей страницы. Следующая операция состояла в нанесении на выпуклые части собранного текста специальной типографской краски. Затем, пока краска ещё не высохла, пластину с текстом плотно прижимали к чистому листу бумаги. Получался оттиск — рождалась страница будущей книги. С появлением компьютеров в процессе печати, и особенно подготовки книги к печати — *макетировании,* произошли революционные изменения.

Набор в типографии начала XX в.

С чего начинается макетирование книги? С выбора формата. Формат бывает разный, достаточно зайти в книжный магазин, чтобы убедиться, что книги отличаются не только размера-

ми, но и формами. Например, книга о чае может иметь очертания заварочного чайника — современные технологии это позволяют.

В зависимости от формата книги художник придумывает макет, т. е. выбирает размеры полей, взаимное расположение текста, иллюстраций, сочетание шрифтов, цветовые решения, разрабатывает обложку и большие иллюстрации на разворот книги — *шмуцтитулы,* придумывает различные элементы дизайна, например указывающие на повторяющиеся элементы текста или требующие особого внимания читателя. Так, иногда квадратики в тексте обращают внимание читателя на примечание, или цитату, или занимательный материал, которые иллюстрируют мысль основного текста, помогают его понять. Чем гармоничнее макет, тем проще и приятнее читать саму книгу.

Сейчас достаточно редко можно встретить макет, разработанный вручную, без применения графических редакторов и других специальных программ. Но даже если художник делал его без использования компьютера, созданный макет всё равно переносится в компьютер в виде специальных *мастер-страниц,* шаблонов, по образу и подобию которых в *верстальной программе* будут собраны все страницы книги.

Текст книги должен легко читаться. Для этого художниками придумано много различных шрифтов. Есть, например, специальные шрифты для заголовков или для рекламы. Такой текст должен сразу бросаться в глаза, кричать: «Прочти меня!». В популярных журналах не так много внимания уделяется тексту, как общему оформлению, дизайну. Глянцевые страницы модных журналов порой требуют сильного напряжения зрения читателя. Создатели подобных изданий и не ставят перед собой задачу привлечь читателя содержательным текстом — пусть человек просто разглядывает красивые картинки. Напротив, текст, напечатанный «книжным» шрифтом, можно читать долго, не уставая. Особое внимание уделяется шрифту для школьной учебной литературы. Плохо выбранный шрифт не только делает книгу трудноусваиваемой, но и вреден. Такой шрифт способен привести к ухудшению зрения, повышению утомляемости школьника.

РЕДАКТОР ТЕКСТОВ

При вызове редактора на экране появляется окно для его работы. В окне редактора, как правило, имеется:

- *Заголовок* с названием редактора и именем редактируемого текста.
- *Меню* — полоса под заголовком со списком ключевых слов.
- *Приборная панель* — полоса с кнопками, соответствующими некоторым командам редактора. Кнопка «нажимается» щелчком «мыши» по ней, при этом вызывается нужная команда.
- *Панель форматов* — полоса, на которой расположены окошко с именами шрифтов, окошко с размерами шрифтов и несколько кнопок с параметрами шрифта и формата. Справа у этих окошек имеется кнопка, при её «нажатии» окошко раскрывается, появляется список шрифтов или размеров шрифтов, и пользователь щелчком «мыши» на нужной строке может выбрать подходящий.
- *Линейка*.
- *Окно редактирования текста* — белый прямоугольник с мигающим курсором, где изображается текст. Если текст полностью не помещается в окне редактирования, то появляются полосы прокрутки: справа — полоса вертикальной прокрутки, перемещающая текст вверх/вниз; внизу — полоса горизонтальной прокрутки, перемещающая текст вправо/влево.
- *Статусная панель*, содержащая много полезной информации, например подсказку о работе редактора, а также информацию о местоположении курсора в тексте: номер строки и номер символа.

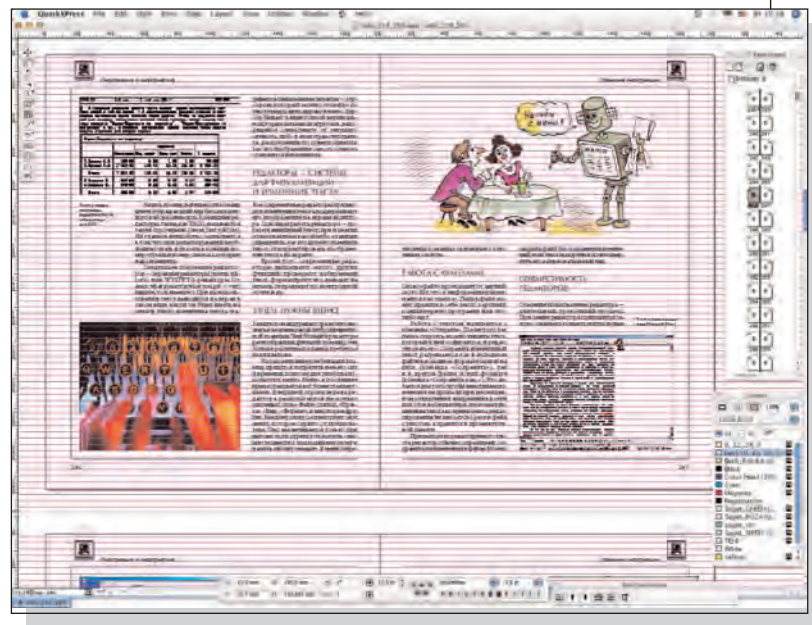


Фото с экрана программы макетирования текста с подготовленными страницами этой книги.

Книги редко бывают совсем без иллюстраций. Порой просто невозможно что-то объяснить, не проиллюстрировав написанное. Способов создания книжных иллюстраций множество.

Можно рисовать красками, карандашами или чем-то иным на бумаге либо любой другой поверхности или делать рисунки в компьютерных программах, с помощью которых можно создавать специальный файл с изображением.

Ещё один способ — использовать фотографии, сканируя их в компьютер с отпечатка, с другой книги или прямо с негатива, применяя *слайд-сканеры*. Или без сканера — фотографируя на цифровую камеру, создающую, в отличие от обычного фотоаппарата, изображение не на плёнке, а на цифровом носителе в виде компьютерного файла.

Можно делать *коллажи,* т. е. совмещать разные изображения в одном, например взять изображение с фотографии, «вырезав» на компьютере только одного человека, и поместить его на новый фон — побережье океана с живописными пальмами или у входа в отель.

Когда собран основной материал — тексты и иллюстрации, разработан макет, пора браться за изготовление книги. Здесь вступает в дело верстальщик. Он и собирает из отдельных кусочков целое — книгу в том виде, в котором видит её читатель. Следуя задумке дизайнера (а иногда сочетая две специальности — дизайнера и верстальщика), он «заливает» текст, располагает картинки, подписи, комментарии на страницах будущей книги так, чтобы, с одной стороны, было красиво, а с другой — понятно. Это не всегда просто, процесс творческий...

Компьютер верстальщика — его основной инструмент. И достаточно часто это не привычный IBM PC, а Macintosh. В США такие компьютеры широко распространены, особенно в образовании и полиграфии. В России считается, что этот компьютер создан специально для издательской деятельности и художников. Вероятно, здесь сказывается надёжность машин и удобство интерфейса, ведь творческому человеку трудно переделывать несколько раз одно и то же из-за «зависания» программы.

Так на замечательной машине верстальщик создаёт виртуальную книгу, существующую сначала только как компьютерный файл. Потом этот файл попадает в типографию (в современных типографиях ручного набора уже не бывает), и там другие «волшебники» печатают книгу.

ЭЛЕКТРОННЫЕ ТАБЛИЦЫ

«В понедельник 6 уроков и 2 часа самостоятельных занятий, во вторник 5 уроков и 3 часа самостоятельных занятий, в среду 5 уроков и 3 часа самостоятельных занятий, в четверг 6 уроков и 2 часа самостоятельных занятий, в пятницу 5 уроков и 2 часа самостоятельных занятий».

Речь идёт о расписании занятий школьника. Если нужно отыскать в этой записке количество уроков в четверг, то придётся просмотреть практически весь текст. Но это крайне неудобно, вот почему расписание занятий обычно пишут в виде таблицы (от *лат.* tabula — «запись сведений по графам»).

И тогда решить эту задачу уже не составляет большого труда. Нужно только найти клетку таблицы на пересечении строки «Четверг» и графы «Количество уроков».

День недели	Количество уроков	Часы самост. занятий
Понедельник	6	2
Вторник	5	3
Среда	5	3
Четверг	6	2
Пятница	5	2

Числа, содержащиеся в клетках таблицы, могут быть связаны друг с другом. Например, в правой таблице содержимое четвёртой колонки является суммой данных второй и третьей колонок, а содержимое строки «Всего» — суммой данных всех предыдущих строк.

Пока таблица существует на бумаге, трудно использовать связь чисел друг с другом. Когда таблица находится в компьютере, можно поручить машине заполнение некоторых клеток. Так, для второй таблицы компьютер мог бы сам заполнить клетки последней строки «Всего», сложив данные в столбцах и последней колонки «Общее количество часов», просуммировав цифры в двух строках.

Каков же выигрыш от автоматического заполнения таблицы? Во-первых, существенная экономия времени, затрачиваемого на заполнение таблицы. (Пусть заполнение таблицы занимает 4 мин: вычисления — 1 мин, ввод 18 чисел — 3 мин. Тогда при автоматическом заполнении время будет затрачено только на ввод 10 чисел, а это менее 2 мин.) Во-вторых, уменьшается количество ошибок.

В случае второй таблицы, например, возможностей ошибиться при вводе 10 чисел меньше, чем при вводе 18 чисел. В вычислениях компьютер, в отличие от человека, вообще не ошибается (конечно, если не будет сбоя в питании или сильных электромагнитных помех).

Программы, считывающие данные из таблиц и заполняющие таблицы новыми вычисленными данными, называют программами обработки таблиц. Трудно даже представить, сколько времени позволяет сэкономить программа для обработки таблиц размером, скажем, в 1 млн клеток, содержащих 20-значные числа (что-нибудь вроде 45 987 801 734 897 235 947 или 0,06 237 891 209 875 611), особенно если формулы вычисления громоздки и их много! И скольких ошибок ввода и вычислений можно избежать благодаря такой программе!

День недели	Количество уроков	Часы самостоят. занятий	Общее количество часов
Понедельник	6	2	8
Вторник	5	3	8
Среда	5	3	8
Четверг	6	2	8
Пятница	5	2	7
Всего	27	12	39

Конечно, создание программы обработки таблиц требует некоторого количества времени и труда, но, однажды написанная, она ускоряет работу при каждом заполнении таблицы (а заполняют таблицу иногда по тысяче или миллиону раз).

ЭЛЕКТРОННЫЕ ТАБЛИЦЫ

Как проходит процесс разработки программы заполнения таблиц? Пользователь сообщает программисту алгоритм заполнения таблицы, т. е. формулы нахождения тех чисел, которые нужно записать в клетки. Программист переводит эти формулы на язык программирования, а также реализует ввод в программу исходных данных (чисел, вводимых пользователем) и вывод из программы результатов вычислений (т. е. запись полученных чисел в таблицу).

Фото с экрана программы электронных таблиц Excel.

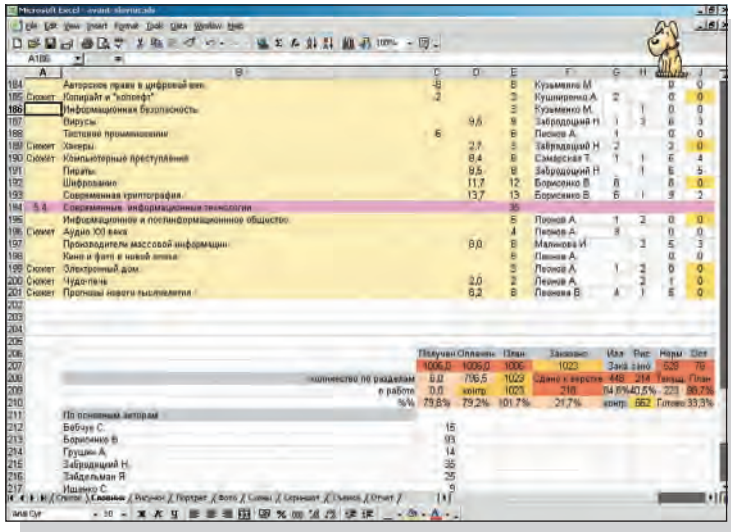
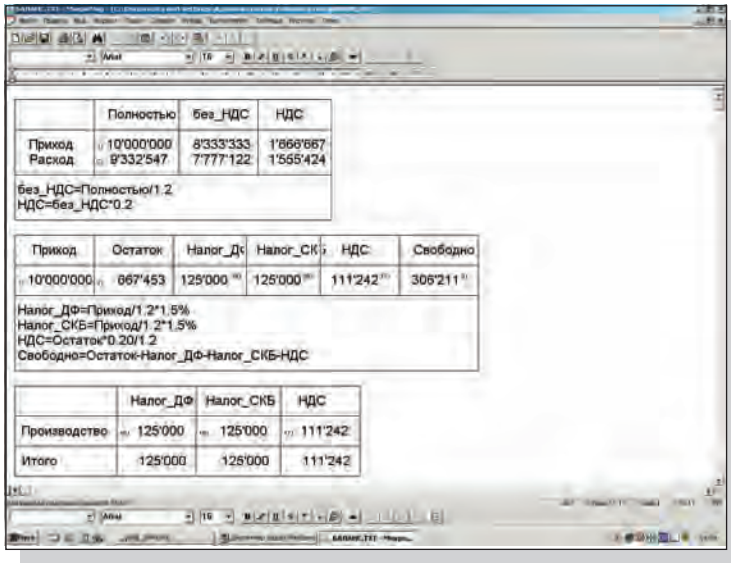


Фото с экрана программы электронных таблиц МикроКальк.



	Полностью	без_НДС	НДС
Приход	10'000'000	8'333'333	1'666'667
Расход	9'332'547	7'777'122	1'555'424

без_НДС=Полностью/1.2
НДС=без_НДС*0.2

Приход	Остаток	Налог_ДФ	Налог_СКБ	НДС	Свободно
10'000'000	667'453	125'000	125'000	111'242	305'211

Налог_ДФ=Приход/1.2*1.5%
Налог_СКБ=Приход/1.2*1.5%
НДС=Остаток*0.20/1.2
Свободно=Остаток-Налог_ДФ-Налог_СКБ-НДС

	Налог_ДФ	Налог_СКБ	НДС
Производство	125'000	125'000	111'242
Итого	125'000	125'000	111'242

При изменении формата таблицы (например, появление дополнительной колонки) или изменении алгоритма заполнения таблицы (изменение формулы) программу обработки таблицы приходится менять, а значит, пользователь опять должен прибегать к помощи программиста.

Процесс требовал упрощения. Придуманные электронные таблицы должны были исключить участие в работе программиста, давая пользователю возможность самому запрограммировать алгоритм обработки его таблицы, а при изменениях алгоритма — самому его перепрограммировать.

Электронная таблица (*англ.* spreadsheet — «табличный процессор») — это программа или программная система, используемая для автоматической обработки данных, хранящихся в таблице, по формулам, задаваемым пользователем.

Каждой клетке, строке или колонке таблицы человек может назначить формулу для вычисления содержимого по данным, находящимся в определённых клетках, строках или колонках. При любом изменении содержимого какой-либо клетки всё содержимое таблицы автоматически пересчитывается.

Первая электронная таблица появилась в 1979 г. Её назвали очень удачно — «VisiCalc» (от *лат.* visio — «зрение», «ви́дение» и calculo — «считать», «подсчитывать»). Легко догадаться, что VisiCalc производит видимые, наглядные вычисления.

Вскоре после VisiCalc стали появляться и другие электронные таблицы. Уже в феврале 1981 г. была разработана популярная в нашей стране электронная таблица SuperCalc. А в начале 1983 г. появилась электронная таблица «1—2—3» фирмы Lotus, которая до сих пор входит в тройку самых конкурентоспособных электронных таблиц (наряду с Excel компании Microsoft и Quattro фирмы Borland).

В начале 90-х гг. электронные таблицы разрабатывались и в России. Одну из них — «МикроКальк», встроенную в текстовый редактор «МикроМир», разработали на механико-математическом факультете МГУ. «МикроКальк» была очень проста в использовании и

N	a	b	c	$x1$	$x2$	Ч. к.
1	1	2	1	–1.00	–1,00	1
2	1	–2	1	1.00	1,00	1
3	1	–1	3	###.##	###.##	0
4	2	4	1	–1,71	–0,29	2
5	2	8	1	–3,87	–0,13	2
6	2	12	1	–5,92	–0,08	2
7	2	16	1	–7,94	–0,06	2
8	2	20	1	–9,95	–0,05	2

$d = b*b - 4*a*c$
$x1 = (-b - SQRT(D)) / (2*a)$
$x2 = (-b + SQRT(D)) / (2*a)$
Ч. к. =
при D > 0 : 2
при D = 0 : 1
иначе: 0

нетребовательна к ресурсам компьютера, это делало программу популярной не только в деловых расчётах, но и в «бытовых» вычислениях (например, при расчёте квартплаты или рациона питания домашних животных).

```
–0,29■
–0,13■■■■■■■■■■■■■■■■■■■■■■■■■■■
–0,08■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
–0,06■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
–0,05■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■
```

ПРОГРАММИРОВАНИЕ ЭЛЕКТРОННЫХ ТАБЛИЦ

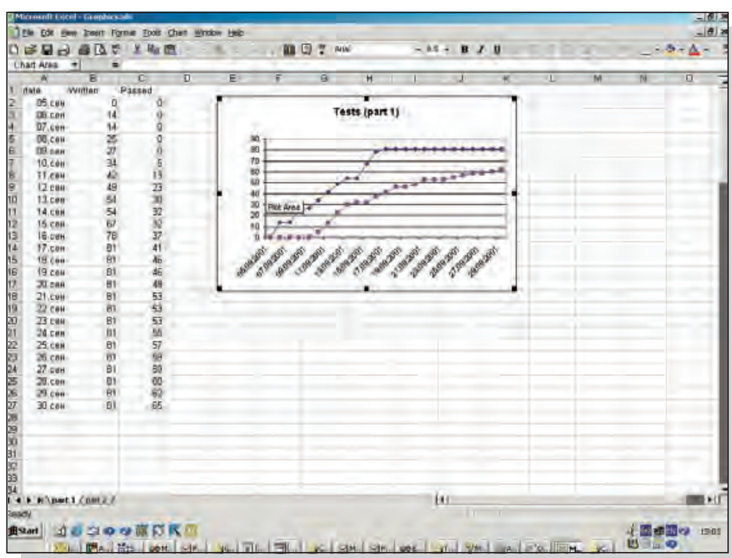
Ниже в таблице запрограммировано вычисление корней квадратного уравнения $ax^2+bx+c=0$ по его коэффициентам (колонки «a», «b» и «c»). Таблица находит не только корни (колонки «$x1$» и «$x2$»), но и их количество (колонка «Ч. к.»).

У третьего уравнения корней нет, так как дискриминант (временная переменная D в формулах) отрицателен. Поэтому вместо чисел в клетках «$x1$» и «$x2$» третьей строки — значки «#», что означает в данном случае ошибку в вычислениях: функция «квадратный корень» (в формулах обозначена как SQRT) не определена для отрицательного числа.

ГРАФИЧЕСКИЕ ВОЗМОЖНОСТИ

Автоматизация вычислений не единственное преимущество электронного представления информации. Компьютеру можно поручить и построение графических изображений.

В период использования алфавитно-цифровых дисплеев рисовать на экране графики и *гистограммы* (графики из столбиков) можно было только с помощью символов, например



Построение графиков в программе электронных таблиц Excel.

квадратиков. На рисунке показана горизонтальная гистограмма, построенная в электронной таблице. Данные для её построения взяты из колонки значений корня «*x*2» уравнений 4—8. По величине горизонтальных столбиков этой гистограммы можно увидеть, что при изменении в четвёртом уравнении коэффициента *b* с 4 до 20 с шагом 4 корень *x*2 сначала увеличивается больше, а потом всё меньше.

С появлением графических дисплеев стало возможным выделять на экране отдельные точки и строить графики. На рисунке два графика запрограммированы таким образом, что каждый из них отображает данные в одной из двух колонок таблицы. Ввод новых данных в таблицу приводит к мгновенному изменению графиков.

СЕТЕВАЯ ПОДДЕРЖКА ТАБЛИЦ

В современных электронных таблицах удаётся запрограммировать не только колонку или строку, но и любую клетку. Причём данные для вычисления содержимого клетки могут браться как из клеток данной таблицы, так и из клеток других таблиц, в том числе и расположенных на удалённых компьютерах.

Описанные рабочие группы могут находиться на удалении друг от друга, например в разных зданиях или даже в разных городах.

Электронные таблицы позволяют заполнять одновременно один и тот же экземпляр таблицы. Заполнение по сети выполняется двумя способами.

Во-первых, можно предоставить группам доступ к таблице. Тогда и программисты, и тестеры смогут заходить в данную таблицу со своих компьютеров и заполнять клетки отведённой им колонки.

Во-вторых, можно запрограммировать клетки таблицы так, чтобы электронная таблица сама считывала числа для заполнения клеток из указанного места.

Это означает, что каждая рабочая группа будет хранить значения показателей работы в каком-то определённом виде (например, файл или таблица) на своих компьютерах, а электронная таблица, открывая сводную таблицу, будет заходить на указанные ей компьютеры и считывать данные для заполнения клеток из указанного ей места.

Возможности при редактировании текстов в современных электронных таблицах практически полностью приближены к возможностям совершенного текстового процессора.

В современных электронных таблицах допустимо не только производить вычисления, но и красиво размещать текст. Например, во многих фирмах в них помещают прайс-листы, счета, анкеты и прочие подобные документы.

Практически во все электронные таблицы встроен достаточно простой язык программирования. Используя его при написании соответствующих программ, из электронных таблиц удаётся получить практически любой офисный программный продукт, вплоть до элементарной базы данных.

СИСТЕМЫ УПРАВЛЕНИЯ БАЗАМИ ДАННЫХ

Часто встречаются люди, проявляющие невнимание к своим родным и друзьям: забывают поздравить с праздником, не помнят дней рождения... Таким следует завести записную книжку, куда помимо фамилии, имени и телефона вносились бы и годовщины, дни именин и прочие важные даты. Хотя и это не поможет в решении проблемы.

Как правило, в записных книжках 28 страничек с прорезями, в которых видны «имена», состоящие из одной буквы русского алфавита. Их последовательность — это «оглавление» записной книжки. На страницу «В», например, запишем:

Вертов Д.	865-55-44
Витя (стрижка собак)	833-44-66
Вилкин Толя (9 октября)	823-12-34
Веломагазин (мастер Коля)	831-67-67

и т. д.

Для того чтобы найти друга с фамилией на букву В, просмотрим всю страничку. А вот мастера Колю отыскать нелегко: он был записан рядом с телефоном веломагазина, когда только состоялось знакомство. По истечении некоторого времени Колю хочется найти на странице «К». В какой-то момент это вызовет такое раздражение, что вы сделаете копию записи со страницы «В» на странице «К»: Коля (мастер из веломагазина) 831-67-67.

А если Коля сменит номер телефона, то придётся внести исправление на двух страницах. Самое трудное — удержать в памяти, что информация про мастера из веломагазина записана в двух местах.

Вилкина Толю с днём рождения можно опять забыть поздравить. У большинства людей нет привычки просматривать перед сном всю записную книжку, чтобы выяснить, нет ли у кого-нибудь из друзей-знакомых завтра дня рождения.

Как же решить эту проблему? Можно взять ежедневник (книжка с 366 страницами, по одной на каждый день года, с «оглавлением»: 1 января, 2 января и т. д.) и переписать всех, по желанию включая и великих людей, на соответствующую дню рождения страницу. И тогда: «Толя, поздравляю с днём рождения всех замечательных людей: тебя, Сен-Санса,

ОБ АЛГОРИТМЕ «ВЕСЫ» (И/ИЛИ)

Алгоритм позволяет эффективно выполнить операцию «объединение» над двумя упорядоченными множествами.

1. Сравниваем очередной элемент одного множества с очередным элементом второго (начиная с первых элементов).

2. Элементы не равны — переписываем меньший элемент в результирующее множество; выбираем для сравнения следующий элемент в множестве, из которого переписан элемент (в другом множестве элемент не меняется).

Иначе (т. е. равны) — переписываем элемент из одного множества в результирующее множество и выбираем из обоих множеств очередные элементы.

3. Если оба множества ещё не просмотрены до конца, то переходим к шагу 1.

Аналогичный алгоритм позволяет выполнить операцию «пересечение» над двумя упорядоченными множествами.

Муравьёва-Апостола и Джона Леннона». (Всё равно, кто будет в этом ряду, главное, чтобы родились в один день.)

Однако вручную делать это трудоёмко, лучше воспользоваться компьютером.

Какие же проблемы возникают при разработке компьютерной записной книжки?

Прежде всего утомительный последовательный просмотр страницы при поиске человека по первой букве фамилии или имени. Если разбить на страницы по 2—3 начальным буквам фамилии, то поиск убыстрится. А если на каждого человека выделить по страничке — совсем быстро.

Кроме того, нельзя найти любую существующую, хранящуюся информацию без полного просмотра всех сведений, так как мгновенно находится лишь то, для чего создано «оглавление». Например, «веломагазин» — на букву В, и совсем не очевидно, что и «мастер Коля» находится на этой же странице.

Ещё одна проблема — появление опасного дублирования (в одной записной книжке или при создании второй, третьей с другими «оглавлениями»), которое может привести к ряду ошибок: в одном месте исправили, а в других местах забыли.

Для решения этих задач и были придуманы *базы данных*. В узком смысле это специальным образом организованное хранилище информации (*файлы* базы данных), где можно не только быстро искать нужные сведения, но и добавлять, исправлять и удалять их. *Система управления базой данных* (СУБД) — это набор программ, позволяющий выполнять перечисленные операции.В широком смысле база данных — это объединённое название файлов и программ.

Чаще всего в базе данных есть файлы *данных* и файлы индексов, в которых хранится информация, необходимая для быстрого поиска.

В каждом файле данных информация хранится неразрывными порциями, которые называют *записями*. Запись, в свою очередь, состоит из *полей*. Они бывают *числовые,* состоящие только из цифр; *символьные,* состоящие из любых знаков: букв, цифр, знаков препинания; *специальные,* имеющие определённый формат: дата, деньги, номер телефона и т. д. Набор полей во всех записях одной базы данных одинаковый, и определяется он до её создания (это входит в сложный процесс — *проектирование*).

При проектировании базы данных для личного телефонного справочника в запись надо включить фамилию, имя, телефон, число и месяц рождения. Если через год окажется, что необходимы отчество, место работы и год рождения, то можно ввести дополнительные поля, которые в уже существующих записях будут иметь пустые значения.

Каждая запись получает уникальный номер, чаще всего зависящий от положения в файле. Это позволяет хранить записи в произвольном порядке, удалять их и вносить новые на освободившееся место. Номер является достаточной информацией, чтобы прочитать запись из файла данных. Однако найти необходимую запись можно, только выполняя последовательный просмотр всех записей.

Быстрый поиск возможен, когда созданы «оглавления» — *индексы*. Нужен индекс для поля или нет, определяется во время проектирования базы данных. Например, если не предполагается искать знакомых по имени, то нет смысла создавать индекс для поля «Имя».

Можно считать, что индекс состоит из пар: значение поля, уникальный номер записи. Первая часть в паре называется *ключом,* вторая — *ссылкой*.

Для того чтобы отыскать человека с фамилией Страхов, достаточно просмотреть перечень ключей индекса «Фа-

Фамилия	
Страхов	**172**
Староверов	021
Богданова	015
Староверов	127
Страхова	044
Богданов	201
Страхов	**135**
Стойнова	413
Староверов	093

Ключ

Ссылка

милия», найти уникальные номера записей в файле данных (172 и 135) и по ним прочитать записи про двух человек.

Если в базе данных хранится информация про 4 тыс. человек, то при каждом поиске может потребоваться «пробежать» в индексе по всем ключам, т. е. по 4 тыс. записей. Это гораздо быстрее, чем просматривать такое же количество записей в файле данных.

Для ускорения поиска в индексе *(файле индексов)* нужно хранить пары ключ — ссылка в порядке возрастания ключей и повторяющиеся ключи заменить на пару ключ — список ссылок. При этом список ссылок тоже расположить по возрастающей:

Фамилия	
.....	
Богданов	201
Богданова	015
.....	
.....	
Староверов	021 093 127
Стойнова	413
Страхов	**135 172**
Страхова	044
.....	

Даже если не применять никаких специальных алгоритмов поиска, время его всё равно сократится в среднем в два раза.

Но если воспользоваться известным алгоритмом *«деление пополам» (бинарный поиск),* то нужно будет прочитать не более 12 ключей, чтобы найти искомый среди 4 тыс. ($11<\log_2 4000<12$), т. е. ускорение произойдёт в 340 раз!

Хранение ключей, упорядоченных по возрастанию, в индексе позволяет легко реализовать поиск по диапазону значений, больше или меньше указанного значения.

Пусть индекс «День рождения», полученный из полей «Число» и «Месяц» (в ключе сначала указывается месяц, потом число), выглядит так:

День рождения	
....	
0228	244
0321	056 084 142
0329	**135**
0331	**044 179**
0402	**127 278 299 305 432**
0419	021 093
0423	112

21 марта

19 апреля

Для ответа на вопрос «У кого день рождения с 25 марта по 5 апреля?» нужно найти все ключи с 0325 по 0405 в индексе «День рождения». Прямым поиском СУБД найдёт 0329 (ближайший больший ключ от начального значения), а затем, последовательно читая ключи, будет собирать ссылки, выполняя операцию «объединение» над списками ссылок, чтобы исключить дублирование, пока не дойдёт до ключа 0419 (ближайший больший ключ от конечного значения). В результате получится следующий список ссылок: 044 127 135 179 278 299 305 432, с помощью которого можно прочитать требуемые записи из файла данных.

Другой вариант поиска — по нескольким индексам находят записи, удовлетворяющие нескольким вопросам одновременно. Например, запрос «Страховы, родившиеся в марте». Для этого в индексе «Фамилия» обнаруживают ключи, начинающиеся на «Стра», а в индексе «День рождения» ключи от 0301 до 0331.

Тогда СУБД по первому индексу найдёт следующий список ссылок: 044 135 172, по второму индексу получит такой список ссылок: 044 056 084 135 142 179; затем, выполнив операцию «пересечение» между списками, получит в итоге ответ: 044 135.

Добавление/удаление записи из файла базы данных сопровождается изменением индексов. Если в файл данных под номером, например, 516 вносят запись «Стаханов, родившийся 4 мая», то по всем полям, имеющим индексы, создаются пары ключ — ссылка, сохраняя порядок:

Фамилия	
.....	
Богданов	201
Богданова	015
.....	
.....	
Староверов	021 093 127
Стойнова	413
Стаханов	**516**
Страхов	135 172
Страхова	044
.....	

День рождения	
....	
0228	244
0321	056 084 142
0329	135
0331	044 179
0402	127 278 299 305 432
0419	021 093
0423	112
0405	**516**
....	

А если ключ уже был в индексе, то к нему добавляется только ссылка в список ссылок.

Так же как при добавлении, при удалении пары ключ — ссылка удаляются из соответствующих индексов. Если при ключе имеется лишь одна ссылка, тогда исключается вся пара, в противном случае — только ссылка из списка ссылок. Например, запись 021 — «Староверов», дата рождения 19 апреля:

Фамилия	
.....	
Богданов	201
Богданова	015
.....	
.....	
Староверов	093 127
Стойнова	413
Стаханов	516
Страхов	135 172
Страхова	044
.....	

Здесь была ссылка на запись 021

День рождения	
....	
0228	244
0321	056 084 142
0329	135
0331	044 179
0402	127 278 299 305 432
0419	093
0423	112
0405	516
....	

И здесь была ссылка на запись 021

И лишь после этого удаляется запись из файла данных.

Процесс изменения файла индексов обычно продолжительнее, чем добавление записи в файл данных.

Давайте продемонстрируем это на таком примере — ошибочно введённой фамилии Староверов вместо правильной — Старовер. Если СУБД каждый индекс записывает в отдельный файл, то изменение одного поля в файле данных (замена фамилии Староверов на Старовер) приведёт к изменениям в файле индексов «Фамилия»:

Фамилия	
.....	
Староверов	**093**
Староверов	127
Стойнова	413
Стуханов	516
.....	

Понадобится добавить новую пару и произвести изменения в существующей паре. Объём файла индексов может быть гораздо больше объёма файла данных.

Есть иной способ поиска в упорядоченном файле, основанный на создании групп смежных записей. Первая запись в такой группе хранится в таблице, которая является указателем к отдельным частям файла, как бы привычным «оглавлением». Чтобы найти запись «Староверов» (поиск по букве «С»), можно попасть на нужную часть индексного файла, а дальше осуществить бинарный поиск:

Фамилия	
.....	
Путников	912
Самарин	012
.....	127
Староверов	**127**
.....	
Сытников	991
.....	

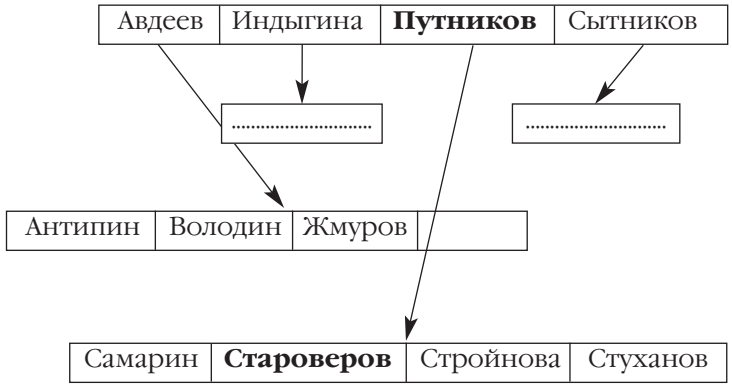
Группа «С»

Как бы ни была хороша скорость бинарного поиска, она всё-таки не даёт подобных показателей при добавлении/удалении пар из файла индексов, так как требует сохранять порядок, на что затрачиваются существенные временные ресурсы компьютера. Целесообразнее использовать нелинейное хранение индексов в виде B-дерева.

B-дерево имеет вид перевёрнутого дерева, листья которого находятся внизу, а единственный корень — вверху. Промежутки с ответвлениями — между корнем и листьями — называют *узлами* дерева. Листья — узлы нижнего уровня. Корень — узел верхнего уровня. В узлах B-дерева располагаются ключи, например «Фамилии».

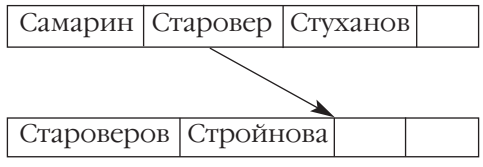
В данном примере B-дерево получилось двухуровневым. Верхний уровень содержит ключи, которые являются точками раздела. Каждый ключ в этом узле указывает на узел одним уровнем ниже. Ключ «Авдеев» является ссылкой на файл данных и, кроме того, указывает, что ключи от «Авдеев» до «Индыгина» находятся на следующем уровне. При поиске записи «Староверов» в корне дерева сначала будет обнаружено, что ключ «Путников» указывает на следующий уровень B-дерева, в узлах которого и надо искать ключ «Староверов» (так как «Староверов» находится по алфавиту между «Путников» и «Сытников»), а затем на следующем узле и будет обнаружен ключ «Староверов», являющийся ссылкой на запись в файле данных.

Добавлять или удалять элементы из B-дерева может оказаться как простой, так и сложной задачей. Например, новый ключ «Жучкин» добавить



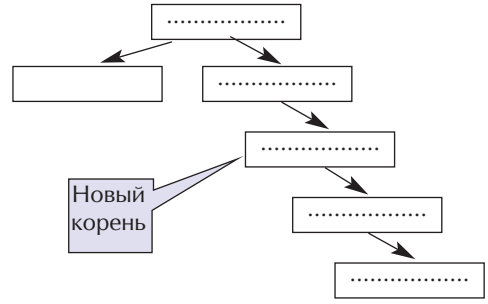
легко, так как место рядом с ключом «Жмуров» свободно. Однако при добавлении ключа «Старовер» потребуется создать ещё один уровень узлов, так как на этом узле больше нет места. Узел

придётся делить на верхний и нижний:

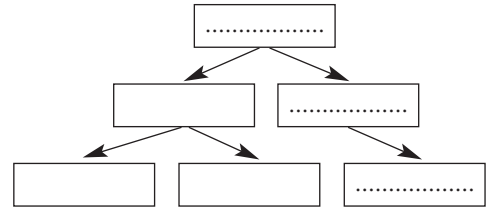


В-дерево надо поддерживать в сбалансированном состоянии, чтобы его «ветки» «росли» равномерно, иначе поиск потребует в некоторых случаях больше времени.

В изображённом дереве хочется «взяться» за «новый корень» (выделенный узел на рисунке) и «потянуть»,



тогда оно слегка изменится в левой части:



Поиск в таком дереве будет наиболее эффективен при любых значениях искомого ключа. Существуют алгоритмы, позволяющие преобразовывать несбалансированные деревья в сбалансированные.

ХЕШИРОВАНИЕ

Так же как и в хеш-реализации структуры данных множества, при поиске в файле данных удобно использовать *хеширование*. Каждой записи ставится в соответствие некоторое число, которое эффективно вычисляется с помощью какой-нибудь простой процедуры. Для построения индексов фамилий достаточно пронумеровать все буквы алфавита, чтобы любой букве соответствовало число от 1 до 33:

А—1; Б—2; В—3 ... Я—33.

И дальше фамилия рассматривается как последовательность кодов букв (от 1 до 33). Например, фамилия Бабенко в таблице запишется так:

2	1	2	6	15	12	16

А «длина» будет равна 7, по числу букв.

```
алг цел индекс (целтаб
фамилия[1:20], цел длина)
  дано
  надо
нач цел K
| знач:=0
| нц для Кот 1 до длина
| | знач := знач * 33 + фамилия [K]
| кц
кон
```

Полученное значение — уникальное число, которое и можно считать ссылкой на запись. Однако такое число будет огромным. Для приведённой в примере фамилии Бабенко индекс составит порядка 10^{10}! Данный подход требует невероятно больших объёмов памяти, что неэффективно, так как не все комбинации букв составляют фамилии. Обычно для решения этой проблемы используется алгоритм вычисления псевдоуникального числа, который и называют *хеш-функцией*. Удобно считать сумму кодов букв в фамилии, вычисленной по модулю размера таблицы *N*, которая отведена под фамилии *(хеш-таблицы)*:

```
алг цел хеш_функция (целтаб
фамилия[1:20], цел длина, N)
  дано
  надо
нач цел K
| знач:=0
| нц для Кот 1 до длина
| | знач := mod (знач + фамилия [K],N)
| кц
кон
```

(Функция mod(*x*,*y*) возвращает остаток от деления нацело *x* на *y*, её ещё называют модулем, а записывают выражение так: *y* mod *y*.). Чтобы различные ключи давали совпадения не слишком часто (ведь это псевдоуникальные, а не уникальные числа), необходимо оставлять около четверти хеш-таблицы пустой. Если всё-таки возникли совпадения — *коллизии*, то можно подобрать временно свободное место в той же таблице, последовательно перебирая места. Для большей эффективности удобно провести вторичное хеширование, выбрав вторичную хеш-функцию аналогично первой:

```
алг цел вторичная хеш_функция
(целтаб фамилия[1:20], цел длина,
N, M, шаг)
  дано
  надо
нач цел K
| знач:=0
| нц для Кот 1 до длина
| | знач := mod (знач + фамилия [K],M)
| кц
| знач := mod (знач * шаг + хеш_функция
|   (фамилия, длина, M), N)
кон
```

Число *M* выбирают взаимно простым с *N*, обычно $M = N-2$. Если при хешировании возникла коллизия, то вычисляют значение вторичной хеш-функции с «шагом», равным 1. Если снова появляется коллизия, то «шаг» устанавливают 2 и т. д. Идея состоит в том, чтобы избегать коллизий «по модулю», беря в качестве последовательных индексов хеш + вторичная хеш, хеш + 2 вторичная хеш, хеш + 3 вторичная хеш ...

При хешировании близко расположенные по алфавиту фамилии могут быть произвольно разбросаны по таблице, вообще говоря, они не обязаны быть упорядочены по алфавиту. В результате поиска записи «Староверов» невозможно получить ответ на запросы о близких фамилиях типа Старовер или Стройнова.

РЕЛЯЦИОННЫЕ БАЗЫ ДАННЫХ

Небольшие группы связанных данных обычно называют записью, например, в файле данных «Продукт-цена» группу, состоящую из наименования продукта, образует запись. Элемент записи называют полем.

Три основных типа организации данных определяют характер связей между записями, элементами и ключами. Фактически с появлением компьютеров в базах данных использовалась иерархическая *модель*. Её можно проиллюстрировать на примере продуктового магазина.

Первый уровень включает таблицу отделов, где в качестве главного элемента выступает категория продуктов. Выбрав нужную категорию, например, «Молочные продукты», можно перейти к таблице следующего, второго, уровня.

Второй уровень включает таблицы, перечисляющие названия продуктов по отделам, без детализации. Так, «Молоко» относится к отделу №1 — «Молочные продукты», но конкретное наименование и цена хранятся в таблицах следующего уровня иерархии. Подобная структура базы данных удобна продавцу, но не подходит для покупателя, которого в первую очередь интересуют цены на продукты. Только «пройдя» все уровни иерархии, база данных может «дать ответ», например, сколько стоит молоко «Домик в деревне» 1,5%. Пока не будут добавлены указатели между первым и третьим уровнями, минующие иерархическую структуру базы данных, получить ответ на запрос о цене продукта сложно (с точки зрения времени выполнения операции).

Простота организации, при которой база данных может быстро дать ответ на определённые вопросы позволила иерархическим базам данных получить широкое распространение на «медленных» ЭВМ. Более того, на этапе проектирования базы данных нельзя предусмотреть, на какие запросы будет неэффективно получать ответы.

Отдел	Категория продуктов
1	Молочные продукты
2	Мясные продукты
3	Бакалея
4	Рыбные продукты
5	Хлебобулочные изделия

Мясные продукты	Мясо Полуфабрикаты Птица …

Молочные продукты	Молоко Сметана Творог Сыр Кефир Масло …

Продукт	Цена
«Домик в деревне» 0,5%	13 р.
«Домик в деревне» 1,5%	15 р.
…	…
«Весёлый молочник» 3,2%	19 р.15 к.
…	…

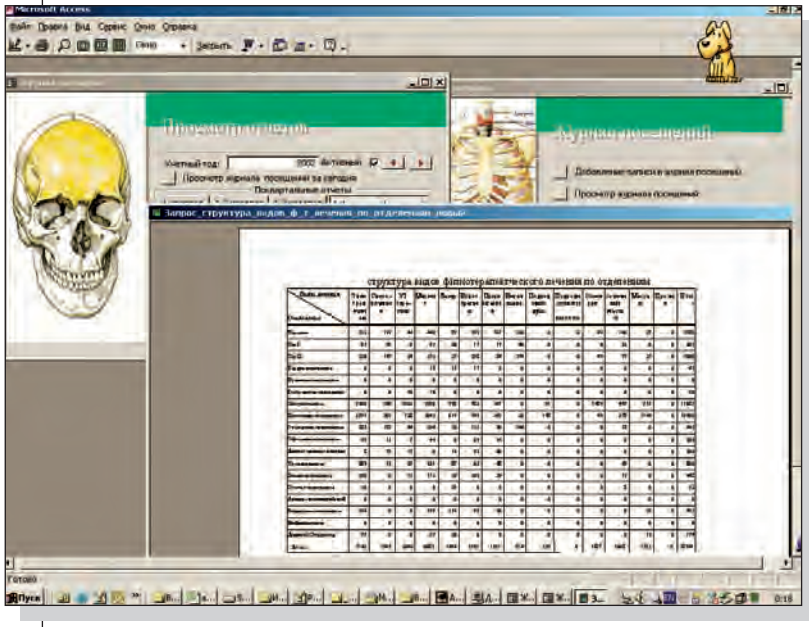


Фото с экрана базы данных.

ЧТО ТАКОЕ ОДИН ГИГАБАЙТ

В одном гигабайте 2^{30} байт, или, как пишут в современных книгах (считая, что 1кбайт = 1000 байт, а не 1024), около миллиарда байтов памяти. В такой памяти можно хранить:

- 500 тыс. страниц текста, т. е. около 1 тысячи романов, — целая библиотека;
- около 1 тысячи цветных слайдов высочайшего качества;
- до 5 тыс. цветных цифровых фотографий растра 1280 × 1024 в формате JPEG;
- аудиозапись более чем 100-часовой речи (при этом запись равноценна по качеству телефонному разговору);
- более чем 10 ч музыкальной записи, использующей алгоритмы сжатия MP 3. Она неотличима на слух от записи на компакт-диске (эта технология используется в FM-радиовещании);
- музыкальную запись в том же формате MP 3, но продолжительностью более 20 ч. Правда, за счёт увеличения объёма качество ухудшается и соответствует качеству записи хорошего кассетного магнитофона;
- без малого 2-часовой музыкальный фрагмент CD-качества;
- почти 15-секундный цветной фильм высочайшего качества записи без использования алгоритмов сжатия;
- более чем 2-часовую видеозапись с использованием алгоритмов сжатия изображения и звука MPEG (качество записи лишь немного уступает телевизионному);
- протокол операций с банковским счётом крупной фирмы более чем за 5 тыс. лет.

Стоимость устройства, способного хранить около 1Гбайт информации (для сравнения: на трёх CD ROM содержится 2Гбайт информации), в настоящее время не превышает 1 доллара США. А стоимость аренды пространства для хранения 1Гбайт в компьютерных сетях составляет всего несколько долларов США в год. Можно ожидать, что через 10—15 лет всё сказанное останется верным , если место гигабайта займёт терабайт (1Тбайт = = 2^{40} байт).

Сетевая модель отличается от иерархической большей гибкостью. В ней допустимы множественные связи между файлами, позволяющие сразу получить доступ к нужному файлу. В нашем примере про магазин можно построить ссылки между первым и третьим уровнями. Однако для этого требуется использование целого набора специальных программистских приёмов (маленьких хитростей, существенно усложняющих базу данных), так как при разработке надо в деталях представлять всю усложнившуюся (с межуровневыми связями) структуру базы. Меткая характеристика «Сетевая база — это самый верный способ потерять данные» очень точно характеризует сетевую базу данных как весьма сложную структуру.

Считается, что структура *реляционной базы данных* в отношении запросов обладает большей гибкостью. Здесь отсутствует иерархия элементов, при поиске информации в базе данных всякий элемент используется в качестве *ключа*. Под *записью* понимается уже не совокупность отдельных элементов (один из которых считается главным), а строка таблицы.

В нашем примере таблица уже выглядит так:

Продукт	Отд.	Категор. продуктов	Цена	Единица
«Домик в деревне» 0,5%	1	Молоч. продукты	13 р.	Пакет 1л
«Домик в деревне» 1,5%	1	Молоч. продукты	15 р.	Пакет 1л
«Весёлый молочник» 3,2%	1	Молоч. продукты	19 р.	Пакет 1л
Камбала	4	Рыбные продукты	97 р.	1 кг
Кальмары	4	Рыбные продукты	68 р.	Уп. 0,5 кг

Запись «“Домик в деревне” 0,5% Молочные продукты 13 р. Пакет 1л» можно понимать как отношение между элементами «Продукт», «Отдел», «Категория продуктов», «Цена» и «Единица». Это отношение в дальнейшем может пополняться такими элементами, как, например, «Имя поставщика»,

Представление о реляционной базе данных как о наборе таблиц неточно (в нём не учитываются так называемые вторичные индексы). В реальных базах данных записи достаточно велики и их много, поэтому всю таблицу не удаётся распечатать на одном листе. Работа с базами данных на компьютере отличается от работы с таблицами на бумаге и больше напоминает работу с картотекой: записи, как и бумажные карточки, легко переставлять, добавлять и удалять.

Эдгар Фрэнк Кодд

«Срок хранения», «Объём по складу» и т. п.

Из полного набора отношений в таблице можно легко построить новую, упрощённую таблицу отношений, например о продуктах в молочном отделе:

«Домик в деревне» 0,5%	13 р.
«Домик в деревне» 1,5%	15 р.
«Весёлый молочник» 3,2%	19 р.

Или дать сведения о стоимости продукта:

Камбала	97 р.	1 кг

Реляционный принцип организации баз данных наиболее удобен в том случае, когда нельзя заранее сказать о характере запросов.

В простейших случаях (при реализации телефонного справочника или в примере с продуктовым магазином) для представления данных достаточно одной таблицы, в более сложных (в электронном депозитарии — хранилище информации об акционерах различных предприятий) может потребоваться несколько таблиц, связанных между собой с помощью ссылок.

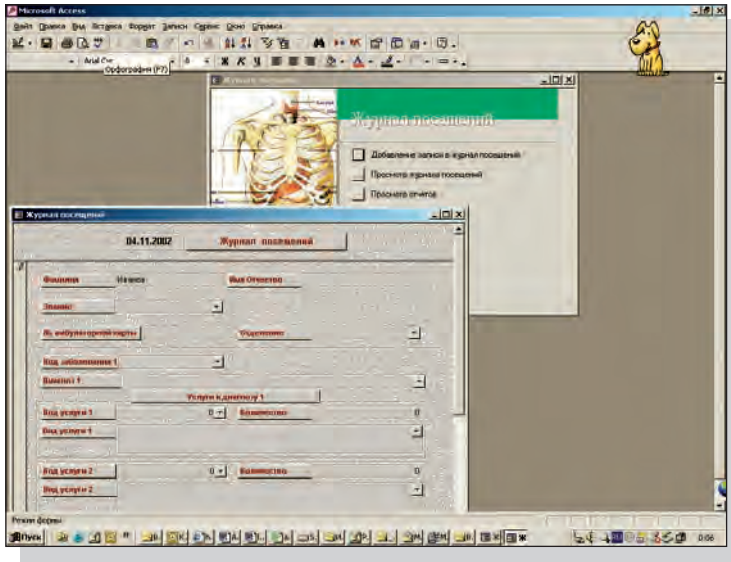


Фото с экрана базы данных

Реляционные базы данных, наиболее широко использующиеся в персональных компьютерах, были впервые предложены сотрудником фирмы IBM Э. Ф. Коддом в 1970 г. Математик разработал принципы этой модели, основываясь на теории множеств.

Вот как в общих чертах они выглядят:

• Данные представляют собой упорядоченный набор в виде строк, называемых *отношениями*.

• Строки обязательно отличаются друг от друга хотя бы значением одного поля.

• Операции выполняются целиком над всем отношением, результат также является отношением. Этот принцип называется замыканием.

Разработчики информационно-поисковых систем обычно уделяют большое внимание тому, как хранимые данные будут показаны на экране при поиске, добавлении или изменении информации. Формат представления данных на экране компьютера называют экранной формой. Как правило, требуется несколько экранных форм. Например, при внесении записи в базу данных продуктового магазина изменений удобно видеть лишь одну строку, но в развёрнутом виде, а при просмотре списка продуктов отдела — несколько строк с краткой информацией о каждом товаре.

Экранную форму можно представить себе как окно, через которое человек увидит какую-то часть базы данных. В него могут быть вставлены «цветные стёкла»: некоторые поля изображаются на экране в немного изменённом виде (не так, как они хранятся в базе данных, а как их удобнее видеть человеку).

Система управления базами данных, показывающая записи только в том порядке, в каком они вводились, практически бесполезна. Поэтому все

СУБД позволяют задавать порядок для просмотра записей. Числовые поля, как и текстовые, можно сравнивать. В телефонной книге вполне естественно расположить записи по алфавиту: из двух слов раньше идёт слово, у которого первая буква «меньше» (т. е. стоит в алфавите раньше). Сравнивать даты совсем просто — одна дата считается «меньше» другой, если относится к более раннему времени.

В реляционных базах данных для задания порядка следования записей применяются ключевые поля (как правило, первое поле или несколько первых полей записи). С точки зрения пользователя, записи располагаются в таком порядке, что содержимое ключевых полей возрастает от начала базы к её концу.

Если ключевых полей несколько, то при определении порядка записей используется тот же метод, что и при лексикографическом сравнении текстов. Вначале записи упорядочиваются по содержимому первого ключевого поля; записи с одинаковым содержимым этого поля расставляются по второму ключевому полю и т. д.

Возможность быстро найти нужную информацию и посмотреть её на экране — это основное достоинство баз данных, но для работы бывает необходимо напечатать отчёт о состоянии базы данных. Размеры экрана ограниченны, и, глядя на него, трудно получить представление обо всей базе данных. Если в экранную форму включается информация об одной или нескольких записях, то в отчёт чаще всего входят данные о большой группе записей (или даже обо всех) в базе данных. Сюда также обычно добавляются некоторые автоматически подсчитанные данные.

Все информационные системы позволяют при поиске и составлении отчётов использовать достаточно сложные условия. Очень часто запись должна удовлетворять сразу нескольким простым условиям. Точно так же, как с использованием арифметических операций составляются арифметические выражения, с помощью логических операций («и», «или», «не») можно составлять логические выражения (составные условия). Пусть требуется найти все продукты с названием «Сметана». Для поиска «сметаны» в отделе «Молоко» добавим, что её цена не превышает 15 р.

В реляционной модели данные представлены отношениями, но нигде не фиксируется, как данные физически располагаются в памяти. Отделение концептуального уровня от логического (представления данных) было революционным переворотом в программировании баз данных, так как ранее программирование сводилось к управлению устройствами (дисками), предназначенными для хранения данных.

При составлении отчётов часто употребляется специальный язык запросов. Запрос о «сметане» может выглядеть примерно так:

```
ВЫБРАТЬ ВСЕ ЗАПИСИ ИЗ ТАБЛИЦЫ продуктовый магазин ПРИ ЗНАЧЕНИИ ПОЛЯ продукт= «сметана» и ЗНАЧЕНИИ ПОЛЯ цена<15.
```

Таблица

Камбала	97 р.	1 кг

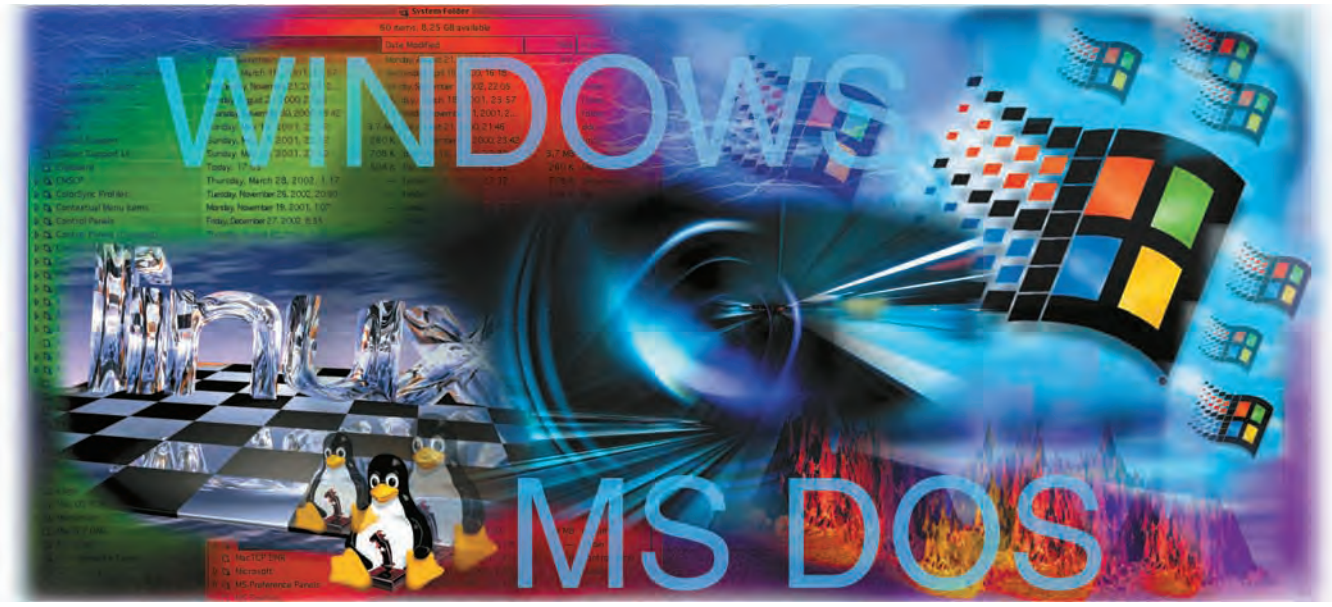
будет результатом такого запроса:

```
ВЫБРАТЬ ПОЛЯ продукт, цена, единица ИЗ ТАБЛИЦЫ продуктовый магазин ПРИ ЗНАЧЕНИИ ПОЛЯ продукт= «камбала».
```

Если бы оказалось, что в базе данных не одна запись «камбала», то и в запросе получилось бы несколько строк. В языке запроса важно указать то, что надо получить, а не то, как это сделать.

Конечно, в реальных базах данных гораздо сложнее и процесс проектирования базы данных, и получение отчётов, и сами формы и запросы. Чтобы создать настоящую информационно-поисковую систему, нужно ещё многое выполнить, ведь от того, как данные распределены по таблицам, зависит эффективность работы всей системы. В простейших случаях современные средства проектирования позволяют создать небольшие системы, основываясь только на общих принципах построения реляционных баз данных. Например, чтобы составить личный каталог аудиокассет, не надо уметь программировать и знать сложные языки запросов.

На практике в отчёты обычно требуется включать более сложную статистику, нежели просто сумма данных в каком-то поле. Поэтому в системах для каждого типа отчёта может потребоваться написать свою программу его генерации, в большинстве случаев достаточно простую.



ОПЕРАЦИОННАЯ СИСТЕМА КОМПЬЮТЕРА

ВОЗНИКНОВЕНИЕ ОПЕРАЦИОННЫХ СИСТЕМ

Загрузка колоды перфокарт для работы ЭВМ в пакетном режиме.

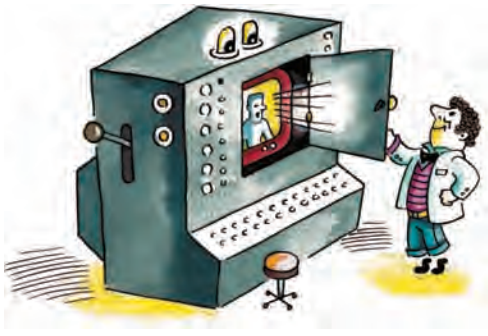
Первые ЭВМ использовались для решения исключительно математических задач, а программами для них служили написанные в машинных кодах вычислительные алгоритмы. Программисту при кодировании программ приходилось самостоятельно управлять ЭВМ и обеспечивать выполнение своей программы. Однако с развитием электроники аппаратура совершенствовалась, а в программах помимо вычислений возникали алгоритмы ввода-вывода информации, которые выделялись в самостоятельные подпрограммы. Так появились *библиотеки ввода-вывода* — набор служебных подпрограмм, облегчающих программирование ЭВМ. Поскольку стоимость машинного времени была очень высокой, возникла потребность в одновременном выполнении нескольких программ: когда одна программа вы-

полняла операции ввода-вывода, другая занимала процессор ЭВМ вычислениями, и наоборот. Выполняющуюся в ЭВМ программу и данные, с которыми программа работала, назвали *заданием,* а одновременный прогон нескольких программ — *очередью заданий.*

В первое время алгоритмы *переключения* заданий на процессоре включались прямо в сами программы в виде подпрограмм. Набор таких подпрограмм назывался *монитором* или *супервизором.* Кроме того, в него входили библиотеки ввода-вывода, подпрограммы распределения оперативной памяти — в общем, все алгоритмы, не имеющие прямого отношения к задачам вычислений.

Ещё одна проблема заключалась в том, что программы, выполняющиеся в ЭВМ, часто содержат ошибки, в результате которых программа может постоянно занимать процессор («зависать») или ошибочно записывать в оперативную память, где размещаются другие программы, результаты своей работы. Пока программа одна, это ещё терпимо, однако если программ несколько, сбой в одной из них приводит к сбою всего набора программ. Следует помнить, что развитие вычислительной техники шло по пути от ЭВМ для конкретной задачи до универсальной ЭВМ для многих задач, поэтому возникшая проблема требовала неотложного решения.

Выход был найден в создании специальных аппаратных механизмов, защищающих память программ от случайного доступа со стороны других программ. Управление этими механизмами уже нельзя включать в сами программы: память, где размещаются алгоритмы, управляющие защитой, должна быть сама защищена от программ, иначе сбой в исполнении программы может повредить и её. Возникла потребность в специальной программе, которая управляла бы механизмами защиты памяти. Так появился *резидентный монитор* — к монитору добавили подпрограммы управления защитой памяти и оформили всё в виде отдельной программы.



ЧТО ТАКОЕ ОПЕРАЦИОННАЯ СИСТЕМА

Представление о том, что такое *операционная система,* эволюционировало во времени. Вот определение, взятое из книги Г. Катцана «Операционные системы. Прагматический подход» (1973 г.): «Операционная система — это организованный набор программ и данных, разработанный специально для управления ресурсами вычислительной системы, облегчения создания программ для ЭВМ и также для управления последними».

Определение Г. Катцана — взгляд на операционную систему с точки зрения аппаратуры: есть аппаратура ЭВМ, и на ней нужно выполнять программы, а для облегчения этой задачи существует операционная система. Современный взгляд на операционную систему включает также точку зрения пользователя как потребителя ресурсов ЭВМ.

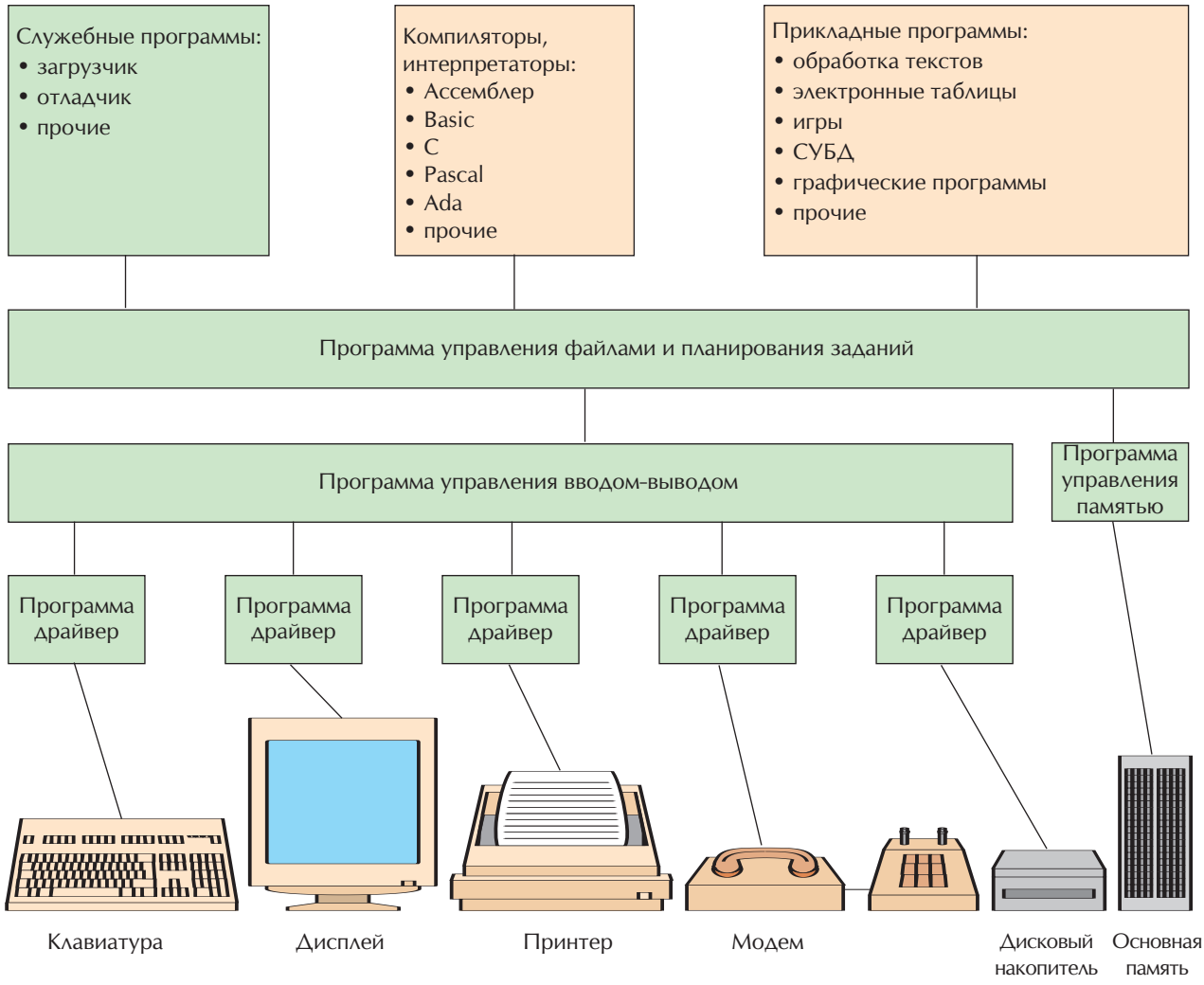
«Операционная система — это программа, выступающая посредником между пользователем и аппаратурой ЭВМ. Назначение операционной системы — обеспечить пользователю среду для удобного и эффективного выполнения программ» (А. Зильбершатц, П. Б. Гэлвин. «Принципы операционных систем». 1998 г.).

«Операционная система — это программа, управляющая выполнением прикладных программ и выступающая в качестве интерфейса между пользователем и аппаратурой ЭВМ. Можно считать, что к операционной системе предъявляются три требования: удобство, эффективность и возможность развития» (У. Сталлингс. «Операционные системы». 1995 г.).

Наконец, у Эндрю Таненбаума в книге «Современные операционные системы» (1992 г.) даны два определения операционной системы: одно с точки зрения аппаратуры, другое с точки зрения пользователя.

«Операционная система как расширенная ЭВМ. Функция операционной системы — предоставить пользователю эквивалент расширенной ЭВМ, или виртуальной ЭВМ, которую удобнее программировать, чем основную ЭВМ...

Операционная система как администратор ресурсов. Цель операционной системы — отслеживать, кто использует данный ресурс, удовлетворять запросы к ресурсам и протоколировать их использование, а также разрешать конфликты при доступе к ресурсам различных задач и пользователей».



Резидентный монитор — это уже зачатки операционной системы в том виде, в котором её воспринимают сегодня. Теперь прикладные программы стали содержать только реализацию своего алгоритма, обращаясь за вспомогательными алгоритмами к монитору, используя при этом фиксированный набор правил, называющийся *прикладным программным интерфейсом*. Впервые стало возможно говорить о *переносимости* прикладных программ с одной ЭВМ на другую. С возникновения программного интерфейса начался процесс отделения правил работы с ЭВМ от аппаратуры, а ЭВМ стали «приближаться» к человеку.

Программный интерфейс позволил создавать абстрактные понятия, совершенно невыразимые в аппаратуре. Например, появилось понятие *файла* и соответственно *файловой системы* — набора интерфейсов и структур данных для организации хранения информации. Возникло понятие *процесса* (или *задачи*) как потребителя ресурсов и единицы работы ЭВМ.

Резидентный монитор всегда находится в оперативной памяти ЭВМ (потому его и назвали резидентным), поскольку его подпрограммы нужны для выполнения всех остальных программ, для которых стали применять термины «прикладные программы» или «пользовательские программы».

Конечно, понятия «файл» и «процесс» существовали задолго до появления резидентного монитора. Однако возникновение его как отдельной специальной программы позволило зафиксировать правила работы прикладных программ со служебными подпрограммами. Действительно, пока программы содержали монитор внутри себя, всегда существовала возможность заменить в момент разработки программы одну подпрограмму работы с файлами на другую (например, более эффективную), тем самым, явно или неявно, переопределив понятие файла. Резидентный монитор, зафиксировав правила работы с файлами и процессами, заодно и определил, что в точности эти понятия обозначают.

Следующий виток развития операционных систем характеризуется увеличением набора правил работы и как следствие усложнением самих операционных систем. К резидентному монитору были добавлены вспомогательные программы, облегчающие выполнение частых операций: копирование файлов, редактирование текстов, компиляция программ с языка программирования в машинный код и др. Всё вместе назвали *операционной системой,* а набор вспомогательных программ — редакторы, компиляторы, программы работы с файлами — *системными утилитами*. При этом грань, разделяющая системные утилиты и прикладные программы, оказалась довольно условной. Термин «резидентный монитор» трансформировался в *ядро операционной системы,* причём в ядре стали выделять несколько важных частей:

— планировщик процессов,
— подсистема управления памятью,
— файловая система,
— подсистема управления вводом-выводом,
— программный интерфейс.

К процессу относят аппаратные регистры компьютера, включая счётчик программ; описание адресного пространства процесса — участков памяти ЭВМ, которые процесс может читать, писать или выполнять; описание файлов, с которыми процесс работает, и всю другую информацию, необходимую для выполнения программы.

ПЛАНИРОВЩИК ПРОЦЕССОВ

Сердцем подсистемы управления процессами является *планировщик процессов,* реализующий алгоритм переключения процессов, т. е. замены процесса, выполняющегося на процессоре, на другой процесс, ожидавший своей очереди. Обычно для обозначения этой процедуры употребляют сложный термин «*переключение контекста процесса*». Работа алгоритма такого переключения контекста зависит от многих параметров, определяющих, какому процессу и когда именно должен быть предоставлен процессор.

Для иллюстрации работы планировщика, как и для понимания работы операционной системы в целом, важно различать *стратегию* и *механизм*. Стратегия работы определяет, что надо делать, а механизм — как надо делать.

Шахматисты тоже делят время матча между собой.

Понятие кванта предполагает неделимость. Квант выполнения неделим в том смысле, что, пока процесс, выполняющийся на процессоре, не использовал отведённый ему планировщиком срок, тот не отдаст процессор другому процессу.

Стратегия работы планировщика, например, может заключаться в том, что процесс явно (вызовом специальной подпрограммы) или неявно (при выполнении операции ввода-вывода, при завершении программы) обращается к операционной системе с указанием, что процессор свободен и его можно использовать для выполнения других процессов. В этом случае говорят, что операционная система реализует *кооперативную многозадачность*, т. е. процессы, возможно, «голосованием», сообща решают, какой из них будет выполняться следующим. Кооперативная многозадачность использовалась, например, в мониторах первых ЭВМ, а также в операционных системах персональных компьютеров — MS DOS и Windows версий 3.1.

Компьютер Alpha фирмы DEC.

Другая стратегия — *вытесняющая многозадачность* — состоит в том, чтобы каждому процессу гарантированно обеспечить выполнение на процессоре, не обделяя никого. Планировщик может прервать выполнение процесса и вытеснить его другим процессом. При этом он выделяет фиксированный промежуток времени, называемый *квантом выполнения*, в течение которого процесс непрерывно выполняется на процессоре. Обычно квант выполнения подбирается так, чтобы процесс не слишком долго занимал процессор, но мог за отведённое время сделать что-нибудь полезное.

Для современных процессоров типичное значение кванта выполнения составляет 10 мс. Только для очень быстрых процессоров, например для процессора Alpha фирмы DEC, квант равен 1 мс.

После истечения кванта планировщик выполняет *алгоритм перепланирования*, выясняя, какой процесс будет выполняться на процессоре следующим.

В операционных системах *разделения времени* все процессы считаются равноправными и по очереди получают свой квант. Для пользователя это выглядит так, будто процессор выполняет все процессы одновременно. Такой алгоритм называется *алгоритмом перепланирования по круговой системе* (*англ.* round robin scheduling).

Достижение равноправности процессов в алгоритме перепланирования не такая простая задача, как может показаться на первый взгляд. Пусть один процесс — это текстовый редактор, а другой — программа вычисления значения π с точностью до миллионного знака после запятой. С точки зрения операционной системы алгоритм работы текстового редактора и алгоритм вычисления устроены примерно так:

```
алг Текстовый редактор
нач
| нц пока не нажата клавиша выхода
| | ждать ввода символа с клавиатуры
| | обработать символ
| | изобразить символ на экране
| кц
кон
```

```
алг Вычисление π
нач
| нц пока не определили число π
| | вычислять
| кц
кон
```

Таким образом, один процесс, выполняющий вычисления, всё время занимает процессор или, учитывая наличие другого процесса, хочет его занять (постоянно борется за ресурс процессора). Другой процесс — текстовый редактор — в основном выполняет операцию ввода и конкурирует за процессор от случая к случаю. Действительно, даже профессиональная машинистка набирает текст со скоростью 600 символов в минуту, или 1 символ в 100 мс. Если квант выполнения составляет 10 мс, то текстовый редактор, ожидая ввода очередного символа, фактически не конкурирует за процессор в течение 10 квантов. Когда процессы равноправны, планировщик отдаст один квант выполнения редактору, после чего отдаст следующий квант процессу, вычисляющему π. Установлено, что человек считает, будто символ появляется на экране мгновенно, если задержка между вводом и изображением символа не превышает 20 мс (или два кванта выполнения). Если редактор не успеет обработать символ за один квант, ему придётся ждать времени на процессоре ещё один квант. В результате человек ощутит задержки при вводе символов.

Чтобы это исправить, следует отдать процессу-редактору более одного кванта (справедливо отдать десять — ровно столько времени он не выполнялся на процессоре). Получается, что для обеспечения равноправности надо сделать процессы неравноправными!

Для достижения такой справедливости в операционную систему вводится понятие *приоритета процесса*. Этот приоритет выражается целым числом, показывающим, как долго процесс не выполнялся на процессоре.

Планировщик после истечения очередного кванта уменьшает приоритет выполнявшегося процесса, после чего запускает алгоритм перепланирования, который выбирает среди конкурирующих процессов процесс с наибольшим приоритетом.

В примере, после того как пройдёт десять квантов, в течение которых процесс-редактор ждёт ввода символа и выполняется вычислительный процесс, приоритет вычислительного процесса окажется равен 0, а приоритет процесса-редактора — 10. Поэтому на протяжении следующих десяти квантов приоритет процесса-редактора будет выше и планировщик отдаст предпочтение ему. Затем приоритеты процессов станут одинаковыми, и планировщик выберет очередной процесс случайным образом.

Описанный алгоритм называют *алгоритмом планирования по круговой*

ФОРМУЛА ВЫЧИСЛЕНИЯ ПРИОРИТЕТОВ

Иногда необходимо явно указать, что данный процесс приоритетнее остальных. Для этого вводят фиксированный приоритет процесса, на основе которого рассчитывается динамический приоритет. Для операционной системы UNIX BSD 4.4 формула выглядит так:

$$P_{дин} = P_0 + \left[\frac{T}{4}\right] + 2 \times P_{фикс},$$

где $P_{дин}$ — динамический приоритет процесса; P_0 — базовый системный приоритет пользовательских процессов; $P_{фикс}$ — фиксированный приоритет процесса; T — время выполнения процесса, вычисляемое, в свою очередь, по следующим рекуррентным формулам:

$$T = \frac{2 \times T_{ср}}{2 \times T_{ср} + 1} T + P_{фикс},$$

если процесс занимал процессор;

$$T = \left[\frac{2 \times T_{ср}}{2 \times T_{ср} + 1}\right]^{T_{ожид}} \times T,$$

если процесс находился в состоянии ожидания, $T_{ср}$ — усреднённая загрузка процессора; $T_{ожид}$ — время, прошедшее, пока процесс находился в состоянии ожидания.

системе с обратной связью. При таком алгоритме работы планировщика *интерактивные задачи*, т. е. задачи, обрабатывающие ввод информации, имеют преимущество над неинтерактивными задачами, например вычислительными алгоритмами. Этот алгоритм планирования применён в операционных системах Windows и UNIX.

Алгоритмами совершенно иного типа являются алгоритмы *перепланирования систем реального времени*. Операционные системы реального времени применяются для решения задач обработки информации, где скорость самой обработки должна быть соизмерима со скоростью поступления информации. Хорошим примером может служить автоматизированная система управления жизнеобеспечением пациентов в больнице, где подчас требуется быстрая (по человеческим меркам) реакция на изменение состояния людей, или автоматизированная система управления роботами на промышленном производстве, мгновенно изменяющая движение манипулятора.

Системы реального времени, в которых неуспех при обработке события в заданное время приведёт к фатальной ошибке, называют *системами жёсткого реального времени*. Упомянутые системы жизнеобеспечения больных и управления производством относятся именно к такому классу операционных систем.

Иногда жёсткие требования гарантированной доставки события в заданный интервал времени не так важны или, по крайней мере, их нарушение не приводит к фатальным последствиям. Например, задача проигрывания звукового файла на компьютере — это задача *мягкого реального времени*. Алгоритм проигрывания выглядит следующим образом:

```
алг Проигрывание
нач
| нц пока не конец файла
| | прочитать очередную порцию данных
| | записать прочитанную порцию
| |     в регистры звуковой карты
| кц
кон
```

◄ Российский Центр управления полётами.

Здесь тоже важно успеть за некоторый интервал времени записать в регистры звуковой карты очередную порцию данных, иначе звуковой карте нечего будет играть. Для проигрывания звука с приемлемым качеством (частота оцифровки — 44 кГц, стерео, 16-битная звуковая карта, т. е. 4 байт с частотой 44 кГц) нужно успевать считывать с диска и записывать данные со скоростью 200 кбайт/с. Порцию данных в алгоритме можно считать равной 8 кбайт. То есть надо успеть выполнить один полный шаг цикла за 8/200 = = 1/25 с. Если не уложиться в отведённое время, это приведёт к потере звука на 1/25 с. Ситуация неприятная, но не фатальная. Большинство людей просто не заметят потерю звука. Соответственно операционная система мягкого реального времени должна по возможности обеспечивать доставку события прикладному процессу за фиксированный интервал времени.

Механизм работы планировщика может быть устроен следующим образом. Все процессы в системе получают никогда не изменяемый приоритет выполнения. В алгоритме перепланирования всегда выбирается процесс с наибольшим приоритетом. В системах жёсткого реального времени надо выполнять алгоритм перепланирования при наступлении любого события. Тем самым гарантируется, что процесс, ждущий события, будет запущен на выполнение при условии, что его приоритет достаточно высок.

В системах мягкого реального времени, как и раньше, процессорное время разбито на кванты выполнения

РАЗМЕРЫ ОПЕРАЦИОННЫХ СИСТЕМ

По мере того как в операционные системы (ОС) включались всё новые функции, они стали расти в объёме: операционная система разделения времени Compatible Time Sharing System, внедрённая в Массачусетсском технологическом институте в 1963 г., состояла из 32 тыс. слов памяти (по 36 бит каждое). OS/360 фирмы IBM в 1964 г. содержала уже более миллиона команд кода. Монстр Multics, разработанный в том же Массачусетсском технологическом институте в 1975 г. совместно с Bell Laboratories, имел более 20 млн команд.

Операционные системы для персональных ЭВМ, напротив, были компактны: ранние версии MS DOS не превышали 50 тыс. команд, а UNIX 7.5 Bell Laboratories середины 80-х гг. состоял примерно из 100 тыс. машинных команд. Со временем UNIX для персональных ЭВМ остался достаточно компактным, чего нельзя сказать про серию Windows-95, 98, 2000.

Машинный зал компьютера IBM 360.

Кабина Boeing 707.

и алгоритм перепланирования запускается по истечении кванта. Если событие наступит в неудачный момент (например, в начале очередного кванта выполнения), скорее всего, не будет возможности его обработать, но и процессорное время не израсходуется впустую, на постоянное выполнение алгоритма перепланирования. Операционная система — вспомогательная программа, и там, где этого не требуется, занимать ресурс процессора выполнением алгоритмов самой операционной системы не стоит.

ПОДСИСТЕМА УПРАВЛЕНИЯ ПАМЯТЬЮ

На другую важную часть операционной системы — *подсистему управления памятью* — возложены две взаимосвязанные задачи: защита выполняющихся в ЭВМ процессов друг от друга и распределение памяти ЭВМ между процессами.

Процессор общается с памятью в терминах адресов: у каждого байта памяти есть номер, или *физический адрес,* который однозначно определяет данный байт (первый байт памяти имеет адрес 0, второй — 1 и т. д.). Набор из адресов всех байтов памяти называют *физическим адресным пространством*. Для решения задач управления памятью используются аппаратные средства, именуемые *виртуальной памятью*. Её функция состоит в том, чтобы обеспечить дополнительное адресное пространство и механизмы преобразования адресов из одного пространства в другое. Это дополнительное адресное пространство называют *виртуальным адресным пространством*, а адреса в нём — *виртуальными адресами*. В отличие от физического виртуальное

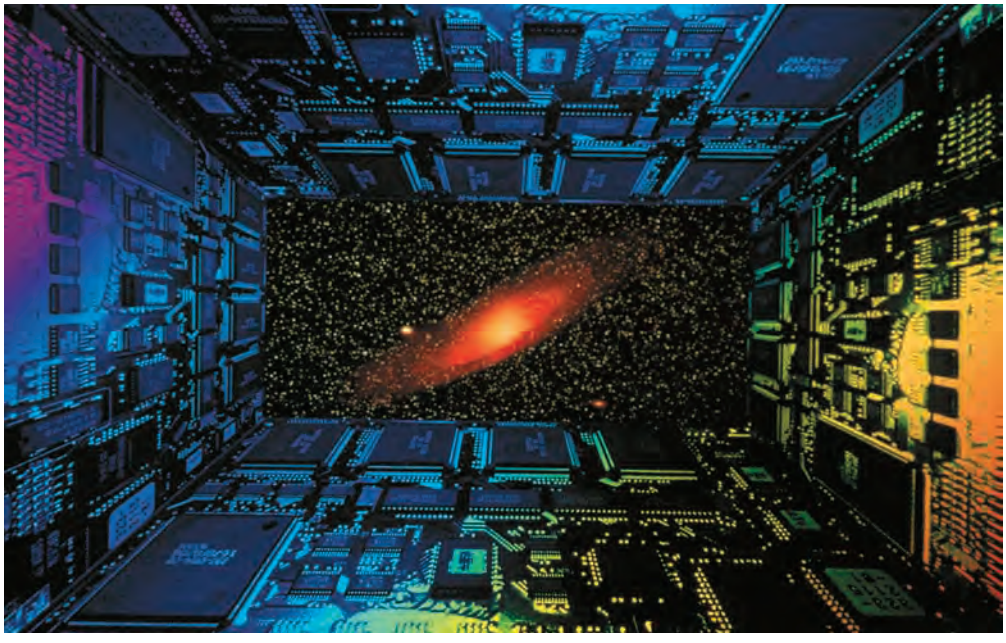
адресное пространство просто набор ничему не соответствующих адресов. Однако внести смысл в его существование можно, присваивая по некоторому правилу адресам виртуального пространства физические адреса. Тогда виртуальный адрес станет номером реального байта памяти.

Первая задача — защитить процессы друг от друга, т. е. защитить память одного процесса от случайных посягательств на неё другого процесса. Каждому процессу ставится в соответствие своё виртуальное пространство — набор адресов от 0 до максимального номера, допустимого процессором, и уникальное для каждого процесса правило преобразования виртуальных адресов в физические. Для обозначения виртуального пространства процесса применяется термин *«адресное пространство процесса»*. Работая в терминах своего адресного пространства, процесс не в состоянии записать что-либо в пространство другого процесса. Действительно, любой адрес, распознаваемый процессором, принадлежит адресному пространству процесса, других адресов нет.

Для того чтобы использовать механизм виртуальной памяти, необходимо включить в ЭВМ специальное устройство — *диспетчер управления памятью,* — осуществляющее преобразование пары номер процесса и виртуальный адрес в физический адрес в памяти. Для современных процессоров диспетчер управления памятью — это отдельный блок в составе процессора, хотя ещё в начале 90-х гг. XX в. диспетчер часто был отдельной микросхемой в составе ЭВМ.

Обычно правила преобразования записываются в виде таблицы, в которой хранятся соответствия между диапазонами физических и виртуальных адресов. Адреса группируют в наборы по нескольку тысяч последовательных адресов в каждом. Например, в наборы по 4096 адресов,

Термины «виртуальная память», «виртуальный адрес» — калька с английских терминов virtual memory и virtual address. Однако значения прилагательного «виртуальный» в русском и английском языках прямо противоположны. Английское virtual означает «фактический», «действительный». В русском же прилагательное «виртуальный» — синоним не существующего, но возможного. Например, виртуальная реальность. Можно сказать, что английский термин virtual memory — это набор адресов и правила их преобразования в адреса физической памяти, тогда как в русском варианте «виртуальная память» — скорее просто абстрактная память, ничему не соответствующая.



что соответствует 4 кбайт памяти или 4 кбайт виртуального адресного пространства. Такой набор называется *страницей.* Если это страница физических адресов, то её именуют *физической страницей* или *страницей памяти*. Соответственно страницу виртуальных адресов называют *виртуальной страницей*. Если пронумеровать страницы, подобно тому как нумеруются байты памяти, получится адрес страницы. В таблице для диспетчера управления памятью, называемой *таблицей страниц,* хранятся адреса физических страниц, а диспетчер вычисляет по адресу виртуальной страницы место в таблице, где находится адрес соответствующей физической страницы.

Использование механизма виртуальной памяти требует небольших изменений в планировщике, а именно, в алгоритм переключения (контекста) процесса нужно добавить переключение таблицы страниц. То есть при замене одного процесса на другой надо указать диспетчеру адрес в памяти, где размещается таблица страниц нового процесса. Точный формат таблицы страниц определяется архитектурой диспетчера (в современных процессорах — архитектурой процессора).

Размещение таблицы страниц в памяти приводит к тому, что диспетчер вынужден обращаться к памяти за информацией при каждом переводе виртуального адреса в физический. Скорость работы процессора намного выше скорости обращения в память, поэтому при трансляции адресов процессор простаивает. Чтобы уменьшить простой процессора, поступают следующим образом. В процессоре организуют память небольшого объёма — *буфер преобразования адреса,* в котором хранят некоторые соответствия между виртуальными и физическими страницами, т. е., по сути, копию отдельных элементов таблицы страниц. При преобразовании виртуального адреса в физический процессор ищет соответствие в буфере преобразования адреса и, только если не находит его там, обращается уже к таблице страниц в памяти. На практике с помощью такого нехитрого приёма удаётся заметно снизить количество обращений к памяти.

Механизмы защиты процессов друг от друга используются и для распределения физической памяти между процессами. Задача формулируется так: разместить в ограниченном объёме физической памяти как можно больше адресных пространств процессов.

Адресное пространство каждого процесса заполнено неодинаково, в нём есть заведомо неиспользуемые места — «дырки», под которые не надо отводить физическую память. Также некоторая часть адресного пространства почти наверняка не будет использоваться часто. Например, код и данные, задействованные в начале работы программы (нужно проинициализировать рабочие переменные) и в конце работы (нужно освободить память, вывести результат и т. п.). Соответственно для участков адресного пространства, используемых лишь при завершении работы, физическую память можно не отводить до тех пор, пока это не потребуется, а память, используемую только в начале работы программы, можно забрать после применения.

Итак, в каждый промежуток времени процесс использует для работы часть объёма своего адресного пространства. Используемые за промежуток времени виртуальные страницы памяти называют *рабочим множеством* процесса. Понятно, что рабочее множество зависит от длительности временного промежутка времени (чем он больше, тем больше рабочее множество процесса).

А если за промежуток выбрать всё время работы программы, рабочее множество будет равно используемому адресному пространству.

Чтобы сэкономить физическую память, можно отводить её только для виртуальных страниц из рабочего множества выполняемых процессов. При определении рабочего множества удобно выбрать интервал времени, равный кванту выполнения. Наилучший алгоритм распределения памяти формулируется так: при выделении процессу очередного кванта выполнения разместим в физической памяти все страницы рабочего множества процесса за этот квант.

Однако, хотя алгоритм действительно является наилучшим, его, к сожалению, нельзя применить. Узнать, какие страницы входят в рабочее множество, можно лишь по истечении кванта, а в алгоритме требуется знать это вначале.

Трудно предсказать будущее, но стоит воспользоваться уже имеющимся опытом. На практике почти всегда оказывается, что рабочее множество не сильно меняется в течение нескольких квантов выполнения. Соответственно в начале работы процесса его рабочее множество можно считать пустым и выделять физическую память под виртуальные страницы по мере необходимости, а также принудительно забирать физическую память для тех страниц, доступа к которым не было на протяжении какого-нибудь интервала времени. Более того, если за интервал взять бесконечность, то принудительно забирать физическую память мы будем, только когда свободной физической памяти для работы не останется. В этот момент забирается память у самой «старой» виртуальной страницы, т. е. у той, к которой дольше всего не было доступа.

Механизм защиты страниц позволяет определить, когда виртуальной странице понадобится отвести страницу физической памяти. В момент старта процесса можно подготовить незаполненную таблицу соответствия между виртуальными и физическими страницами. Дополнительно придётся ещё хранить информацию: нужно отличать виртуальные страницы, для которых соответствия в таблице трансляции ещё не построено, от «дырок» в адресном пространстве, для которых такого соответствия не должно существовать в принципе.

Если процессор при обращении по виртуальному адресу не найдёт соответствия для виртуальной страницы, возникнет *исключительная ситуация — прерывание*. При возникновении прерывания процессор начинает выполнять специальную подпрограмму — *обработчик прерывания*.

ИСПОЛЬЗОВАНИЕ ВИРТУАЛЬНОЙ ПАМЯТИ ДЛЯ ЗАЩИТЫ ВНУТРИ ПРОЦЕССА

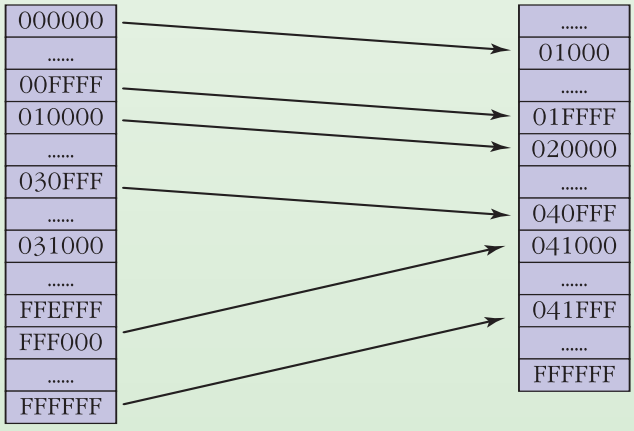
Механизм виртуальной памяти используют также для защиты процесса от самого себя. Например, защищают сегмент кода программы от случайных записей. Для защиты процессов друг от друга достаточно просто занять все возможные адреса, устранив, таким образом, возможность обращения к чужой памяти. Для защиты внутри процесса эта техника не работает, так как постоянно требуется доступ к коду и данным программы. Однако каждый сегмент специфичен: сегмент кода достаточно лишь читать и выполнять, к сегменту данных нужен доступ только на чтение и запись (но не на выполнение!).

Для разграничения доступа к сегментам в специальном байте управляющей информации сегмента хранят биты — признаки доступа: чтение, запись и выполнение. Естественно, процессор должен распознавать эти биты и проверять доступ. Так, для страниц сегмента кода можно установить биты доступа «чтение» и «выполнение», а для страниц сегмента данных и стека — биты доступа «чтение» и «запись». При этом программа не сможет случайно записать данные в код и случайно выполнить данные вместо кода, посчитав, что тут содержится программа.

ПРИМЕР РАБОТЫ МЕХАНИЗМА ВИРТУАЛЬНОЙ ПАМЯТИ

Пусть для простоты примера в компьютере имеется 16 Мбайт памяти. Тогда последний байт памяти имеет физический адрес 16777215 (16x1024x1024 — 1) в десятичной системе исчисления или 0xFFFFFF в шестнадцатеричной системе исчисления (буква «x» обозначает, что число записано в шестнадцатеричной системе), в которой принято записывать адреса. В этом случае физическое адресное пространство выглядит так: 0x000000 — 0xFFFFFF (т. е. набор адресов от 0 до 0xFFFFFF).

Виртуальное адресное пространство пронумеровано также от 0x000000 по 0xFFFFFF (объём — 16 Мбайт). Тогда виртуальным адресам ставятся в соответствие физические по следующему правилу:



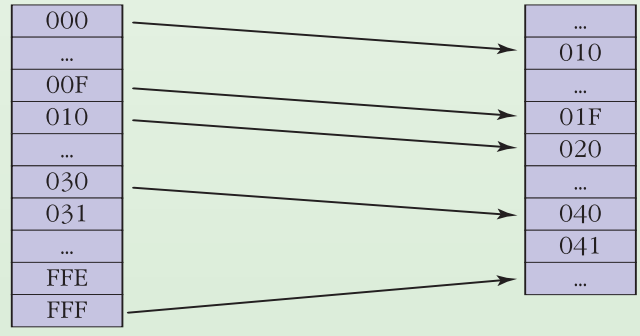
- виртуальные адреса с 0x000000 по 0x00FFFF будут соответствовать физическим адресам 0x010000 — 0x01FFFF со сдвигом (т. е. виртуальный адрес 0 соответствует физическому 0x10000, виртуальный адрес 1 соответствует физическому 0x10001 и т. д., наконец, виртуальный адрес 0xFFFF соответствует физическому 0x1FFFF);
- виртуальные адреса 0x010000 — 0x030FFF будут соответствовать физическим 0x020000 — 0x040FFF;
- виртуальные адреса 0xFFF000 — 0xFFFFFF будут соответствовать физическим 0x041000 — 0x041FFF;
- виртуальные адреса 0x031000 — 0xFFEFFF ничему соответствовать не будут.

Теперь процессор может, например, прочитать байт по виртуальному адресу 0x025367. Этому виртуальному адресу по правилам соответствует физический адрес 0x35367 и соответственно 217961-й байт памяти ($35367_{16} = 217961_{10}$).

А сейчас надо записать правило для диспетчера управления памятью, который, собственно говоря, и должен вычислять физический адрес по виртуальному. Причём записать как можно более компактно, поскольку диспетчер это правило должен где-то хранить. Если на каждый виртуальный адрес хранить в таблице его физический, то потребуется 48 Мбайт памяти = 16 Мбайт памяти × × 3 байт (размер адреса); это в три (!) раза превышает объём памяти.

Поэтому используют разбиение памяти на страницы размером 4 кбайт, пронумеровав их, начиная с 0. Зная адрес, легко вычислить номер страницы: нужно отбросить последние три цифры в шестнадцатеричной системе исчисления. Например, адрес 0x35367 находится в странице 0x3516 или 5310 в десятичной системе исчисления.

Теперь правила соответствия памяти в терминах страниц выглядят следующим образом:



- виртуальные страницы 0x000 — 0x00F соответствуют физическим 0x010 — 0x01F;
- виртуальные страницы 0x010 — 0x030 соответствуют физическим 0x020 — 0x040;
- виртуальная страница 0xFFF соответствует физической 0x041;
- виртуальные страницы 0x031 — 0xFFE никаким физическим страницам не соответствуют.

Тогда для хранения соответствия страниц достаточно одномерной таблицы (массива) в памяти.

Пусть таблица размещается в начале физической памяти ЭВМ. Для вычисления адреса физической страницы для данной виртуальной нужно умножить номер виртуальной страницы на два, так как номер страницы кодируется 2 байтами. Например, процессор должен прочитать байт по виртуальному адресу 0x113A5, номер виртуальной страницы получается отбрасыванием трёх младших знаков, т. е. 0x11. Умножив 0x11 на два и получив 0x22, можно прочитать значение по этому адресу из таблицы, получится номер физической страницы — 0x21. Далее, снова добавив три младшие цифры виртуального адреса, получим физический, по которому и надо прочитать байт — 0x213A5.

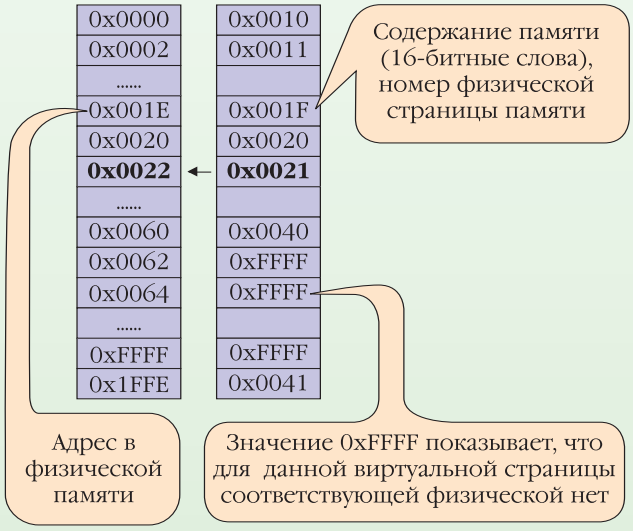
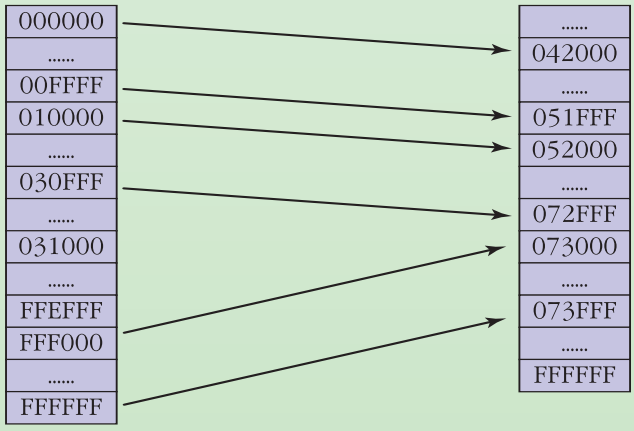


Таблица в памяти занимает всего 8 кбайт, или всего 0,5 % 16 Мбайт памяти.

Теперь представим два процесса *A* и *B*. У каждого из них собственное виртуальное адресное пространство 0x000000 — 0xFFFFFF. Остаётся только назначить физическую память для адресных пространств обоих процессов. Пусть адресное пространство процесса *A* использует уже знакомую таблицу. Распределяя память для процесса *B*, надо обратить внимание, что есть две области свободной физической памяти: память с адресами 0x000000 — 0x00FFFF и 0x042000 — 0xFFFFFF соответственно. Первая область памяти использована для размещения таблицы страниц процесса *A*. Во второй разместится процесс *B*:



- виртуальные адреса с 0x000000 по 0x00FFFF будут соответствовать физическим адресам 0x042000 — 0x051FFF;
- виртуальные адреса 0x010000 — 0x030FFF будут соответствовать физическим 0x52000 — 0x072FFF;
- виртуальные адреса 0xFFF000 — 0xFFFFFF будут соответствовать физическим 0x073000 — 0x073FFF;
- виртуальные адреса 0x031000 — 0xFFEFFF ничему соответствовать не будут.

Для размещения таблицы страниц процесса B опять требуется 8 кбайт памяти.

Использование трёх наборов адресов (их ещё называют сегментами) виртуальной памяти процесса не случайно. Дело в том, что, как правило, память каждого процесса-программы распределена следующим образом:

- память для команд, выполняемых процессором, — *сегмент кода;*
- память переменных — *сегмент данных;*
- память для стека, в котором хранятся временные переменные подпрограмм, адреса возврата из подпрограмм, — *сегмент стека*.

Соответственно виртуальное пространство процесса в примере состоит из трёх сегментов:

0x000000 — 0x00FFFF — сегмент кодов;

0x010000 — 0x030FFF — сегмент данных;

0x031000 — 0xFFEFFF — «дырка» в пространстве, не используется;

0xFFF000 — 0xFFFFFF — сегмент стека процесса.

Сегмент данных может меняться в процессе выполнения программы, так как требуется память под новые данные. Сегмент данных удобно размещать сразу после сегмента кодов (который не изменяется), а сегмент стека — в самом конце пространства.

***

На практике таблица страниц выглядит сложнее. В настоящее время большинство процессоров 32-разрядные, это означает, что адрес состоит из 32 разрядов, или 4 байт (в байте 8 разрядов). Соответственно виртуальное адресное пространство, с которым может работать 32-разрядный процессор, имеет размер 2^{32} байт, или 4 Гбайт. Однако объём физической памяти должен быть ещё больше. Например, у 32-разрядного процессора Intel Pentium III размер физического адреса — 36 разрядов. У 64-битных процессоров аппетиты ещё больше, размер виртуального адресного пространства — 2^{64} байт, или 16 Пбайт. Поскольку человечество ещё не решило, что делать с таким количеством физической памяти, реальная разрядность физического адреса меньше. У 64-разрядного процессора Alpha физический адрес занимает 48 разрядов.

В данном случае это будет прерывание по нарушению защиты страниц. Обработчику предстоит вначале выяснить, не произошла ли ошибка, должно ли быть установлено соответствие между виртуальной страницей и какой-либо физической. Если должно, обработчик выделяет физическую страницу и изменяет таблицу, записывая новое отображение.

Чтобы определить самую старую страницу, лучше всего если процессор при обращении к виртуальной странице будет записывать куда-нибудь текущее время, измеряемое, например, в числе инструкций, выполненных с момента включения ЭВМ. Операционная система через некоторый интервал времени просматривает таблицы страниц всех процессов и освобождает физическую страницу. На практике хранить полное время доступа к странице достаточно дорого, и операционная система использует менее эффективные алгоритмы.

В результате работы алгоритма распределения памяти освобождаются физические страницы. Данные, которые в них содержатся, могут ещё потребоваться, и информацию надо сохранить. Как правило, у ЭВМ есть диск, на котором можно выделить место для хранения содержимого освобождённых физических страниц. Это место на диске называют *областью подкачки*. Хранение содержимого освобождённых страниц именуют *подкачкой страниц*. Иногда в литературе для обозначения этого процесса можно встретить пришедший из английского языка термин «свопинг» (*англ.* swapping — «перекачка»).

При использовании подкачки страниц алгоритм распределения памяти немного усложняется. Перед тем как выделить физическую страницу в обработчике прерывания, надо выяснить, не было ли содержимое страницы ранее сохранено в области подкачки, и, если было, «подкачать» содержимое обратно в память, т. е. прочитать обратно в память данные с диска.

Размер области подкачки зависит от загруженности ЭВМ и объёма оперативной памяти. Раньше широко использовалась формула: размер области подкачки в два раза больше объёма физической памяти. Сейчас физическая память даже у персональных ЭВМ достигает нескольких гигабайт, поэтому размер области подкачки делают равным объёму памяти или даже ещё меньше.

А как быть, если в области подкачки нет свободного места? Есть два способа решить эту проблему. Один, консервативный, состоит в том, чтобы при выделении очередной порции памяти зарезервировать место в области подкачки, а если его нет, то память не выделять. Это означает, что, если в области подкачки свободного места нет, в операционной системе нельзя будет запустить новый процесс, а уже работающие процессы не получат дополнительной памяти. Однако на практике это при-

водит к тому, что область подкачки используется неэффективно: свободное место в области подкачки есть, но использовать его нельзя, оно зарезервировано.

Другой способ состоит в том, что операционная система берёт на себя «повышенные обязательства» и выделяет виртуальную память, пока есть свободное место в физической памяти и в области подкачки (по-английски эта стратегия называется overcommitment). Если в области подкачки свободного места нет, придётся совершать неприятные действия: принудительно освободить место, что означает просто аварийно уничтожить какой-либо процесс.

Можно разработать и другие алгоритмы, которые пытаются предсказать будущее на основе прошлого опыта. Например, вместо того чтобы забирать память у самой «старой» страницы, забирать память у наименее используемой страницы. (Это не то же самое!)

В области алгоритмов распределения памяти постоянно ведутся исследования, здесь всегда есть место для улучшений.

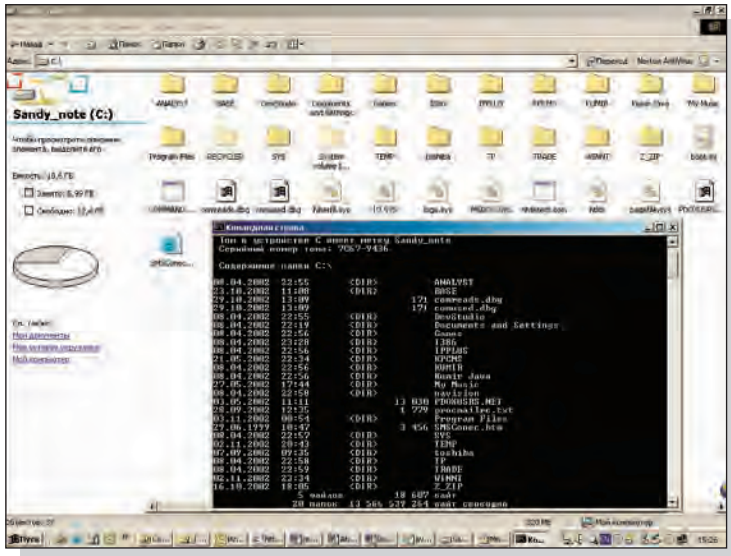
ФАЙЛОВАЯ СИСТЕМА

Ещё одна важнейшая часть операционной системы — *файловая система.* Этим термином обозначаются два совершенно разных понятия. Во-первых, файловой системой называют набор алгоритмов операционной системы для работы с файлами. Во-вторых, файловая система — это и структура хранения файлов на запоминающих устройствах, например на дисках.

В самом общем (и самом непонятном) виде можно определить *файл* как упорядоченный набор *записей,* а запись — как последовательность байтов.

Удобно рассматривать файловую систему как структуру хранения файлов на запоминающих устройствах. Например, диск рассматривается как массив *блоков,* каждый блок — это массив байтов фиксированной длины, обычно 512 байт. То есть нельзя

Понятие записи необходимо, поскольку с файлами работают именно в терминах записей. Например, с текстовым файлом работают как с набором строк символов; запись в текстовом файле — это строка. Видеофайл состоит из кадров, поэтому запись в видеофайле — это кадр. Иногда с файлом удобно работать как с последовательностью байтов (часто в этом случае употребляют термин «поток байтов»), тогда каждый байт можно считать отдельной записью.



Файловая система Windows.

Файловая система UNIX.

адресоваться к каждому байту диска напрямую: доступны только адреса блоков. Адрес блока можно считать 32-битным числом, адреса такой длины хватит, чтобы работать с дисками объёмом до $2^{32} \cdot 512 = 2$ Тбайт.

```
[root@linus root]# ls -la
total 328
drwxr-x---   14 root     root         4096 Oct 28 12:34 .
drwxr-xr-x   19 root     root         4096 Oct  8 12:42 ..
-rw-r--r--    1 root     root         2516 Sep 24 20:52 anaconda-ks.cfg
-rw-------    1 root     root        21061 Oct 24 15:52 .bash_history
-rw-r--r--    1 root     root           24 Jun 11  2000 .bash_logout
-rw-r--r--    1 root     root          234 Jul  5  2001 .bash_profile
-rw-r--r--    1 root     root          176 Aug 23  1995 .bashrc
-rw-r--r--    1 root     root          210 Jun 11  2000 .cshrc
drwx------    3 root     root         4096 Oct 24 15:52 Desktop
drwx------    3 root     root         4096 Sep 27 16:37 .gnome
drwx------    2 root     root         4096 Sep 27 16:15 .gnome_private
-rw-r--r--    1 root     root         1112 Oct 24 15:52 .gtkrc-kde
-rw-------    1 root     root          199 Oct 24 15:52 .ICEauthority
-rw-r--r--    1 root     root         8236 Sep 24 20:47 install.log
-rw-r--r--    1 root     root            0 Sep 24 20:04 install.log.syslog
drwxr-xr-x    3 root     root         4096 Oct 17 18:09 .java
drwx------    4 root     root         4096 Sep 25 17:42 .kde
-rw-------    1 root     root          157 Sep 25 17:42 .kderc
drwxr-xr-x    3 root     root         4096 Sep 25 17:42 .mcop
-rw-------    1 root     root           31 Oct 24 15:52 .mcoprc
drwxr-xr-x    3 root     root         4096 Oct  4 22:14 .mozilla
drwxr-xr-x    5 root     root         4096 Oct 11 15:21 .netscape
drwx------    2 root     root         4096 Sep 27 18:45 nsmail
-rw-r--r--    1 root     root       112295 Sep 27 19:15 print.pdf
drwxr-xr-x    2 root     root         4096 Sep 24 21:34 .qt
drwx------    2 root     root         4096 Sep 25 17:43 .ssh
-rw-r--r--    1 root     root          196 Jul 11  2000 .tcshrc
-rw-r--r--    1 root     root           11 Sep 27 17:56 test.koi
drwxr-xr-x    3 root     root         4096 Oct 15 19:42 tmp
-rw-------    1 root     root         6988 Oct 28 12:25 .viminfo
-rw-r--r--    1 root     root            8 Oct 24 15:52 .wmrc
-rw-------    1 root     root           56 Oct 28 12:32 .xauthbZTBBK
-rw-------    1 root     root         1164 Oct 24 15:52 .Xauthority
-rw-------    1 root     root           57 Oct 28 12:04 .xauthypc8fA
-rw-r--r--    1 root     root        15507 Sep 24 21:34 .xftcache
-rw-r--r--    1 root     root         1126 Aug 23  1995 .Xresources
-rw-------    1 root     root         1971 Oct 25 19:50 .xsession-errors
[root@linus root]#
[root@linus root]#
[root@linus root]#
[root@linus root]#
[root@linus root]#
[root@linus root]#
```

В современных файловых системах вводят понятие *логического блока,* размер которого делают кратным размеру *физического блока.* Например, сейчас наиболее распространённый размер логического блока — 1 кбайт, встречаются также размеры 4 и 8 кбайт. Необходимость введения логического блока объясняется спецификой диска: прочитать с диска несколько подряд идущих блоков быстрее, чем прочитать те же блоки подряд по одному.

Для каждого файла нужно хранить его *атрибуты:*

- имя файла (в файловой системе UNIX это просто число);
- тип файла: обычный файл или каталог;
- кто владелец файла и к какой группе пользователей он относится (два числа — идентификатор или номер владельца и идентификатор группы);
- права доступа к файлу — чтение, запись, выполнение — для трёх категорий пользователей: владельца файла, пользователя из той же группы, что и владелец, и остальных (9 бит: по 3 бита с информацией о защите для трёх категорий пользователей);
- информацию о работе с файлом — время создания файла, его последнего изменения и время последнего доступа к нему (три больших числа, интерпретируемые как количество секунд, прошедших с 1 января 1970 г., или, как иногда говорят, с начала Эпохи, — эта дата считается днём возникновения UNIX; в 32-битном числе можно закодировать даты с 1970 по 2038 г.);
- число ссылок на файл, или, по-другому, количество символьных имён у файла;
- номера блоков, в которых размещается файл, — список блоков файла;
- размер файла в байтах, поскольку списка блоков недостаточно; по размеру файл может быть не кратен блоку.

Все атрибуты, за исключением списка блоков, занимают небольшой объём — 32 байт. Поэтому можно отвести на диске область фиксированного

объёма, где следует хранить информацию обо всех файлах сразу. Такая область называется *массивом индексов*. Соответственно *индекс файла* — набор его атрибутов. При этом имя файла есть индекс в этом массиве. Число файлов на диске будет ограничено, и его надо знать заранее при создании на диске файловой системы. Можно просто выбрать большое число и надеяться, что его хватит.

Поиск файла по имени — тривиальная операция, поскольку имя файла — это номер индекса файла в массиве индексов. Однако пользователь работает с символьными именами файлов, и надо хранить соответствие между символьным именем и номером индекса. Самое простое — разместить эту информацию в каталоге. Собственно в каталоге только эта информация и будет храниться. То есть каталог — это файл, состоящий из записей в виде символьного имени и номера индекса. Осталось выделить специальный каталог — *главу иерархии,* поскольку файловая система имеет иерархическую структуру. Главой иерархии называют *корень файловой системы*. Для него можно выделить специальный индекс, например первый, и специальное имя. Традиционно имя корня файловой системы ОС UNIX — «/». Теперь любой файл в файловой системе доступен по символьному имени.

Например, для того чтобы найти файл с именем «/Мои документы/Разное/МойФайл.txt», надо

1) прочитать первый индекс, найдя блоки корня файловой системы;

2) читать блоки корня, пока не найдена запись, в которой находится символьное имя «Мои документы». Пусть этому имени соответствует индекс номер 17;

3) прочитать индекс с номером 17, найдя блоки этого каталога;

4) читать блоки файла 17, пока не найдена запись, в которой находится символьное имя «Разное». Пусть этому имени соответствует индекс номер 103;

5) прочитать индекс с номером 103, найдя блоки этого каталога;

6) читать блоки файла 103, пока не найдена запись, в которой находится символьное имя «МойФайл.txt».

ИМЕНА

В некоторых файловых системах, например в FAT MS DOS или Windows, имя файла может быть обычным символьным именем, например «МойФайл». В файловых системах ОС UNIX имя файла — просто число, т. е. все файлы пронумерованы. При этом существуют и символьные имена, которые являются ссылкой на это имя файла, и таких ссылок может быть сколько угодно. Иными словами, символьные имена «МойФайл» и «ТвойФайл» могут ссылаться на один и тот же файл. В файловых системах RSX и VMS файл полностью определялся символьным именем, расширением и версией. Имеются в виду файлы вида «МойФайл.txt;1», «МойФайл.txt;2». Здесь «МойФайл» — это имя, txt — расширение, 1 или 2 — версия. Сейчас подобные имена с расширением и версией можно встретить в файловой системе ISO9660, используемой для хранения файлов на CD ROM.

Современные файловые системы имеют *иерархическую структуру*. Файлы хранятся в *каталогах* или папках. Поэтому для доступа к файлу информации о его имени недостаточно: нужно знать *маршрутное имя*, включающее список всех каталогов, которые необходимо пройти, чтобы добраться до файла. Например, в FAT MS DOS маршрутное имя файла выглядит так:

D:\Моидокументы\Разное\МойФайл.txt

Здесь D — диск, на нём размещается файл с именем «МойФайл.txt»; Мои документы — имя папки, где находится папка с именем Разное (в ней содержится нужный файл); символ \ — разделитель.

В файловой системе ОС UNIX это же маршрутное имя будет записываться так:

/mnt/disk/Моидокументы/Разное/МойФайл.txt

Здесь /mnt/disk — каталог, с которого начинается файловая система диска; символ / — разделитель элементов маршрутного имени.

Наконец, в файловой системе VMS маршрутное имя записывается в виде

DISK$D:[МОИДОКУМЕНТЫ.РАЗНОЕ]МОЙФАЙЛ.TXT;1

Здесь «DISK$D» — имя диска, на котором размещается файл; символы «[» и «]» ограничивают поле каталогов в пути к файлу; символ «.» является разделителем каталогов.

ФАЙЛЫ В ОС UNIX

Подход заключается в том, чтобы посмотреть на типичную длину файла в файловой системе и создать требуемое число индексов в зависимости от объёма диска. Например, в современных ОС UNIX типичная длина файла 10 кбайт. Значит, нужно на каждые десять логических блоков файловой системы отводить один индекс, если логический блок имеет размер 1 кбайт.

Со списком блоков дело обстоит сложнее. Надо уметь адресовать каждый байт файла. Адресация в файле традиционно устроена как смещение относительно текущей позиции указателя чтения/записи; смещаться можно как вперёд, так и назад, поэтому адрес будет 32-разрядным числом со знаком. И максимальный объём файла 2 Гбайт, ровно столько можно адресовать 32-разрядным числом со знаком.

В списке блоков может присутствовать до 2 млн элементов (число элементов в списке блоков = объём файла/размер блока = 2 Гбайт/1 кбайт = $2 \cdot 2^{20}$), а для хранения списка понадобится 8 Мбайт ($2 \cdot 2^{20} \cdot 4$ байт в 32-разрядном адресе блока = 8 Мбайт).

Номера нескольких первых блоков хранятся прямо в индексе файла. Поскольку считается, что типичный файл занимает десять блоков, на него отводится в индексе место для десяти блоков. Далее вводится понятие (аналогичное директории страниц) «*блок косвенности*». В нём хранятся не записи файла, а номера блоков в списке. Блок косвенности будет 11-м блоком в индексе файла. Так как размер блока принят равным 1 кбайт, размер адреса — 32 бит; в блоке косвенности можно хранить номера 256 блоков. Всего 266 блоков, или первые 266 кбайт файла.

Вообще в одном блоке косвенности может храниться информация о 256 кбайт файла. Для хранения файла размером 2 Гбайт нам нужно 8 тыс. блоков косвенности ($2 \cdot 2^{20}$ блоков в списке/256 блоков в блоке косвенности = $8 \cdot 2^{10}$ блоков косвенности), что потребует 32 кбайт в индексе. Это меньше, чем 8 Мбайт в случае хранения просто номеров блоков, но всё ещё много, чтобы хранить все блоки косвенности в каждом индексе.

Второй блок косвенности содержит номера первых блоков косвенности (которые уже содержат номера блоков файла) и является 12-м блоком в индексе файла. Во втором блоке косвенности можно хранить номера 256 первых блоков косвенности. Итого, $10 + 256 + 256^2 = 65\,802$ блока, или чуть больше 64 Мбайт файла. Для хранения 2 Гбайт нам понадобится 32 вторых блока косвенности ($2 \cdot 2^{20}$ блоков в списке/256^2 блоков во втором блоке косвенности = = 32 вторых блока косвенности), что потребует 128 байт в индексе. Третий блок косвенности содержит номера вторых блоков косвенности. Он будет 13-м блоком в индексе. В третьем блоке косвенности могут храниться до 256 вторых блоков косвенности. Всего, $10 + 256 + 256^2 + 256^3 = 16\,843\,018$ блоков, т. е. максимальный размер файла может составлять 16 Гбайт. Третьего блока заведомо хватит для хранения файла объёмом 2 Гбайт. Структура списка блоков в индексе получилась следующей.

Номера первых десяти блоков файла.

Первый блок косвенности. → Номера следующих 256 блоков файла.

Второй блок косвенности. → Номера 256 первых блоков косвенности. → Номера следующих 256^2 блоков файла.

Третий блок косвенности. → Номера 256 вторых блоков косвенности. → Номера 256^2 первых блоков косвенности. → Номера следующих 256^3 блоков файла.

Для хранения номеров блоков понадобится зарезервировать 52 байт в индексе (13 номеров блоков × 32 бит = 52 байт). Общий размер — 84 байта.

Однако при работе с большими файлами за компактный индекс придётся заплатить производительностью. Действительно, чтобы считать 11-й килобайт файла, надо прочитать два блока — первый блок косвенности, чтобы узнать номер блока, и сам блок. Для чтения каждого блока из первого мегабайта потребуется прочитать три блока — второй блок косвенности, первый блок косвенности и сам блок. Наконец, для чтения блока из сотого мегабайта потребуется прочитать четыре блока — третий блок косвенности, второй блок косвенности, первый блок косвенности и сам блок.

Объём современных дисков составляет десятки гигабайт; если выделять один индекс на каждые 10 кбайт, нам понадобится 1 млн индексов, соответственно размер массива индексов составит 84 Мбайт, или менее 1 % объёма диска, что немного.

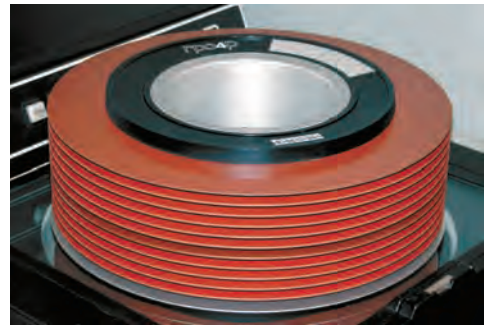
Пусть этому имени соответствует индекс номер 356;

7) прочитать индекс с номером 356, найдя блоки этого файла;

8) прочитать блоки требуемого файла.

Если внимательно посмотреть на этот алгоритм, можно заметить, что для любого файла в файловой системе придётся читать с диска одни и те же данные. Действительно, шаг первого алгоритма придётся выполнять для каждого файла. Файлы пользователя, как правило, находятся только в принадлежащем ему каталоге, например «Мои документы», поэтому часто придётся выполнять шаг 2. Чтение с диска достаточно медленная операция, поэтому закономерно возникает потребность ускорить поиск и чтение часто используемых файлов.

Можно воспользоваться идеей кэширования данных и применить её к файлам. В памяти отводят место под часто используемые блоки файлов — *буферный кэш*. Алгоритм работы с буферным кэшем (алгоритм вытеснения блоков) очень похож на алгоритм распределения памяти. Если алгоритм, который уничтожает самую «старую» страницу, применить к буферному кэшу, то будет выбрасываться самый старый блок. Более эффективное решение — выбрасывать из кэша блок, редко используемый. Буферизовать блоки можно также и при их записи на диск. Операция записи медленнее, чем операция чтения. Это связано с особенностями оборудования. Во-первых, современные диски сами кэшируют блоки, и чтение может происходить из кэша диска. Во-вторых, при чтении с диска данные записываются в память, а при записи на диск — читаются из памяти, как ни странно, операция чтения памяти более медленная, чем операция записи. Можно кэшировать и записи на диск. Однако здесь надо действовать осторожно: при кэшировании блоков, записываемых на диск, велик риск их потерять в случае сбоя ЭВМ. Они могут остаться в кэше и не записаться на диск. Проблема заключается в том, что пользователь может резонно полагать, что его данные уже сохранены на диске, и очень удивится, когда их там не обнаружит. Поэтому некоторые ОС вообще не кэшируют операции записи, предпочитая работать пусть медленно, но зато надёжно. ОС UNIX традиционно кэширует блоки при записи, однако через равные промежутки времени (как правило, через 30 с) принудительно записывает все блоки на диск. Более чем 30-летняя практика использования данного алгоритма показала, что это приемлемо для большинства пользователей.



Магнитный диск.

А ЕСЛИ ИХ МНОГО?

В современных ОС существует потребность работать с несколькими файловыми системами одновременно. Например, на диске размещают файловую систему UNIX или NTFS в случае ОС Windows, поскольку эти системы специально созданы для эффективной работы с файлами на дисках, или, как говорят, оптимизированы по быстродействию. Для работы с CD ROM используют файловую систему ISO9660, позволяющую наиболее рационально хранить файлы, т. е. она оптимизирована по пространству хранения. Дискеты, как правило, содержат файловую систему FAT, которая использовалась в ОС MS DOS, поскольку все попросту привыкли, что дискета содержит FAT. Кроме того, существует файловая система NFS, разрешающая работать с дисками других ЭВМ, подключёнными к локальной сети или доступными через Интернет.

Файловые системы имеют различные соглашения о том, как именовать файлы, или, как говорят, различные пространства имён. У пользователя, естественно, есть желание работать с единым пространством имён, т. е. именовать файлы так, как он привык, не думая о типе используемой файловой системы. Конечно, операционная система берёт на себя работу по преобразованию имён и сведению их к единому виду. За выполнение преобразования имён отвечает специальная часть ОС — виртуальная файловая система. По существу, виртуальная файловая система — это прослойка между ОС и алгоритмами, работающими со структурой конкретной файловой системы. То есть виртуальная файловая система получает запросы от ОС для работы с файлами (например, прочитать данные из файла или удалить файл), определяет, в какой файловой системе находится данный файл, и транслирует запрос в операции этой файловой системы.

ПОДСИСТЕМА УПРАВЛЕНИЯ ВВОДОМ-ВЫВОДОМ И ДРАЙВЕРЫ УСТРОЙСТВ

Основная задача *подсистемы управления вводом-выводом* как части операционной системы — обеспечение доступа к периферийным устройствам, таким, как диски, клавиатура, видеокарта, сетевая плата. Название «периферийные устройства» возникло из представления об ЭВМ как наборе основных устройств: процессора, оперативной памяти и вспомогательных устройств, выполняющих ввод или вывод информации. Прилагательное «периферийное» в названии этих устройств происходит от английского peripheral — «второстепенный». Другие общеупотребительные названия периферийных устройств — «внешние устройства» и «устройства ввода-вывода».

Устройств ввода-вывода великое множество, и операционной системе помимо эффективной работы с ними требуются единые правила доступа к ним, без специальных знаний об особенностях каждого конкретного устройства. Действительно, с точки зрения пользователя, все типы клавиатур ничем не отличаются друг от друга. Вернее, он скорее скажет, что они имеют различный дизайн корпуса и различное число кнопок, а не различные типы подключения к ЭВМ и различные правила общения с ними. Например, клавиатура может подключаться к последовательному порту ЭВМ, или к порту PS/2, или через шину USB. Каждое из перечисленных типов подключений имеет специфические для них правила взаимодействия, или, как говорят, *интерфейс с устройством*.

Для того чтобы сделать взаимодействие с устройствами одинаковым, в ОС определяют единые правила взаимодействия, а для каждого типа устройства создают набор алгоритмов, переводящих единые правила в специфические правила общения с этим типом устройств, по сути управляющих данным устройством. Такой набор алгоритмов называется *драйвером устройства,* а правила работы с этими алгоритмами — *интерфейсом с драйвером*.

Понятно, что единый и одновременно эффективный интерфейс со всеми драйверами для абсолютно всех типов устройств изобрести невозможно. Поэтому существует несколько типов правил общения между драйверами и остальными частями ОС (несколько типов интерфейсов с драйвером) и соответственно несколько типов драйверов (разные интерфейсы — разные алгоритмы). Обычно выделяют следующие типы:

- драйвер символьного устройства, например драйвер клавиатуры. Основные операции для работы с таким устройством — чтение и запись последовательности байтов;
- драйвер блочного устройства, например драйвер диска. Основные

Обычные часы с флэш-памятью подключаются к компьютеру через USB.

операции для работы с таким устройством — чтение и запись блоков, т. е. массивов байтов фиксированной длины;

• драйвер сетевого устройства, например адаптера для входа в Интернет. Основные операции для работы с таким устройством — приём и отправка пакетов, т. е. массивов байтов определённой структуры.

Как выглядит интерфейс с драйвером, например с драйвером символьного устройства? Прежде всего операционной системе необходимо определить, присутствует ли устройство в ЭВМ. Соответственно у драйвера должна быть подпрограмма, которая проверяет конфигурацию ЭВМ и выдаёт ответ «да» или «нет»:

```
мой_драйвер_опросить() возвращает да/нет
```

Если устройство найдено, его необходимо *проинициализировать*, т. е. запрограммировать так, чтобы оно находилось в состоянии «выключено». В момент вызова функции инициализации операционная система ещё не готова работать с устройством; например, для клавиатуры это означает, что по нажатии клавиши она не должна посылать код клавиши в ЭВМ.

```
мой_драйвер_инициализировать()
```

Следующая подпрограмма, которая будет вызвана операционной системой, — подпрограмма начала работы с устройством. Вызывая эту подпрограмму, операционная система сообщает, что она готова работать с устройством. Например, готова принимать символы от клавиатуры. Соответственно подпрограмма начала работы должна включить устройство.

```
мой_драйвер_открыть(устройство)
```

Теперь могут вызываться подпрограммы чтения, записи и управления.

Подпрограмма чтения возвращает байты, выданные устройством к моменту вызова подпрограммы. Если устройство ещё ничего не выдало, то будет возвращён 0.

CDROM, подключаемый к компьютеру через универсальный разъём USB.

```
мой_драйвер_читать(буфер) возвращает количество прочитанных байтов
```

Подпрограмма записи записывает указанное количество байтов в устройство, в случае клавиатуры этой подпрограммы не будет.

```
мой_драйвер_записать(байты) возвращает успех/неуспех
```

Подпрограмма управления устройством выполняет все остальные операции. Например, у клавиатуры есть световые индикаторы, которые можно зажигать или выключать.

```
мой_драйвер_управление(команда, параметры_команды)
```

Наконец, в момент завершения работы с устройством будет вызвана подпрограмма, которая его выключит.

```
мой_драйвер_закрыть(устройство)
```

Ещё у драйвера есть подпрограмма обработки прерывания; она обычно и выполняет основную работу при возникновении прерывания от устройства. Если клавиатура по нажатии клавиши возбудит прерывание, подпрограмма обработки драйвера клавиатуры перенесёт код клавиши во внутренний буфер символов драйвера, содержимое которого будет скопировано подпрограммой чтения.

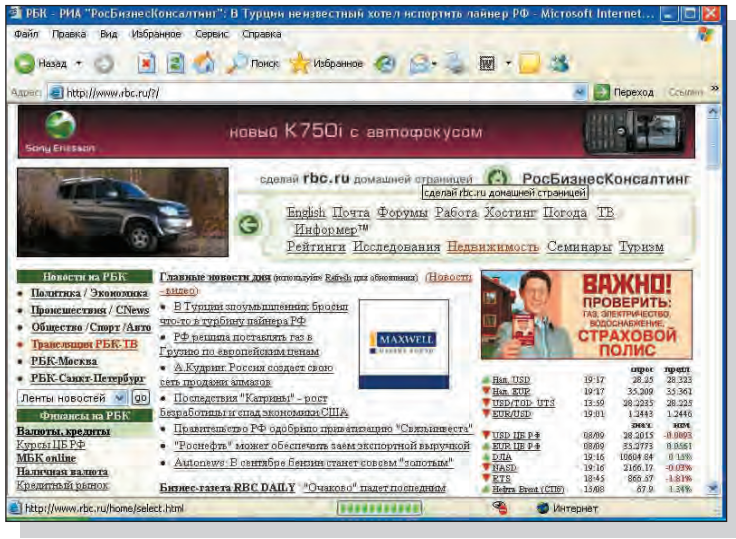
```
мой_драйвер_прерывание()
```

ПРИКЛАДНОЙ ПРОГРАММНЫЙ ИНТЕРФЕЙС

С помощью *прикладного программного интерфейса* определяют правила работы программиста с операционной системой. В эти правила входят перечень доступных программисту подпрограмм — *системных вызовов,* описание правил вызова этих подпрограмм, т. е. имя подпрограммы, количество её входных параметров и возвращаемых значений, описание того, что данная подпрограмма делает.

Прикладной программный интерфейс в виде фиксированного, неизменяемого набора правил работы программ с операционной системой впервые возник в резидентных мониторах как набор вспомогательных подпрограмм ввода-вывода и управления. Действительно, пока программы содержали монитор внутри себя, всегда существовала возможность заменить в момент разработки программы одну подпрограмму ввода-вывода на другую, например более эффективную, с иным интерфейсом. В случае резидентного монитора вспомогательные подпрограммы находятся вне программы пользователя и заменить подпрограмму невозможно.

Начиная с этого момента прикладные программы содержат только реализацию основного алгоритма, обращаясь за вспомогательными алгоритмами к операционной системе через прикладной программный интерфейс. Это очень важная веха в развитии операционных систем. Фиксированный программный интерфейс одной операционной системы можно воспроизводить на разных операционных системах для разных ЭВМ. Для прикладных программ это означает возможность без изменений выполняться на разных ЭВМ, т. е. можно говорить о *переносимости* прикладных программ с одной ЭВМ на другую. Иными словами, появилась возможность создавать программы, которые не зависят от аппаратуры и операционной системы.



Переносимость (или *мобильность*) прикладных программ — основа концепции *открытых систем*. Под термином «система» здесь понимается не только операционная система, но и аппаратура, и программы, которые применяются для автоматизации какой-либо прикладной задачи. Операционная система является одной из компонентов этой *информационной* системы. Другими компонентами могут быть система управления базами данных — СУБД, графическая оболочка, сетевые средства — механизмы, необходимые для приложений, но могут, вообще говоря, не предоставляться операционной системой.

Открытая система — это способ построения информационной системы, при котором компоненты системы взаимодействуют друг с другом через стандартные интерфейсы. С точки зрения пользователя, открытая система обязана соответствовать набору открытых стандартов, определяющих интерфейсы, поддерживаемые форматы данных и т. п., обеспечивающих возможность перехода от системы к системе.

Открытый стандарт — стандарт, не зависящий от технологии, специфического аппаратного и программного обеспечения или от продуктов конкретного производителя. Открытый стандарт одинаково доступен и пользователям систем, и програм-

мистам, которые разрабатывают системы, и производителям аппаратуры. Открытый стандарт находится под контролем общественности, и поэтому все, кого он затрагивает, могут участвовать в его разработке.

Считается, что открытая система обладает следующими свойствами:

- *переносимость (мобильность)* — возможность переноса программ и данных при замене аппаратуры и возможность работы с ними пользователей без их переподготовки при изменениях в системе;
- *расширяемость (масштабируемость)* — возможность добавления новых функций в систему или изменения некоторых уже имеющихся при неизменных остальных частях системы;
- *интероперабельность* — этот громоздкий термин означает способность к взаимодействию с другими системами;
- *дружественность к пользователю.*

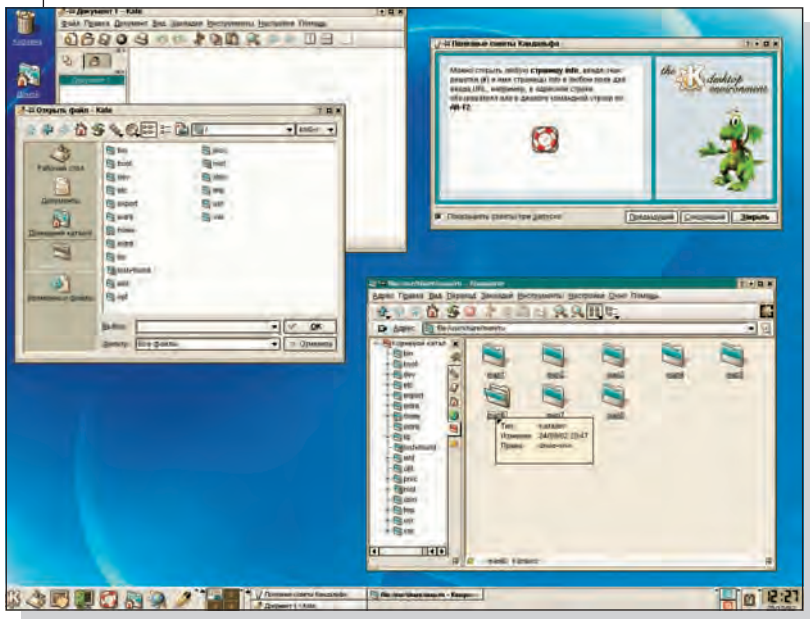
Конечно, система, обладающая такими свойствами, выглядит очень привлекательно как для пользователей, так и для программистов. Безусловно, этими свойствами, взятыми по отдельности, обладали программы, эксплуатировавшиеся задолго до появления концепции открытых систем. Однако в открытой системе, во-первых, эти свойства рассматриваются как взаимосвязанные, а, во-вторых, их реализация гарантируется выбранным набором открытых стандартов.

Для операционной системы, как компонента информационной системы, общепризнанным международным стандартом является POSIX. Полное наименование — ISO/IEC 9945 Information Technology — Portable Operating System Interface, но в литературе часто используется сокращённое название POSIX. Разработка стандарта началась в 1985 г. в Институте инженеров по электротехнике и электронике США (Institute of Electrical and Electronics Engineers, IEEE), и стандарт был известен как IEEE Std 1003. Позднее, в начале 90-х гг. XX в., стандарт был опубликован и Международной организацией по стандартизации (International Organization for

Standardization — ISO) и, таким образом, стал международным.

Наиболее интересна первая часть стандарта, определяющая прикладной программный интерфейс операционной системы. В 1996 г. очередная редакция первой части стандарта была выпущена одновременно Американским национальным институтом стандартов (American National Standards Institute — ANSI) совместно с IEEE и ISO совместно с Международной электротехнической комиссией (International Electrotechnical Commission — IEC) под названиями ANSI/IEEE Std 1003.1-1996 и ISO/IEC 9945-1:1996 соответственно. Одна из «свежих» редакций стандарта была опубликована IEEE в 2001 г. под названием IEEE Std 1003.1-2001.

Стандарт POSIX определяет прикладной интерфейс, характерный для ОС UNIX. Даже сокращение POSIX отражает этот факт: разработчики ОС UNIX в своё время требовали, чтобы операционные системы, созданные на базе их исходных текстов, имели в названии окончание IX. Поэтому часто UNIX употребляют как синоним термина «открытая система». В действительности, во-первых, не все версии ОС UNIX удовлетворяют этому стандарту, во-вторых, ряд операционных систем, никак не связанных с ОС UNIX, поддерживает прикладной интерфейс POSIX. Например, Windows NT и её последующие модификации Windows 2000 и Windows XP, Open VMS, операционная система реального времени VxWorks содержат



Стандартная оконная система ОС Linux.

наряду со своим собственным прикладным интерфейсом интерфейс POSIX.

Хотя POSIX и является международным стандартом, выпущенным авторитетными организациями, ответственными за стандартизацию, для построения полноценной открытой информационной системы его недостаточно. Поэтому наряду с POSIX получили распространение и другие стандарты. Как правило, такие стандарты включают в себя POSIX и добавляют ряд собственных спецификаций. Например, широко известный консорциум The Open Group, объединяющий ведущие компании рынка информационных технологий, создал спецификацию UNIX98, в которую помимо POSIX входят стандарты на графический интерфейс, сетевые средства, язык Java и сервисы Интернета.

ОС UNIX

Сегодня UNIX по праву считается эталоном в мире операционных систем. На протяжении 30 лет любое достижение в области операционных систем, вычислительной техники и информационных технологий так или иначе связано с ОС UNIX. В чём заключается такой успех? Что позволяет UNIX не только оставаться на плаву, но и быть лидером в ряду универсальных операционных систем?

Созданию ОС UNIX предшествовал проект вычислительной системы Multics (1964 г.), объединивший разработчиков из Массачусетсского технологического института, компании General Electric и Bell Telephone Laboratories — научно-исследовательского центра компании AT&T. Цель проекта — операционная система разделения времени, которая могла бы не только работать непрерывно и безотказно 7 дней в неделю, 24 ч в сутки, предоставляя максимальный спектр возможностей, но и, по словам разработчиков, «удовлетворить почти все требования, выдвигаемые к большой компьютерной системе на сегодняшний день и ближайшее будущее». Можно сказать, что удалось выполнить лишь половину поставленной цели: ОС Multics действительно была надёжной и безотказной системой, в компании Ford она проработала с 1974 по 1997 г.

В апреле 1969 г., за несколько месяцев до появления первой версии ОС Multics, руководство компании AT&T решило выйти из проекта и, более того, закрыло все исследования в области построения операционных систем. В Bell Telephone Laboratories эта новость вызвала разочарование, ведь разработчики нуждались в операционной системе, которая позволяла бы не только эффективно решать многие задачи, но и была бы центром, вокруг которого формируется сообщество пользователей. Опыт эксплуатации операционных систем разделения времени показал, что совместная работа нескольких пользователей — это не просто интерактивная работа, а скорее питательная среда для обмена идеями — основы научной деятельности человека.

Группа специалистов из Bell Telephone Laboratories: Кен Томпсон,

Деннис Ритчи, Джо Оссанна, Руд Кенедей и Дуг Мак-Илрой — продолжила разработку отдельных частей экспериментальной операционной системы, в том числе и файловую систему.

В течение 1969 г. Кен Томпсон придумал игру «Космическое путешествие», которая по существу представляла собой программную модель физики движения космических тел в Солнечной системе. Играть в неё на лабораторной ЭВМ было весьма расточительно: расход вычислительных ресурсов, потребляемых игрой, составлял 75 долларов в час. К счастью, Томпсон нашёл подходящую вычислительную аппаратуру — ЭВМ PDP-7 с графическим терминалом. PDP-7 выглядел карманным компьютером по сравнению с сервером. В ходе переноса игры на PDP-7 была создана так называемая *кросс-технология,* когда программы для одной ЭВМ пишутся с использованием другой ЭВМ.

Для операционных систем кросс-технология — необходимый шаг. В начале разработки требуется *создать исполняемый файл* операционной системы (образ ОС), которой ещё не существует. Это напоминает известный вопрос: «Что появилось раньше: курица или яйцо?». Решение состоит в разработке инструментальных средств, способных создавать исполняемые файлы на другой ЭВМ.

Для «Космического путешествия» понадобилась специальная *среда исполнения*. Здесь и пригодилась файловая система Bell Telephone Laboratories. Дополненная простейшей подсистемой управления процессами, она сформировала простейшую операционную систему разделения времени. Эта ОС была многопользовательской (к PDP-7 подключались два терминала!), но каждый пользователь мог одновременно запустить только одну программу. Для системы были созданы небольшой набор пользовательских утилит, расчитанных на работу с файлами (текстовый редактор, утилиты копирования, удаления, печати файлов), и командный интерпретатор — программа, с помощью которой пользователи запускали все остальные программы. Именно в таком виде вышла первая версия операционной системы на ЭВМ PDP-7. Чуть позже, в 1970 г., Брайан Керниган предложил имя для этой ОС — UNICX (*англ.* Uniplexed Information and Computing Service — «Простая информационновычислительная служба»), это антоним Multics (*англ.* Multiplexed Information and Computing Service — «Сложная информационно-вычислительная служба»). Вскоре название редуцировалось до UNIX.

Ключевой момент в истории UNIX — изобретение Дугом Мак-Илроем и Рудом Кенедеем в 1973 г. *программных каналов* (связывающих поток вывода одной программы с потоком ввода другой, чтобы получить *конвейер команд*), после чего эта ОС и стала ОС UNIX в современном понимании.

Для иллюстрации работы программного канала можно рассмотреть следующий пример. Есть книга, которая состоит из глав, каждая глава расположена в отдельном файле. Нужно напечатать книгу с нумерацией страниц. В UNIX это может выглядеть так:

```
cat глава1 глава 2 глава3 глава4
глава5 | mpage -H | lpr
```

Здесь `cat` — программа, которая читает файлы `глава1` — `глава5` и записывает их в свой поток вывода; `mpage` — программа, которая форматирует текст, разбивает его на страницы и, добавляя номер страницы, записывает всё это в свой поток вывода; `lpr` — вывод на печать потока ввода. Знак «|» связывает поток вывода с потоком ввода через программный канал.

С появлением каналов ОС UNIX приобрела новое качество: из операционной системы она превратилась в *набор инструментальных средств* программиста, значительно облегчающий программирование прикладных задач. UNIX предложила новый подход в использовании вычислительной техники: комбинирование универсальных программ вместо разработки единой самодостаточной системы специального назначения.

Изобретение программных каналов — такое же значимое событие, как и изобретение первого спутника

Кен Томпсон.

Деннис Ритчи.

Дуг Мак-Илрой.

Билл Джой.

связи, кнопочного телефонного аппарата, мобильной связи, оптоволоконных линий связи. Все эти изобретения упоминаются на выставке достижений в области телекоммуникаций, расположенной в холле здания лабораторий Белла (г. Мюррей-Хилл, штат Нью-Джерси, США).

В 1973 г. ОС UNIX была переписана на язык программирования Си, изобретённый Керниганом и Ритчи специально для UNIX. Это был ещё один революционный шаг: языки такого высокого уровня при создании операционных систем не использовались.

В октябре 1973 г. Томпсон и Ритчи представили доклад на симпозиуме по принципам построения операционных систем. Доклад в июле 1974 г. был опубликован в журнале «Communications of the ACM». В это время компании AT&T запрещалось продавать программное обеспечение, так как AT&T являлась монополистом на рынке телефонных услуг, и по законам федерального правительства США компания не могла вести иную коммерческую деятельность. Поэтому UNIX вместе с исходными текстами мог получить любой желающий. Это и определило последующий успех UNIX: при наличии исходных текстов программное обеспечение легко настроить под те или иные потребности, добавив при необходимости нужные компоненты или исправив существующие ошибки.

ОС UNIX заинтересовала университеты, где стали использовать её исходные тексты в процессе обучения программированию. Возникло сообщество пользователей UNIX, о чём и мечтали её разработчики.

UNIX стала мобильной операционной системой в 1978 г. Тогда же в австралийском университете Волонгонга программисты Джюрис Рейнфельдс и Ричард Миллер осуществили перенос UNIX на ЭВМ Interdata 7/32. Чуть позже в Принстонском университете Том Лайон закончил перенос UNIX на ЭВМ IBM 360, а Деннис Ритчи и Стив Джонсон выполнили перенос UNIX на ЭВМ Interdata 8/32, доработав для этого язык Си.

В январе 1979 г. вышла очередная, 7-я редакция UNIX с лицензией, запрещающей использование исходных текстов в качестве учебного пособия. Компьютерный мир вступал в эру рабочих станций, вместо шкафа высотой 2 м использовались недорогие ЭВМ с процессором в виде кристалла. Соответственно производителям требовалась мобильная операционная система, чтобы минимизировать затраты на производство программного обеспечения для новой аппаратуры. AT&T продолжала выпускать новые версии UNIX уже на коммерческой основе (самые популярные UNIX System III — 1981 г. и UNIX System III — 1983 г.), но имя UNIX перестало ассоциироваться с Bell Telephone Laboratories, авторами UNIX стали университеты.

Наиболее широкую известность получила группа разработчиков из Калифорнийского университета г. Беркли, выпускавшая свой вариант UNIX, известный как BSD (Berkeley Software Distribution).

В 1976 г. выпускник Беркли Кен Томпсон провёл свой академический отпуск, работая преподавателем университета. Он установил там Исследовательский UNIX и написал компилятор для языка Pascal. В 1977 г. студент Билл Джой собрал коллекцию программного обеспечения и выпустил первую версию BSD.

Популярность BSD UNIX обусловлена тремя факторами:

- демократичная лицензия: BSD UNIX мог получить любой желающий вместе с исходными текстами;
- возможность работы в сети («стандартная система Интернет» — это немаловажная характеристика);
- удачный выбор платформы: VAX — «ЭВМ первой половины 80-х гг.». Помимо всего перечисленного, BSD UNIX — это хорошая система, её файловая система и подсистема управления памятью долгое время оставались самыми быстрыми в мире UNIX.

В 1982 г. Билл Джой ушёл из Беркли, основав компанию SUN Microsystems. Эта компания до настоящего времени выпускает рабочие станции и серверы, первоначально предназ-

наченные для работы под управлением UNIX (в версии от SUN UNIX называлась SunOS, а потом была переименована в Solaris). Ориентация на UNIX позволила продукции SUN лидировать во второй половине 80-х гг. Достижение SUN в области UNIX — создание сетевой файловой системы NFS, которая разрешает работать с файловыми системами удалённых ЭВМ, объединённых сетью.

Следующий этап развития UNIX связан с Linux, который придумал студент второго курса отделения вычислительной техники факультета естественных наук Университета Хельсинки Линус Торвальдс. Во время учёбы он изучал курс по теории операционных систем, в основе которого была ОС Minix — UNIX-подобная ОС, созданная знаменитым специалистом в области операционных систем Эндрю Таненбаумом специально для учебных целей. Поначалу Торвальдс хотел, чтобы система была сделана немножко лучше, чем Minix, и не претендовал на что-то серьёзное — это было просто хобби. В сентябре 1991 г. Linux версии 0.01 был выложен для свободного копирования в Интернете.

Стремительному развитию проекта Linux способствовали два фактора. Во-первых, в начале 90-х гг. уже не существовало проектов свободнораспространяемых операционных систем, которые находились бы в начальной стадии развития. Следовательно, войти в группу разработчиков и принять участие в таких проектах было довольно сложно: требовалось обладать запасом знаний о конкретных деталях той или иной операционной системы. Linux же была небольшой операционной системой и идеальным полигоном для испытаний новых идей. Во-вторых, новые версии ОС Linux выпускались очень быстро, изменения между версиями были небольшими, что позволяло легко находить и устранять ошибки в коде. Рекордной можно считать серию ядер Linux 2.1, в рамках которой было выпущено чуть менее 150 версий.

Сейчас Linux — это система миллионов пользователей.

Линус Торвальдс.

Эндрю Таненбаум.

МУЛЬТИМЕДИА

ПЕРЕХОД К ЦИФРОВОМУ ПРЕДСТАВЛЕНИЮ ИНФОРМАЦИИ

Информационный кризис XX в. показал, что традиционные, аналоговые способы хранения информации устарели. Необходим был переход к цифровому представлению данных с использованием компьютеров для хранения и обработки информации. На рубеже тысячелетий стало окончательно ясно, что будущее — за цифровыми технологиями.

Перевод информации в цифровой вид не всегда лёгкая задача. Информация может быть различной по форме: аудио, видео, текст и даже произведение искусства. Возникают две проблемы: первая связана с представлением различных типов информации, другая — с организацией ввода информации в компьютер.

Самой простой оказывается задача представления текстов, где мысли, образы, понятия уже закодированы с помощью слов, которые, в свою очередь, состоят из букв, а их легко преобразовать в цифровую форму.

Сложнее обстоит дело с иероглифическим письмом, где фигурные знаки обозначают целые понятия и слова или отдельные слоги и звуки речи. В японском языке используется несколько тысяч иероглифов, представляющих корни слов, и несколько десятков знаков «каны», передающих отдельные слоги и служащих для записи грамматических окончаний. Присвоить каждому иероглифу свой код — возможное, хотя и неэффектив-

Компьютеры работают по цифровой технологии, в них информация представлена не значениями непрерывно меняющихся величин, а устойчивыми состояниями физических устройств. Для цифровой обработки информации нужны физические устройства, которые обладают двумя устойчивыми состояниями и способны по сигналам извне переходить из одного состояния в другое.

ное решение. Чаще цифровое представление иероглифических текстов основывается на фонетическом принципе.

Ввод текста в компьютер посредством клавиатуры по скорости и удобству (текст сразу представлен в цифровой форме) часто превосходит обычное письмо. Однако столь современный метод ввода текста неприемлем, если текст существует в напечатанном или рукописном виде. Во-первых, неизбежны ошибки, во-вторых, просто нецелесообразно так непроизводительно использовать человеческий труд. Подобные тексты могут быть введены с помощью сканера, считывающего текст как графическую информацию, картинку. Специальная программа распознавания переводит графику в текстовый документ. Задача распознавания образов ещё далека от совершенства, и плохой почерк до сих пор остаётся загадкой для компьютера.

Современные системы обработки текстов помимо ввода и редактирования обладают такими возможностями, о каких создатели первых текстовых редакторов лишь мечтали. Перечислим некоторые из них: автоматическая проверка орфографии в процессе набора, интеллектуальная вёрстка текста, создание абсолютно новых, не имеющих «бумажных» аналогов гипертекстовых и мультимедийных документов, содержащих не только текст, но также звуковое сопровождение и анимацию. Всё это стало осуществимо благодаря переводу текстовой информации в цифровую форму.

АУДИОИНФОРМАЦИЯ

Компакт-диск, или DVD-диск, может хранить информацию о звуке, изображении, письме в виде миллиардов двоичных нулей и единиц. Долговечность лазерного диска (сотни лет) во многом объясняется самими принципами цифровой технологии. В отличие от виниловой пластинки лазерный диск не боится пыли, а мелкие царапины не ухудшают качества воспроизведения.

Иногда приходится слышать о механичности, бездушности цифрового звука. В современных аудиоустройствах ламповые усилители сигнала с чисто аналоговыми характеристиками заменены на более дешёвые и экономичные, но и более «дискретные» транзисторы. Вероятно, эта «правильность» и может обусловливать подобное восприятие оцифрованного звука. Любители качественного натурального (high-end) звука продолжают пользоваться дорогими ламповыми усилителями и виниловыми пластинками.

Звуки представляют собой большое число простейших синусоидальных колебаний разной частоты. Человек в среднем воспринимает звук от 50 Гц до 18 кГц. АудиоCD записываются с частотой дискретизации 44,1 кГц 16 бит стерео. Это значит, что для каждого из двух стереоканалов амплитуда сигнала замеряется 44,1 тыс. раз в секунду, для хранения её цифрового представления используется 16 бит. При такой дискретизации гармонические колебания с частотой, большей 22,05 кГц, отсекаются при записи, что вообще-то не страшно, поскольку даже эта граница практически равна границе восприятия человека. Кодирование с использованием психоакустических особенностей восприятия звука позволяет ещё сильнее уплотнить звук на компакт-диске (примерно в 5—7 раз), сделав эти потери незаметными для человека.

В 1992 г. фирма Sony представила новый носитель записи звука мини-диск (MD).

Цифровая обработка звука, так же как и текста, применяется практически на каждом шагу. Большинство

Если вы не можете разобрать собственный почерк, но набираете текст на компьютере со скоростью свыше 70 слов в минуту, то вы, скорее всего, не гений, а просто современный человек...

Томограф.

музыкальных центров имеют встроенный цифровой процессор, представляющий собой особую микросхему для обработки звуковых сигналов в цифровой форме. Современные DVD-проигрыватели и видеомагнитофоны также имеют встроенные процессоры для обработки звука, по качеству не уступающие Dolby AC3-системам, которые устанавливают в кинотеатрах. На цифровую запись и обработку звука перешли многие радиостанции и телестудии. Даже дома записи на MD (мини-диски) часто производят не старым, аналоговым, способом (с потерей качества при перезаписи), а цифровым (точная копия исходной записи).

Ещё одним направлением постоянных исследований является распознавание голоса, чтобы использовать для ввода текста ЭВМ.

Оборудование для создания объёмной фотографии поступило в продажу уже в 2002 г.

ИЗОБРАЖЕНИЕ

Для ввода изображений с фотографий или гравюр можно использовать всё те же сканеры, так как они позволяют вводить их с очень хорошим разрешением. Уже нет смысла сначала делать фотографии обычным фотоаппаратом, а потом сканировать отпечатанное изображение или слайд. Современные цифровые фотоаппараты не уступают по качеству снимков своим аналоговым собратьям. То же относится и к цифровому видео.

Сложной пока остаётся проблема представления пространственных объектов (однако подробные объёмные фотографии уже возможно создавать). Существует два основных способа такого представления — задание сетки точек на поверхности рассматриваемого объекта и приближение объекта различными геометрическими фигурами, имеющими строгое математическое представление. Первый способ применяется при описании сложных объектов, таких, как скульптура или сам человек. Второй способ широко используется в системах автоматического проектирования (САПР), в станках с числовым программным управлением (ЧПУ) для производства необходимых деталей.

Медики на основе серии параллельных снимков, получаемых на специальном приборе — томографе, строят трёхмерное изображение исследуемых органов и частей тела человека. Без компьютерной обработки таких изображений нельзя проводить современные операции с использованием лазеров, например удаление внутренних опухолей без разрезов или повреждения соседних тканей. Компьютер способен не только видоизменять уже существующую информацию, но и создавать новую.

Обработка графики и видео значительно облегчается при использовании цифровой информации. Профессиональные графические редакторы, такие, как Adobe Photoshop или Corel Draw, предоставляют большой выбор различных эффектов, цифровых фильтров, настроек в домашних условиях. А ведь раньше для этого требовалась

целая фотолаборатория со сложным и дорогим оборудованием.

Немалую роль играет цифровая технология в кино. Речь здесь идёт даже не об удобстве цифрового нелинейного монтажа. Фильм «Звёздные войны. Эпизод I» был частично снят цифровой видеокамерой. Режиссёр фильма Джордж Лукас справедливо утверждал, что никто не сможет распознать этот цифровой кусок. А следующий фильм — «Звёздные войны. Атака клонов» — стал одним из первых фильмов с использованием только цифровых технологий. Процесс спецэффектов упростился, и изображение, сделанное с помощью компьютера, ничем не отличается от реального.

ВЕКТОРНАЯ И РАСТРОВАЯ ГРАФИКА

Айвен Сазерленд, пионер в применении компьютера для построения изображений, как-то сказал: «Дисплей, подключённый к ЭВМ, представляется мне окном в Алисину страну чудес... С помощью дисплея я сажал самолёт на палубу авианосца, следил за движением элементарной частицы в потенциальной яме, летал на ракете с околосветовой скоростью и наблюдал за таинствами внутренней жизни вычислительной машины».

В 80-х гг. рынок персональных компьютеров был фактически захвачен клонами компьютеров IBM PC. Вехой в развитии самих IBM PC явился разработанный в 1987 г. стандарт VGA (*англ.* Video Graphics Array — дословно «видеографическая матрица» или «массив»). Видеоадаптеры, платы осуществляющие преобразование информации в изображении на экране, стали использовать более высокое разрешение (640×480 точек) и бóльшую глубину представления цвета на мониторе компьютера (256 цветов вместо 16). Появилась возможность видеть на экране не только стилизованные картинки низкого мультипликационного качества, но и фотографии. Белая мышка на цветном экране выглядела как живая.

Экран персонального компьютера представлял собой растр точек, или пикселей (от *англ.* picture element). Стандарт VGA комплектовался памятью 256 кбайт, где поэлементно хранились картинки, образованные

Создание незатейливых рисунков на обычных печатающих устройствах относится ещё к первым годам появления ЭВМ. В 1950 г. в Массачусетсском технологическом институте компьютер Whirlwind был оборудован дисплеями на электронно-лучевых трубках, позволяющих выводить изображения.

Растровый и векторный цветок.

Разработанный в 70-х гг. в исследовательском центре фирмы Xerox многооконный интерфейс использовался как на компьютерах этой фирмы, так и на персональных ЭВМ Lisa фирмы Apple Computer, Inc. Интерфейс требовал одновременного расположения на экране монитора графики и текста. Новые системы были построены по принципу WYSIWYG (*англ.* What You See Is What You Get — «Что вы видите, то и имеете»), в соответствии с которым изображение на экране должно быть как можно более похожим на печатный вариант.

Луч в электронно-лучевой трубке попадает на экран, покрытый люминофором. Он выделяет лучистую энергию (интенсивность света зависит от кинетической энергии электронов). Поскольку свет излучается люминофором лишь в течение нескольких миллисекунд, всё изображение необходимо регенерировать (возобновлять) с частотой около 30 раз в секунду, а иногда и чаще.

горизонтальными строками развёртки на дисплее (экране кинескопа).

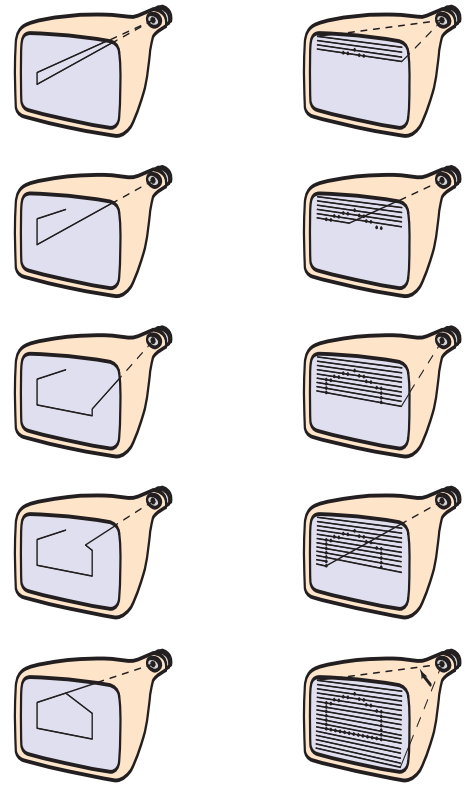
Для увеличения растра, например до 800×600 точек (800 точек по горизонтали и 600 — по вертикали), при сохранении *глубины цвета* 8 бит на точку (256 цветов) требовалось увеличить объём памяти в два раза (800×600 = 480 000 точек, т. е. около 480 кбайт). При растрах 1024×768, 1280×1024, 1600×1200 памяти надо ещё больше. 256 цветов недостаточно для получения полноценной картинки, поэтому глубина цвета была увеличена: 16 бит на точку — high color (*англ.* «лучший цвет»), 24 бит на точку — true color (*англ.* «настоящий, реальный цвет»). Последний назван так потому, что изображение с использованием подобной глубины цвета (16,7 млн цветов) практически идеально для глаза.

Легко подсчитать, что если на каждую точку приходится по 24 бит (т. е. 3 байт), то для хранения картинки 800×600 точек потребуется около 1,4 Мбайт (800×600×3 = 1 440 000 байт). Это неэкономно, если фигура небольшая, например треугольник. Но существуют и другие способы хранения графической информации.

Пусть художник рисует карандашом на бумаге и на запястье у него специальное устройство, передающее в компьютер все движения руки. Тогда законченный рисунок можно запомнить (и потом нарисовать) не только как набор точек (растр), но и как последовательность движений руки художника. Такой метод представления графической информации называют *векторным*. Рисунок задаётся векторами, указывающими координаты начала отрезка и относительное смещение. Если предположить, что координаты отрезков двухбайтовые, то для хранения треугольника в памяти потребуется всего 2×2+2×3×2 = = 16 байт (координаты начальной точки и трёх векторов).

ВЕКТОРНЫЙ И РАСТРОВЫЙ ДИСПЛЕЙ. СХЕМЫ

В векторном дисплее электронный луч способен непрерывно проходить между любыми двумя точками, порождая при этом чёткую прямую линию. Буквы в таких дисплеях состоят из коротких векторов. Дисплей также уме-



ет рисовать различные простейшие фигуры, например дуги и линии, которые задаются двухмерными координатами. Запоминание позиций для последующей перерисовки — дело самого дисплея.

Векторная графика сначала чрезвычайно широко применялась в дисплеях, потому что требовала малых объёмов памяти. Основной её недостаток — невозможность изобразить сплошные области, поскольку объекты (как плоские, так и трёхмерные) представлены в виде «проволочных» каркасов. Если на экране слишком сложная векторная картинка, состоящая из большого числа элементарных фигур, то она не будет успевать обновляться, и изображение начнёт мерцать.

В растровых дисплеях, в отличие от векторных, луч пробегает весь кадр (как в телевизоре), при этом управляют не отклонением, а только интенсивностью лучей в цветном дисплее. Одно из достоинств растровых дисплеев заключается в том, что представление картинки не зависит от её сложности (т. е. числа составляющих её объектов), поэтому проблемы мерцания не существует.

В то время как на векторном дисплее луч вычерчивается так же, как и при работе с линейкой и рейсшиной, работа растрового дисплея основана на электронной аналогии *пуантилизма* (манера живописи мазками в виде точек).

При этом методе не только могут оказаться заметными отдельные пиксели, но и все не вертикальные или не горизонтальные линии имеют ступенчатые края. Подобный эффект «зазубренности» сводится к минимуму при увеличении разрешающей способности, т. е. когда при том же размере экрана возрастает количество элементов растра. Векторные картинки легко масштабируются, но этого нетрудно добиться и на растровом дисплее, при использовании специального программного обеспечения.

К 90-м гг. XX столетия растровые дисплеи (и персональные ЭВМ) практически вытеснили с рынка вычислительной техники векторных конкурентов.

Жорж Сёра. Воскресенье после полудня на острове Гранд-Жатт. 1884—1886 гг. Институт искусства. Чикаго.

ФОРМАТЫ ГРАФИЧЕСКИХ ФАЙЛОВ

Для того чтобы экспортировать или сохранить изображение в файле, на диске были придуманы форматы файлов (методы, описывающие кодирование изображений). Наиболее популярные из них — BMP, PCX, GIF, PNG, TIFF и хорошо знакомый из цифровой фотографии JPEG.

BMP (*англ.* Bit MaP — «битовая карта») — «родной» формат графических файлов для ОС Windows компании Microsoft, поскольку наиболее точно соответствует внутреннему формату системы.

Формат изображения так называемого *растрового массива* зависит от количества битов, используемых для кодирования цвета каждого пикселя (1, 4, 8, 16 или 24 бит на пиксель). Например, при изображении с 256 цветами под каждый пиксель отводится 1 байт (8 бит), который содержит номер цвета в *таблице цветов* файла. Таким образом, если в цветовой модели RGB в таблице цветов файла BMP хранится R/G/B =255/0/0, то значению пикселя 0 (в растровом массиве) будет поставлен в соответствие яркокрасный цвет.

Таблица цветов, или *палитра,* представляет собой массив, где индекс массива — это номер цвета палитры, а его содержимое — тройки чисел, описывающие цвет по яркости трёх

Пуантилизм (от *фр.* pointiller — «писать точками») придуман в XIX в. французским художником-неоимпрессионистом Жоржем Сёра.

Появление Windows 95 и 98 привнесло в формат PCX некоторые изменения: число битов на точку может достигать 32, а размер изображения с 2^{16} бит вырос до 2^{31} бит, т. е. до 2 Гбит. Размер можно задавать отрицательным числом, указывающим на «направление» прорисовки изображения (с левого верхнего угла, а не с левого нижнего).

Можно представить себе графический файл как поток грузовиков, везущих бочки (байты) на склад (видеоплату). При поступлении на склад содержимое бочки делят на части и расставляют на определённые полки. Уходит немало времени, ведь надо открыть каждую бочку, достать необходимую часть, переложить на полку и т. д. А если заранее распределить её по бочкам так, чтобы содержимое одной бочки соответствовало отдельной полке, время на работу сократится. Это и есть аппаратная зависимость.

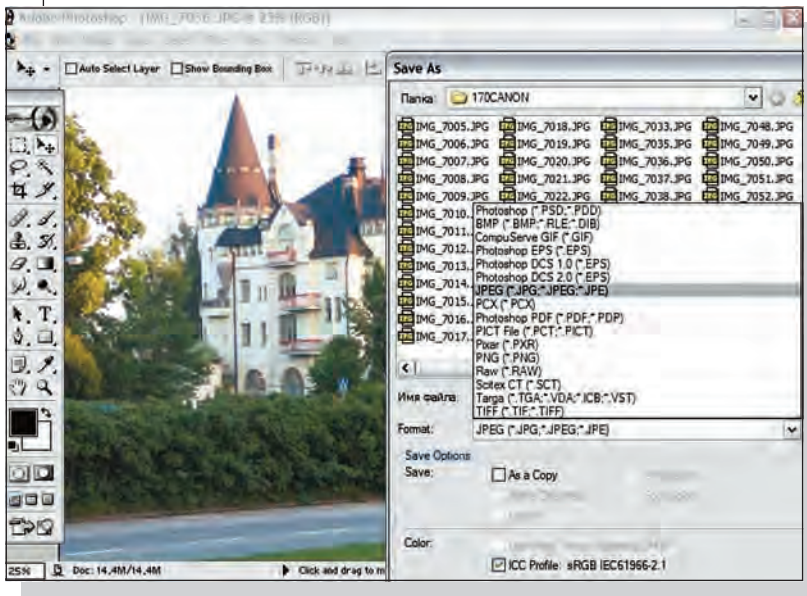
РАЗРЕШЕНИЕ

Разрешение важно при печати, когда требуется сохранить реальный размер оригинала. Когда сканируют фотографию, происходит преобразование изображения в дискретную (цифровую) форму. Изображение получает определённый размер в пикселях, и, чтобы при этом не потерялась информация о реальном исходном размере фотографии, требуется хранить разрешение, указывающее, сколько точек изображения помещалось в 1 см или в дюйме.

Разрешение (как горизонтальное, так и вертикальное) вычисляется делением получившегося размера в пикселях на соответствующий размер исходного изображения.

Если изображение имеет 800 точек по горизонтали, а длина исходной фотографии — 12 см, то при сканировании используется разрешение 800/12 = 67 точек в 1 см.

Чем выше разрешение, тем качественнее получится изображение при печати, но тем больше будет растровый массив и соответственно файл займёт больше места.



основных компонентов: R/G/B (красного, зелёного, синего).

Файлы BMP с глубиной представления цвета 16 и 24 бит на пиксель не имеют таблицы цветов. В них значения пикселей растрового массива непосредственно задают значения цветов RGB. Пиксели в формате изображения хранятся в порядке слева направо, начиная, как правило, с нижней строки изображения. Если число байтов в строке развёртки нечётное, то к такой строке добавляется ещё один байт, чтобы «выровнять» данные растрового массива.

PCX — стандарт, разработанный фирмой ZSoft, используемый для программы PC Paintbrush в ОС MS DOS компании Microsoft. Но PCX не стал так же популярен, как BMP. Тем не менее PCX распространён на PC-технике. Специализированные графические редакторы — Corel Draw, Ulead PhotoExpress, Adobe PhotoShop — поддерживают данный формат.

PCX — это аппаратно-зависимый формат. Он предназначался для того, чтобы информация в файле хранилась, так же как и в видеоплате. Раньше на IBM PC существовали 16-цветные EGA-адаптеры, память которых делилась на непрерывные куски-плоскости *(планы)*. Составляющие цвета пикселя (как правило, 1 бит) находились в соответствующих местах плана. Чтобы поставить точку определённого цвета, требовалось в четырёх местах исправить по 1 бит. Получалось, что цветная картинка как бы состояла из четырёх монохромных (не путать с чёрно-белыми!) планов. Для поддержания совместимости со старыми программами современные видеоадаптеры умеют работать в подобных EGA-режимах.

При true color в PCX всегда есть три плана, а один план бывает только при 8-битном изображении (именно тогда в конце файла находится палитра). Формат PCX допускает хранение в растровом массиве изображения большего, чем картинка, видимая на экране. Левая верхняя граница задаёт отступ от верхнего угла изображения, а разница между границами указывает размер видимой части. Это было приду-

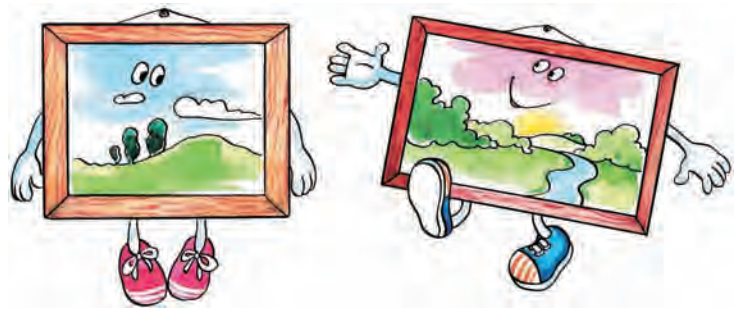
мано для использования данного формата в факсимильных аппаратах. Кроме того, для более правильной привязки изображения к экрану в файле существует два поля, описывающие разрешение 144 точки на дюйм.

GIF (*англ.* Graphics Interchange Format — «формат для обмена графикой») первоначально предназначался для сети CompuServe. Формат GIF устроен иначе, чем PCX и BMP. В GIF графическая информация состоит из блоков, которые следуют один за другим (нельзя сказать, где начинается 17-й блок, пока не прочитан 16-й). Поэтому говорят, что информация хранится в виде *потока данных* (при чтении файла информация льётся непрерывно, как вода).

Формат GIF работает не более чем с 256 цветами. Зато каждый файл может содержать не одно, а несколько изображений. Существует две версии формата GIF, различаемые по своим кодам, — GIF87*a* и GIF89*a*. GIF87*a* выпущена в мае 1987 г. Каждый файл состоит из заголовка, необязательной палитры и собственно информации о картинке. *Глобальная палитра* используется для всех изображений, находящихся в файле, если у какого-то изображения нет собственной палитры. Палитры картинок и глобальная палитра устроены совершенно одинаково (как в BMP): тройки байтов задают красную, зелёную и синюю составляющие цвета́ точки.

Файлы формата GIF при загрузке из сети появляются своеобразно: сначала нечётко, потом всё яснее. Если нет высокоскоростного подключения к Интернету, то можно наблюдать загрузку файла «вживую». Сначала будут переданы 0-, 8-, 16-я строки картинки (т. е. каждая восьмая строка), со второй строки чередование пойдёт через четыре строки: 2-, 6-, 10-я... И последний проход — с первой строки все нечётные: 1-, 3-, 5-я... Такое изображение позволяет увидеть всю картинку, даже если получена лишь половина изображения.

Версия GIF89*a* создана в июле 1989 г. Она дополнена новыми информационными блоками, которые названы *управляющим расширением*.



Выделено четыре типа информационных блоков.

• Управление графикой используется для создания движущихся картинок (*англ.* animated GIF).

• Блок текста позволяет располагать текст поверх изображения. Это особенно удобно, когда нужно поменять текст, оставив неизменным фоновое изображение.

• Комментарии пропускаются при выводе изображения, но они полезны авторам, чтобы указать подробную информацию (дата создания, варианты, эскизы).

• В приложениях хранят специальные данные, которые воспринимаются специальными программами. Другие программы просто пропускают этот блок.

К недостаткам GIF-файлов можно отнести небольшую глубину цвета (256 цветов) и то, что данные растрового массива сжимаются с помощью закрытого алгоритма LZW, т. е. если создаваемая программистом программа использует GIF-формат, то надо иметь платное лицензионное соглашение с фирмой CompuServe.

Как BMP и PCX, GIF использует специальный алгоритм сжатия данных, чтобы объём дисковой памяти при хранении стал меньше. Изображения хранятся в виде потока блоков размером не более 255 байт, каждый блок состоит из байта (длины блока) и собственно данных. Когда картинка заканчивается, следует блок нулевой длины.

Фрагмент фотографии, обработанной при помощи алгоритма сжатия JPEG.

PNG (*англ.* Portable Network Graphic Format — «формат для передачи графики по сети», а произносится как «пинг») создан по инициативе группы программистов из США (23 человека), возглавляемой Томасом Бутеллом.

Большинство форматов графических файлов разрабатывались одним-двумя программистами, и как результат созданный формат было трудно модифицировать под новые требования. PNG-формат разработан специально созданным комитетом, особое внимание уделялось документированию PNG, поэтому все его последующие модификации не требуют доработки уже работающих с ним графических редакторов.

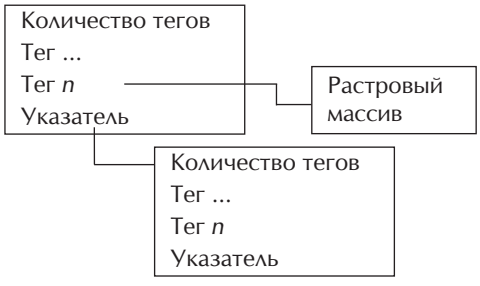
PNG разрабатывался как альтернатива GIF, причём PNG и GIF89*a* имеют много общего, в частности, они организованы как поток данных, могут поэтапно выводить изображение и делать прозрачным основной цвет и т. д. Однако PNG позволяет хранить не только изображение true color (24 бит), но и 48 бит на пиксель. PNG использует открытый алгоритм сжатия LZ77 (он, в отличие от LZW, не требует лицензионных платежей), обеспечивающий сжатие без потерь.

TIFF (*англ.* Tagged Image File Format — дословно «формат изображений, снабжённый "ярлыками" — тегами»), разработанный Aldus Corporation в 1986 г., предназначался в качестве стандарта для хранения чёрно-белых изображений, введённых со сканеров. Это наиболее широко поддерживаемый формат растровой графики, который применяется в настольных издательских системах. Версия TIFF 4.0 позволяла хранить несжатые цветные изображения. Очередная модификация, появившаяся в августе 1988 г., уже использовала цветовую палитру и алгоритм сжатия LZW. В июне 1992 г. версия TIFF 6.0 поддерживала не только RGB-изображения, но и кодировку CMYK, YUV, а также алгоритм сжатия JPEG.

TIFF считают одним из самых сложных форматов графических файлов. Файл состоит из списка так называемых *тегов,* или полей, и растровых массивов. Положение списка и массивов не фиксировано, вероятно, поэтому формат и считается сложным. Устройство TIFF похоже на устройство файловой системы: список — это директория, каждый элемент списка (тег) — имя файла, растровый массив — данные файла.

Фотография без сжатия.

Как и в файловой системе, тег может содержать ссылку (указатель) на растровые данные, а директория — указатель на следующую директорию.



TIFF имеет более 70 различных типов тегов. Тег одного типа хранит, например, информацию о ширине изображения в пикселях, другого — информацию о высоте изображения, следующего — таблицу цветов... Также можно определить собственные типы тегов или игнорировать непонятные.

JPEG назван так по начальным буквам комитета Joint Photographic Experts Group (Объединённая группа экспертов по фотографии; сформирована в 1987 г., входит в Международную организацию по стандартизации — ISO).

Сравнительная таблица графических форматов

Формат	BMP	PCX	TIFF	GIF	PNG	JPEG
Число бит/пиксель	24	24	24	8	48	24
Количество цветов	16,7 млн	16,7 млн	16,7 млн	256	281,475 млрд	16,7 млн
Размер изображения	65 535 x 65 535	65 535 x 65 535	4,3 млрд	65 535 x 65 535	2 млрд x 2 млрд	65 535 x 65 535
Методы сжатия	RLE	RLE	LZW, RLE и др.	LZW	LZ77	JPEG с потерей информации
Количество изображений	1	1	1 и более	1 и более	1	1 и более

ЧТО МОЖЕТ 3D-УСКОРИТЕЛЬ

Минимальный элемент, с которым имеют дело ускорители, — треугольник. Вся изображаемая картинка предварительно разбита на треугольники. Одноцветные треугольники не могут создать эффект объёмной картинки, поэтому на многие из них в процессе прорисовки будут наложены текстуры (изображения, накладываемые на всю поверхность сразу, передающие внешний вид материала, как, например, рисунок на ковре, песок на пляже). Для изображения кирпичной стены здания потребовалась бы прорисовка множества граней для моделирования кирпичей, однако текстура даёт больше реализма и использует меньше вычислительных ресурсов. При наложении текстуры на поверхность надо учитывать перспективу. Эта коррекция необходима, чтобы объекты с текстурой имели различный вид в зависимости от того, насколько они близки к наблюдателю. Для ускорения процесса можно хранить несколько серий однотипных текстур и при выборе текстуры ориентироваться на уже изображённые на экране детали. Соответственно если объект уменьшается, то и размер его текстурного покрытия уменьшается тоже.

Часто при наложении текстур заметны швы между двумя ближайшими объектами. Для подавления этих эффектов применяют специальные методы фильтрации. Ступенчатые края на прямых линиях — давний бич растровых дисплеев. Для получения чётких краёв изображения создают плавный переход от цвета края к цвету фона. Цвет точки, лежащей на границе объектов, определяется как нечто среднее между цветами двух граничных точек, при этом иногда побочным эффектом является смазывание краёв.

Большинство ранее разработанных методов сжатия практически непригодны для сжатия изображений, содержащих сотни тысяч цветов. Формат JPEG является методом сжатия с потерей информации. И хотя он не был определён в качестве стандартного формата графического файла, на его основе создано много новых форматов и улучшено немало старых.

ТРЁХМЕРНАЯ ГРАФИКА

Обычные плоские рисунки моделируются из отрезков, окружностей и многоугольников. В трёхмерной графике к двум координатам добавляют координату z и новые объекты: многогранники, пирамиды, сферы, цилиндры и трёхмерные поверхности. Сплошные объекты можно также формировать из простых. После задания объектов трёхмерной сцены встаёт проблема их визуализации. Если изображение нестатично (например, в игре или тренажёрах), надо не только пересчитывать 30 раз в секунду все положения трёхмерной сцены, но и изображать на экране последовательно сменяющие друг друга картинки.

Даже для техники XXI в. это достаточно дорогой процесс, поэтому трёхмерная сцена представлена в виде совокупности более простых объектов, а специальный графический процессор (так называемый 3D-ускоритель, интегрированный в видеокарту) строит и изображает картинку на экране. То есть современная видеокарта персонального компьютера (совместно с монитором) превратилась из чисто растрового дисплея в интеллектуальный графический дисплей со встроенным 3D-ускорителем.

Чего ожидать в будущем? Все идеи компьютерной графики XX в. воплотились в домашних персональных компьютерах XXI в. Кажется, что предел достигнут.

Основная проблема при построении трёхмерной сцены — как определить, какие из объектов изображения видимы и как они освещены. Необходимо иметь информацию и о взаимном расположении объектов. Обычно для решения этой задачи при-

меняют *Z-буфер*. В *Z*-буфере хранятся значения удалённости всех видимых пикселей (координаты *z*), по одному значению на каждый пиксель растра. Когда подсчитан очередной пиксель изображаемого объекта (с координатами *x*, *y*), полученная координата *z* сравнивается со значением координаты *z*, хранящейся в *Z*-буфере. Если новый пиксель имеет глубину больше записанной в *Z*-буфере, значит, эта точка объекта закрыта каким-либо другим объектом и не отображается на экране. Наоборот, если пиксель оказывается на переднем плане, то значение его координаты *z* записывается в *Z*-буфер. Этот пиксель и будет виден на экране, пока другой объект его не закроет. После вычисления трёхмерной картинки в изображении используются все пиксели, оставшиеся в *Z*-буфере.

Z-буферизация, встроенная в 3D-ускоритель, сильно увеличивает скорость просчёта картинки, несмотря на то что *Z*-буфер занимает большие объёмы памяти (24-разрядный *Z*-буфер при разрешении 640×480 требует около 900 кбайт). Эта память должна быть внутренней памятью 3D-ускорителя. Разрядность *Z*-буфера (число битов на координату z) — один из важнейших показателей; чем выше разрядность, тем выше дискретность координат *z* и точнее выполняется прорисовка удалённых объектов.

Ускорители обычно имеют возможность воспроизведения некоторых эффектов, таких, например, как затуманивание, когда гоночный автомобиль поднял песчаную пыль. Изображаемые пиксели смешиваются с цветом тумана в зависимости от его глубины. С помощью этого же алгоритма объекты погружаются в дымку, что позволяет создать иллюзию расстояния.

Прозрачность ещё один эффект, который обеспечивает 3D-ускоритель. В окружающем мире наряду с непрозрачными объектами существуют прозрачные и полупрозрачные. Эффект полупрозрачности создаётся путём объединения цвета точки как комбинации цветов переднего (полупрозрачного) и заднего планов. Обычно коэффициент прозрачности (alpha) имеет

значение от 0 до 1. При этом цвет вычисляется по следующей формуле:

```
цвет пикселя = (alpha) (цвет пикселя переднего плана) + (1 – alpha) (цвет пикселя заднего плана).
```

(Пиксель заднего плана — тот, что находится в *Z*-буфере.)

При формировании каждого следующего кадра 3D-ускоритель проходит весь путь построения заново, поэтому он должен обладать высоким быстродействием. Чтобы придать движению плавность, применяют двойную буферизацию: один буфер — для отображения картинки, другой же — для построения новой картинки. Пока на экране отображается содержимое одного буфера, в другом идёт подготовительная работа. Когда очередной кадр построен, буферы меняются местами. Таким образом, на экране всё время наблюдается отличная картинка.

2D-ускорители уже с 90-х гг. применяются в видеокартах персональных компьютеров. Они обеспечивают прорисовку курсора мыши «поверх экрана» и роллирование (сдвиг) изображения в прямоугольном окне экрана.

ЦИФРОВАЯ ФОТОГРАФИЯ

Ни для кого не секрет, что обработка изображений на компьютере сегодня дело привычное, если не сказать повседневное. Графика для сайтов, фотомонтаж, вёрстка журналов и газет — всё это и многое другое выполняется на ЭВМ.

В настоящее время существует немало форматов для хранения изображений на компьютере. Самыми распространёнными и общеупотребительными являются, пожалуй, JPEG.

Если картинка, находящаяся в таком файле, имеет большой размер, например около тысячи точек по вертикали и горизонтали, то информация, описывающая такую картинку, будет занимать довольно много места на носителе данных. Для уменьшения размеров файлов применяют сжатие, причём оно может быть как без потерь, так и с потерями информации.

Пример возможностей цифрового фото.

Программы сжатия производят анализ исходного изображения и отбрасывают «ненужные» точки. Чем грубее этот анализ, тем быстрее и сильнее сжимается изображение; при сильном сжатии искажение изображения может стать заметным. Современные методы компрессии (сжатия) изображений позволяют достичь очень хороших результатов при минимальных потерях данных.

Перевод изображения в цифровой формат осуществляется с помощью специальных устройств — *сканеров*. Основной частью сканера является светочувствительный элемент. Он улавливает световой поток, отражающийся от исходного бумажного изображения, и генерирует электрический сигнал, который преобразуется в поток данных, описывающий исходную картинку. Такой поток уже может быть обработан на компьютере.

Поскольку обычного дневного света, как правило, недостаточно для светочувствительного элемента, сканер дополнительно оснащается мощной лампой, освещающей сканируемый участок.

Светочувствительный элемент — вещь достаточно дорогая, поэтому он делается небольшого размера и устанавливается на движущееся шасси, которое шаг за шагом перемещается вдоль изображения, и на ЭВМ передаются данные о частях картинки. Управляющая программа на компьютере «склеивает» отдельные кусочки картинки и позволяет наблюдать на экране конечный результат.

Другим устройством создания цифровых изображений является цифровая фотокамера. Принцип её работы очень напоминает принцип работы обычной плёночной камеры, только вместо плёнки используется светочувствительная матрица, улавливающая световой поток. Чем больше точек может обработать матрица, тем лучше качество картинки. Цифровые камеры оснащены собственными но-

сителями данных, на которые записываются файлы в каком-либо распространённом формате, готовые к переносу и редактированию на компьютере. Носителями могут быть дискеты, компакт-диски, жёсткие диски и энергонезависимая память. Около 35 фотографий размером 1280×1024 точки, что примерно соответствует стандартному фото 9×12, занимают порядка 8 Мбайт флэш-памяти. То есть примерно 3 тыс. подобных фотографий можно разместить на одном компакт-диске.

В последнее время в рамках всё повышающегося спроса на цифровую фототехнику производители стараются максимально увеличить размер растра снимка, якобы повышая его качество, но и увеличивая размер хранимого файла до 5—6 мегапикселей и более в бытовых фотоаппаратах. Но размер матрицы решает далеко не все. Так, выпущенный в 1998 г. фотоаппарат Olympus D620 (C1400L) имел матрицу 1,4 мегапикселя, однако кадры, сделанные на нём, подчас превосходят по качеству снимки современных фотоаппаратов с большими матрицами, не говоря уже о фотографиях, полученных на «мегапиксельных» фотокамерах, встроенных в мобильные телефоны. Качество оптики и плотность расположения элементов в светочувствительной матрице цифровой техники сильно влияют на качество снимков. Успокаивает лишь то, что цифровые «мыльницы» лучше плёночных собратьев.

При использовании цифровых камер не только исчезает процесс проявления фотографии, но и фотографирование становится всё более экономичным процессом. Во-первых, не надо тратить деньги на плёнку и проявку; во-вторых, при цифровой фотографии не стоит бояться сделать плохой снимок, так как его можно сразу увидеть и стереть, освободив цифровую память для новых. Поэтому всё больше родителей дарят своим детям цифровые фотоаппараты, чтобы те не разорили их, пока осваивают новую технику.

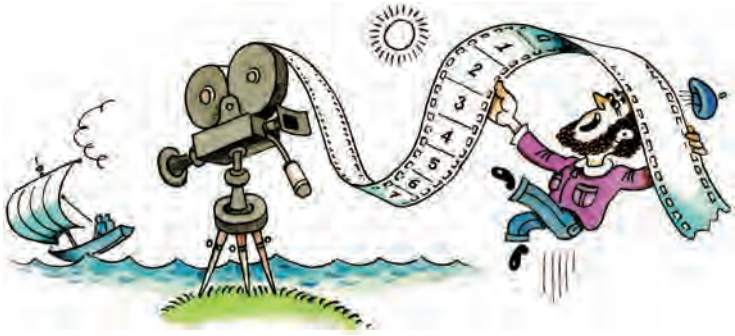
Цифровая макросъёмка.

Для обработки цифровых изображений существует множество программ. Они открывают широкий простор для творчества — от простого изменения размера картинки до наложения на неё фантастических эффектов.

Другим достоинством цифровой фотографии является то, что снимки не портятся, не стареют, на них не остаётся царапин; в считанные минуты фотографии можно отправить в любую точку мира с помощью электронной почты и конечно же разместить на домашней странице в Интернете.

ЦИФРОВОЕ ВИДЕО

На компьютере можно не только играть, решать сложные задачи или использовать его в качестве пишущей машинки, но и смотреть кино, причём порой с качеством, значительно превосходящим изображение на обычном видеомагнитофоне или в телевизоре.



Как известно, видеоизображение — не что иное, как последовательность кадров. Каждая секунда фильма — это набор из 24 отдельных кадров. Таким образом, чтобы сохранить секунду видео, нам необходимо сохранить все 24 картинки.

Пусть размер картинки составляет 512•384 точки. Любая из точек может быть окрашена в один из 16 млн цветов, т. е. для её представления понадобится 3 байт памяти

$$512 \cdot 384 \cdot 3 = 589\,824 \text{ байт.}$$

Итак, для хранения одного кадра нужно около 600 кбайт памяти. А если потребуется сохранить ещё и звук, то размер кадра возрастёт. И тогда для хранения, например, двухчасовой видеозаписи с не очень большим разрешением может понадобиться более 10 Гбайт дискового пространства, что не так уж мало даже для современных носителей информации.

Фото с экрана работы программы для видеомонтажа. Выбор параметров монтажа.

Чтобы хранить видео на цифровых носителях, используют технологии сжатия видеоизображений. Самым распространённым форматом сжатия является формат MPEG. Он разрабатывался группой учёных Motion Picture Engineering Group в конце прошлого века (работы над усовершенствованием данного формата ведутся и сегодня). Сейчас MPEG позволяет сжимать изображение с коэффициентом от 1:10 до 1:100 и более, в зависимости от сложности картинки.

MPEG состоит из трёх частей: Audio, Video, System (последняя выполняет объединение и синхронизацию двух других). Существует несколько разновидностей стандарта MPEG. Наиболее известные из них — MPEG-1, MPEG-2 и MPEG-4. Они отличаются друг от друга степенью сжатия, а также качеством изображения на выходе.

Из нескольких кадров в MPEG складывается один полноценный кадр, остальные же являются кадрами, сформированными искусственно на основе алгоритма предсказания движения. Фактически, хранится только часть информации — за счёт высокой степени сжатия. Чем реже вставляется ключевой кадр, тем хуже качество изображения. Поэтому необходим компромисс между размером файла с фильмом и его качеством.

Звук в MPEG-файле кодируется следующим образом: убираются все звуки, неразличимые для человеческого уха. За счёт такой обработки звукового потока уменьшается размер звуковой дорожки к фильму. При этом некоторые виды MPEG поддерживают несколько звуковых дорожек, т. е., например, во время просмотра можно переключаться с английской версии фильма на русскую и наоборот. Это очень активно используется в современных устройствах DVD (*англ.* Digital Versatile Disk — «цифровой диск различного назначения»).

САМ СЕБЕ РЕЖИССЁР

Конечно, хорошо использовать компьютер для просмотра фильмов, но

ещё интереснее с его помощью создавать собственное кино. Для решения этой задачи потребуются некоторое дополнительное оборудование и терпение. Исходный материал (то, что снято видеокамерой) необходимо перевести в цифровой вид и сохранить на доступном носителе информации. Перевод можно выполнить, используя специальную плату захвата видеоизображения. Если же компьютер уже имеет плату ТВ-тюнера (можно смотреть телепрограммы на экране монитора), скорее всего, из «железа» ничего больше не понадобится. После подсоединения видеомагнитофона или камеры к компьютеру запускается программа для записи видеопотока на диск.

Размер изображения на экране телевизора составляет 576 строк. Но телевизионная картинка формируется из двух частей по 288 строчек. В первой части содержатся все нечётные строки, а во второй — чётные. На телевизионном экране этого обычно не видно, однако при обработке видео на ЭВМ можно наблюдать эффект решётки, особенно в сценах с быстрой сменой картинки. Для удаления решётки применяют специальные программы-фильтры, которые формируют цельную конечную картинку. Этот процесс называется deinterlacing.

Что касается звука, то его можно записывать вместе с видео, если скорость компьютера позволяет обрабатывать одновременно и видео, и звук (в большинстве случаев это именно так), а можно и отдельно.

Редактирование фильма на компьютере не требует специального дорогого студийного оборудования, весь процесс очень нагляден и соответствует принципу WYSIWYG. Сколько фильм будет занимать места на диске, зависит от выбранной программы кодирования. Придётся искать компромисс между качеством и размером (чем лучше качество, тем больше места занимают файлы). Для хранения видео существует много разных форматов, которые улучшают качество и уменьшают размер файлов. Скорее всего, в недалёком будущем появится возможность смотреть видео идеального качества через Интернет.

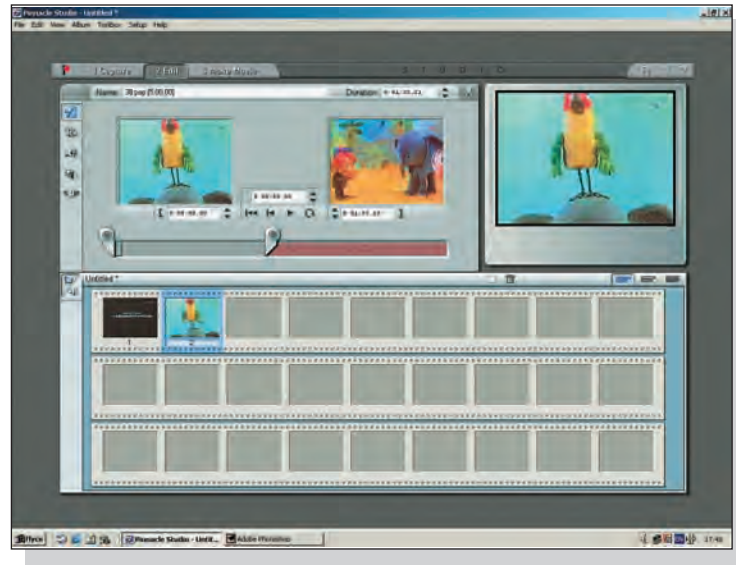
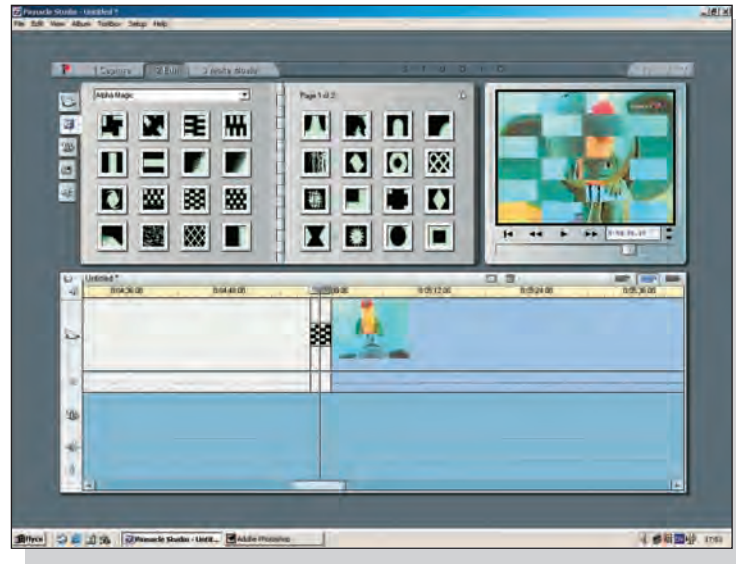


Фото с экрана работы программы для видеомонтажа.
Вверху — монтаж,
внизу — выбор спецэффектов.



Frame (англ.) — кадр, неделимая часть видеопотока.

Frameskip (англ.) — выброс кадра, выпадение. Случается, когда компьютер не успевает обработать кадр, например, если не хватает быстродействия или кадр очень сложный (быстрая смена сцен).

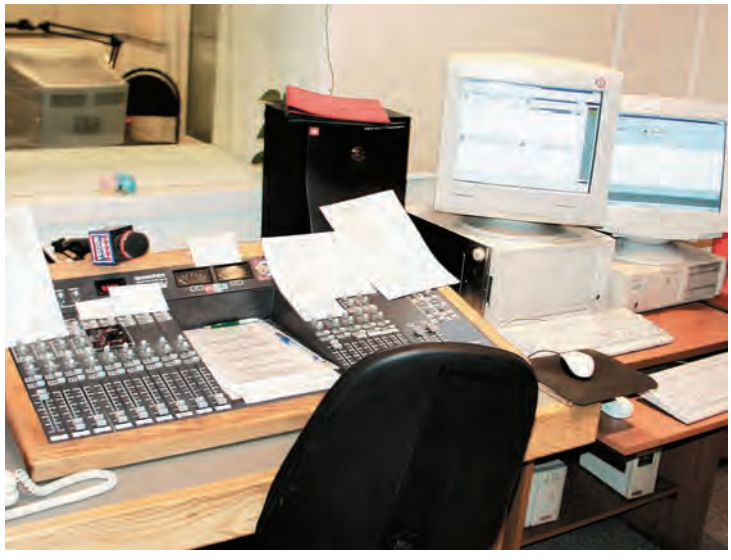
Codec (англ.) — программа для кодирования и раскодирования изображения. Небольшая программа, устанавливаемая на компьютер, для того чтобы другие программы могли «читать и писать» видео в определённом формате.

АУДИОМОНТАЖ

С изобретением в конце XIX в. радио родилась и новая отрасль масс-медиа — радиовещание. Революционным шагом здесь стало использование технологии магнитной записи при подготовке радиопрограмм. Прямой эфир предполагал обязательное присутствие гостей в студии во время передачи. Но гораздо удобнее было участников передачи записать в студии на магнитофон заранее, а уже потом эту запись воспроизводить.

Традиционный монтажный пульт конца XX в. в московском Доме радио (вверху).

Современный монтажный пульт на радиостанции (внизу).



Передача «по просьбе радиослушателей» могла неоднократно повторяться. Но самое главное, её специально «готовили» к эфиру, проводя *монтаж*. Из сплошной записи вырезали куски не столько по цензурным соображениям, сколько из-за оговорок, пауз, повторов одних и тех же слов.

Технология cut and paste (*англ.* «вырежи и вклей») точно отражала положение дел при аудиомонтаже. Аналоговый ленточный магнитофон, клей и ножницы были основными инструментами монтажной студии. Процесс монтажа являлся поистине «творческой» операцией.

На слух вылавливали оговорку, точно работая ножницами, разрезали ленту в начале и конце куска, выбрасывали ненужную часть, аккуратно склеивали. Главное условие — не испортить, не помять, не растянуть магнитную ленту. Порой монтаж 20-минутной передачи занимал несколько часов.

По окончании монтажа подбиралась и заимствовалась отдельно фоновая музыка, и передача переписывалась с двух магнитофонов на третий (речь и музыка). После этого она считалась готовой к эфиру.

Магнитный материал стоил дорого, поэтому ленты освобождали от старых записей, размагничивали и повторно пускали в производственный процесс. Лента иногда не выдерживала постоянных склеек и могла порваться прямо во время эфира или давала слабозаметные щелчки при следующих записях.

Готовый материал, который планировалось использовать не один раз, как правило, переписывался на другую ленту и складывался в архив. Там были созданы специальные условия хранения: поддерживались определённые температура и влажность, чтобы ленты не пересыхали и не портились. За ними регулярно ухаживали, их периодически перематывали, чтобы снять напряжение и не допустить слипания. Записи проверялись и даже переписывались, когда лента ста-

рела. Делалось всё, чтобы записи жили долго. К лентам, хранящимся в архиве, относились так же, как военные к своему оружию на складе.

Такой способ хранения очень дорог. Из-за любого сбоя в процессе хранения записи погибали, так, из-за недостатка финансирования в конце 90-х г. прошлого века (время распада СССР) огромное количество уникальных записей из архива радио было потеряно навсегда.

Но у аналогового метода записи существует очень большой недостаток. При перезаписи, даже на студийной аппаратуре, накладывались посторонние шумы от магнитофона, искажения от усилителя, через которые проходил аналоговый сигнал, шумела и магнитная лента. То есть с каждой перезаписью качество заметно ухудшалось.

В конце XX в. с помощью компьютеров научились хранить, записывать и воспроизводить звук не хуже, чем на профессиональных студийных магнитофонах. Виниловые пластинки и магнитофонные кассеты уступили место в студиях цифровым компакт- и мини-дискам.

Но самые революционные изменения произошли в процессе монтажа. Аналоговому магнитофону, бесконечным лентам, клею и ножницам пришёл на смену персональный компьютер. Технология cut and paste сохранилась, но стала цифровой. Звуковой ряд изображается на экране монитора в виде графика амплитуды звука по времени. Теперь оговорка была видна. Временная шкала легко масштабируется, фрагмент удаётся отметить с точностью до тысячных долей секунды (для ленты это были бы доли миллиметра). Запись сначала прослушивают, как бы с удалённым куском, и лишь потом принимают решение — вырезать или сдвинуть границы. Можно не только семь раз отмерить, а потом отрезать, но и вернуть всё назад, если кусок удалён ошибочно. Нужный кусок легко переносится в любое место фонограммы. Программы монтажа позволяют накладывать фоновую музыку на передачу. В общем, всё, на что прежде требовались часы физического труда, занимает теперь минуты.

Мини-диск (MD) разработан фирмой Sony (Япония). Это цифровой магнитооптический носитель, вмещающий столько же информации, сколько и обычный CD (74 мин). В отличие от CD на него можно неоднократно записывать и удалять с него произвольные куски записи. MD использует при записи сжатие звука в 5 раз методами, при которых потеря качества незаметна.

Из самых простых эффектов, применяемых при обработке передачи, можно назвать затухание и нарастание уровня звука, эффект эха, различных залов и стадионов.

Программы аудиомонтажа обладают поистине безграничными возможностями. Смонтированная передача поступает на компьютер звукорежиссёра, откуда потом и идёт в эфир. Хранение записей также стало более удобным. Данные на мини-дисках и компьютерах не теряют качества со временем. Да и 30-минутная передача требует всего 50 Мбайт дискового пространства, т. е. её вечное хранение практически ничего не стоит.

Что же осталось от старой технологии? Творческий процесс, когда интеллект человека при монтаже из совершенно невзрачной записи делает оригинальную, интересную, запоминающуюся передачу, как литературный редактор сухой научный текст переводит на язык детской энциклопедии. Ведь недаром на радио ходит шутка: «Вы нам только алфавит наговорите, а мы уж сами передачу смонтируем».

Клей впоследствии заменили клеящей лентой, очень похожей на скотч.

В радиовещании звук записывается со сжатием в стандарте MPEG-1 Layer-3 потоком 256 кбит/с.

Студийные магнитофоны.

ВЫЧИСЛЕНИЯ

ВЫЧИСЛЕНИЯ НА КОМПЬЮТЕРЕ

Информация о температуре воздуха, давлении, влажности, облачности, количестве выпавших осадков, силе и направлении ветра в достаточно большом количестве точек Земли даёт возможность составить довольно точный прогноз погоды. Беда в том, что готов он будет, когда уже пройдёт тот период, на который прогноз составлялся, т. е. погоду на лето предскажут следующей осенью.

Ключевой момент в решении этой проблемы — повышение производительности компьютеров, позволяющее увеличить точность и быстроту выполнения задач, а также дающее возможность решать новые классы задач.

Вычисления в компьютере происходят в так называемом арифметическом устройстве (АУ). АУ преобразует информацию; в остальных устройствах компьютера (запоминающем, управляющем, входном, выходном) происходит лишь перемещение информации во времени и пространстве. АУ бывает двух видов — универсальное и специализированное. Универсальное АУ выполняет все арифметические операции, предусмотренные системой команд компьютера, которая обычно включает операции сложения, вычитания, сравнения, умножения, деления и иногда вычисления квадратного корня.

АЛУ (арифметико-логическое устройство) — это простейший вид универсального арифметического устройства. Оно обрабатывает целые числа, выполняет операции сложения, вычитания, сдвига и логические операции.

Специализированное АУ осуществляет группу близких по алгоритму операций. Так, устройство сложения выполняет операции сложения, вычитания, сравнения; устройство умножения — операции умножения; устройство деления — деление и вычисление квадратного корня.

Универсальное АУ состоит из меньшего объёма оборудования (т. е. занимает меньше места на полупроводниковом кристалле, на котором изготовляется современный процессор), чем несколько специализированных АУ, имеющихся в компьютере. Наличие в машине специализированных АУ позволяет осуществлять несколько арифметических операций одновременно, что ускоряет и увеличивает производительность компьютера.

Чем больше оборудования в кристалле, тем сложнее АУ и быстрее их работа, повышающая производительность микропроцессора. Кроме того, можно увеличивать количество арифметических устройств в компьютере, распараллеливая тем самым процесс вычислений. Если раньше некоторые операции (например, деление и вычисление квадратного корня) выполнялись с помощью специальных программ, то сейчас это обычно делается в аппаратуре, что гораздо быстрее программной реализации.

Томограф.

Ещё в 1965 г. Гордон Мур (один из основателей фирмы Intel) заметил, что через каждые несколько лет выпускается новый кристалл полупроводниковой памяти, который имеет объём примерно в два раза больше, чем у предыдущего кристалла памяти. Подобные оценки справедливы и для количества транзисторов в кристалле микропроцессора, и для производительности микропроцессора.

В 90-х гг. появились мультимедийные арифметические устройства, их использование значительно расширило возможности обработки графической информации, возникли компьютерная томография и виртуальная реальность.

Гордон Мур.

ПРЕДСТАВЛЕНИЕ ЧИСЕЛ С ФИКСИРОВАННОЙ ЗАПЯТОЙ

Числа, обрабатываемые в компьютере, делятся на *числа с фиксированной запятой* и *числа с плавающей запятой*. Числа с фиксированной запятой — это обычные целые числа, представленные в двоичном виде. Запятая, отделяющая целую часть числа от дробной, стоит в них на фиксированном, т. е. постоянном, месте. Обычно это место справа от младшего разряда числа. В языках программирования

такие величины обозначаются как целые. Чаще всего целое число занимает 32 или 64 двоичных разряда (бита), при этом один разряд занимает знак числа, а остальные — само число. Минимальное число, которое можно представить с помощью 64-разрядного целого числа, — -2^{63}, а максимальное — $+2^{63} - 1$.

Операции сложения, вычитания и умножения с целыми числами являются точными. Деление и вычисление квадратного корня в результате округления могут дать приблизительный (неточный) результат.

Допустим, надо сложить два одноразрядных числа *A* и *B*. Если сложить 1 + 1, то сумма, которая также хранится в двоичном разряде, станет равна 0 (хотя ожидаемый результат 2). Для того чтобы не потерялся итог, используют не только результат, но и *признак переполнения* в двоичном числе — *перенос*. Перенос также двоичный разряд и равен 0, если переполнение не возникало, или 1, если произошло переполнение. Если обозначить сумму буквой *S*, а перенос буквой *C*, то можно составить следующую таблицу истинности:

A	*B*	*C*	*S*
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

Из неё видно, что сумма *S* равна 1, если одно и только одно слагаемое равно 1. Логический элемент, реализующий таблицу, называется *одноразрядным сумматором*. С помощью этого элемента можно складывать два одноразрядных числа *A* и *B*, получить сумму и перенос, но нельзя складывать числа, имеющие большее количество разрядов. Для этого необходим вход переноса *Cin*, который должен учитывать перенос из соседнего младшего разряда.

Для *полного одноразрядного сумматора* можно написать следующую таблицу истинности:

A	*B*	*Cin*	*C*	*S*
0	0	0	0	0
0	1	0	0	1
1	0	0	0	1
1	1	0	1	0
0	0	1	0	1
0	1	1	1	0
1	0	1	1	0
1	1	1	1	1

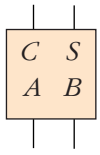
Первые четыре строки этой таблицы (*Cin* = 0) совпадают с предыдущей таблицей истинности. Перенос *C* есть, если складываются две или три единицы; сумма *S* = 1, если складывается нечётное число единиц (одна или три).

Пусть *A* и *B* — это два 4-разрядных двоичных числа, 0 — номер младшего разряда, 3 — номер старшего разряда. Здесь перенос C_0, рождающийся в младшем разряде, подаётся на вход переноса первого разряда; перенос C_1 поступает во второй разряд; перенос C_2 подаётся в старший разряд, т. е. выходы и входы переносов соединены последовательно.

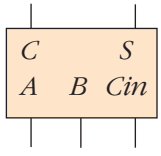
Сумматор с такой организацией цепи переносов называют *последовательным*. Это самый простой тип сумматора, но и самый медленный. В 64-разрядном сумматоре, построенном по такой схеме, старший разряд суммы получается через 64 цикла одноразрядного сумматора!

Сумматор является узлом, во многом определяющим быстродействие компьютера, поэтому повышение быстродействия сумматоров — это всегда актуальная задача. Сейчас созданы сумматоры, которые в десятки раз быстрее последовательного сумматора. Сумматоры используются не только при сложении чисел, но и при умножении, делении и даже вычислении квадратного корня.

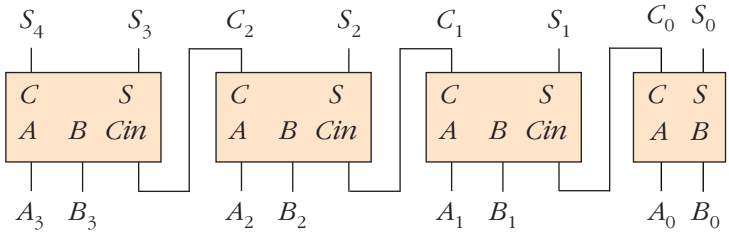
Одноразрядный сумматор.



Полный одноразрядный сумматор.



Последовательный сумматор.



Для вычитания числа с фиксированной запятой вычитаемое преобразуется в дополнительный код (число, в сумме с исходным дающее 0), а затем прибавляется к уменьшаемому.

Умножение можно выполнять, как обычное умножение, в столбик. Пусть E и T — множители, а P — произведение. Сначала первый множитель E умножается на младший разряд множителя T_0. Если $T_0 = 0$, то $E \cdot T_0 = 0$, иначе $E \cdot T_0 = E$, т. е. произведение первого множителя на разряд второго множителя равно либо нулю, либо первому множителю. Это произведение называют частичным произведением. Затем первый множитель сдвигается на один разряд влево и умножается на разряд второго множителя T_1 — получается следующее частичное произведение. Такой процесс идёт до старшего разряда второго множителя T_{63}. Для получения произведения надо сложить частичные произведения, а это можно сделать с помощью уже знакомых сумматоров.

Данный алгоритм умножения — самый простой и довольно медленный. Существует огромное количество алгоритмов, позволяющих ускорить процесс умножения.

Деление — это самая сложная и самая медленная арифметическая операция. Существуют алгоритмы деления, в которых за один шаг определяется сразу несколько цифр частного.

Сейчас широко применяются алгоритмы, где деление заменяется умножением на величину, обратную делителю. Часто обратная величина определяется с помощью таблицы. В некоторых процессорах вообще нет устройства деления, и эта операция выполняется специальной программой.

Для научных вычислений характерно использование очень больших или очень маленьких чисел. Например, расстояния между космическими объектами и их массы очень велики, а расстояния внутри атома и массы элементарных частиц очень малы. В этом случае говорят о широком диапазоне используемых чисел. Числа же с фиксированной запятой имеют узкий диапазон.

Первые компьютеры использовали представление чисел с фиксированной запятой. В то время для решения проблемы узкого диапазона программисты применяли масштабирование: если числа, с которыми должна работать программа, были очень малы, программист умножал их на одно и то же число, чтобы они попали в диапазон чисел компьютера. Если же данные для программы, наоборот, были очень велики, то делили их на одно и то же число.

Для масштабирования программист анализировал данные, промежуточные результаты и конечный результат, чтобы на всех этапах хватило диапазона вычислений. Выход за границы диапазона приводил к грубой ошибке в вычислениях. Например, если число в компьютере имеет 10 разрядов, а перемножаются два 6-разрядных числа, то произведение будет иметь 12 разрядов, т. е. 2 разряда произведения не попадают в разрядную сетку компьютера, результат при этом может получиться бессмысленным.

Процесс масштабирования отнимал у программистов слишком много времени и сил.

Алан Клементс с женой.

«Есть по крайней мере одна настоящая причина для изучения умножения и деления — существует бесконечное количество способов выполнения этих операций, и, следовательно, существует бесконечное количество диссертаций, которые можно защитить, если изобрести новый умножитель».

Алан Клементс (американский учёный в области вычислительной техники). «Принципы компьютерного оборудования». 1986 г.

ПРЕДСТАВЛЕНИЕ ЧИСЕЛ С ПЛАВАЮЩЕЙ ЗАПЯТОЙ

Числа с плавающей запятой в компьютере представляются так:

$$Ч = (-1)^{Зн} \cdot M \cdot 2^{П},$$

где $Ч$ — величина числа, $Зн$ — знак числа, M — мантисса числа, $П$ — порядок числа. В языках программирования числа с плавающей запятой обычно обозначаются как вещественные. Знак числа занимает один бит. Если $Зн = 0$, то $(-1)^{Зн} = (-1)^0 = +1$ и $Ч$ имеет положительный знак. Если $Зн = 1$, то $(-1)^{Зн} = -1$ и $Ч$ имеет отрицательный знак.

На величину M накладывается ограничение:

$$1 \leqslant M \leqslant 2.$$

Мантисса, удовлетворяющая этому неравенству, называется *нормализованной*. $П$ — это целое число со знаком.

Такое представление чисел похоже на запись, которая используется учёными и инженерами. Например, инженер вместо $a = 350\,000\,000\,000$ напишет $a = 3{,}5 \cdot 10^{11}$.

Использование чисел с плавающей запятой позволило значительно увеличить диапазон и точность чисел в компьютере и освободить программиста от утомительного процесса масштабирования.

Так, числа с плавающей запятой формата 64 имеют 11 разрядов порядка, это позволяет представлять числа от -2^{2047} до $+2^{2047}$. Если расстояние между двумя соседними числами с фиксированной запятой составляет 1 (ведь два соседних числа отличаются именно на единицу), то расстояние между двумя соседними числами с плавающей запятой, которые находятся вблизи 0, равно 2^{-2100}.

Сложим два числа с плавающей запятой:

$$Ч1 = (-1)^{Зн1} \cdot M_1 \cdot 2^{П1},$$
$$Ч2 = (-1)^{Зн2} \cdot M_2 \cdot 2^{П2}.$$

Чтобы сложить числа, представляющие собой степени, нужно уравнять показатели их степеней. Пусть $П1 > П2$ и $П1 - П2 = РП$ (разность порядков), тогда можно написать следующее:

$$Ч1 + Ч2 = (-1)^{Зн1} \cdot M_1 \cdot 2^{П1} +$$
$$+ (-1)^{Зн2} \cdot M_2 \cdot 2^{П2} =$$
$$= (-1)^{Зн1} \cdot M_1 \cdot 2^{П1} +$$
$$+ (-1)^{Зн2} \cdot (M_2 \cdot 2^{-РП}) \cdot (2^{П2} \cdot 2^{РП}).$$

Умножили и разделили $Ч2$ на $2^{РП}$, величина $Ч2$ от этого не изменилась,

ОШИБКА ОКРУГЛЕНИЯ

Надо заметить, что арифметика с плавающей запятой значительно сложнее арифметики с фиксированной запятой. Например, для чисел с плавающей запятой результат зависит от порядка выполнения операций, т. е.

$$A + (B + C) \neq (A + B) + C.$$

Это связано с тем, что вычисления с плавающей запятой не точные, а приблизительные. При выполнении операций с плавающей запятой отбрасываются младшие разряды, которым не хватает места в разрядной сетке, и производится округление. В ходе решения задачи ошибка округления накапливается и может привести к получению совершенно неправильного результата.

Обычный метод проверки численных алгоритмов — использование повышенной точности. Так, вместо 32-разрядной точности применяют 80-разрядную, но это не всегда даёт нужный результат. Например, вычислим функцию двух переменных:

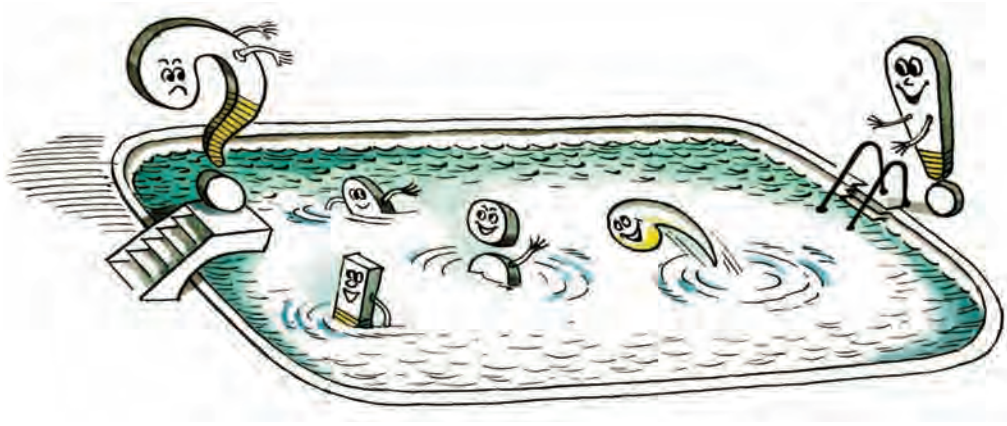
$$f(x, y) = 333{,}75y^6 + x^2(11x^2y^2 - y^6 - 121y^4 - 2) + 5{,}5y^8 + x/(2y)$$

при $x = 77\,617$ и $y = 33\,096$. Получаем следующие результаты с 32-, 64- и 80-разрядной точностью:

32р $f = 1{,}172\,603\ldots$
64р $f = 1{,}172\,603\,940\,053\,1\ldots$
80р $f = 1{,}172\,603\,940\,053\,178\ldots$

Все три результата совпадают по семи старшим десятичным цифрам. Может показаться, что увеличение разрядности действительно повышает точность вычислений. На самом же деле правильный ответ имеет даже другой знак результата!

Интервальная арифметика — это раздел арифметики, помогающий устранять такие особенности вычислений с числами с плавающей запятой. Интервальный метод даёт в ответе не число, а интервал между двумя числами, в котором заключён результат. Если интервал очень большой, это говорит о том, что выбранный алгоритм даёт недостаточную точность.



но изменились составляющие числа *Ч2* — его мантисса и порядок.

Для умножения мантиссы на $2^{РП}$ можно перенести запятую на *РП* разрядов влево либо сдвинуть мантиссу вправо. При этом *РП* младших разрядов мантиссы выйдут за границу разрядной сетки, они не смогут участвовать в вычислениях. Нахождение разности порядков *РП* можно сделать с помощью небольшого сумматора, сдвиг мантиссы вправо производится узлом процессора, который так и называется — сдвигатель.

С другой стороны, $2^{П2} \cdot 2^{РП} = 2^{П2+РП} = 2^{П1}$, т. е. порядок второго числа равен порядку первого, и числа можно складывать. Теперь продолжим преобразование:

$$Ч1 + Ч2 = [(-1)^{Зн1} \cdot M_1 + (-1)^{Зн2} \cdot (M_2 \cdot 2^{-РП})] \cdot 2^{П1}.$$

Осталось сложить M_1 (с учётом знака) и сдвинутую вправо на *РП* разрядов мантиссу M_2, у которой тоже учтён знак. А складывают мантиссы с помощью сумматора. Результат будет иметь порядок, равный большему из двух порядков, — *П1*.

Пусть надо умножить два числа с плавающей запятой:

$$Ч1 = (-1)^{Зн1} \cdot M_1 \cdot 2^{П1},$$

$$Ч2 = (-1)^{Зн2} \cdot M_2 \cdot 2^{П2},$$

$$Ч1 \cdot Ч2 = (-1)^{Зн1} \cdot M_1 \cdot 2^{П1} \cdot (-1)^{Зн1} \times M_2 \cdot 2^{П2} = (-1)^{Зн1} \cdot (-1)^{Зн2} \cdot M_1 \cdot M_2 \times 2^{П1} \cdot 2^{П2} = (-1)^{Зн1+Зн2} \cdot M_1 \cdot M_2 \cdot 2^{П1} \cdot 2^{П2}.$$

Видно, что для умножения чисел с плавающей запятой надо умножить мантиссы, а порядок произведения равняется сумме порядков сомножителей. Знак произведения определяется знаками сомножителей: если знаки одинаковые, то произведение положительно, если знаки разные, то произведение отрицательно.

Пусть надо разделить *Ч1* на *Ч2*:

$$Ч1/Ч2 = [(-1)^{Зн1} \cdot M_1 \cdot 2^{П1}] / [(-1)^{Зн2} \cdot M_2 \times 2^{П2}] = (-1)^{Зн1-Зн2} \cdot (M_1/M_2) \cdot 2^{П1-П2}.$$

Для деления чисел с плавающей запятой надо разделить мантиссы, а от порядка делимого отнять порядок делителя. Знак частного положительный, если знаки делимого и делителя совпадают, и отрицательный – если знаки делимого и делителя различны.

В англоязычных странах вместо плавающей запятой говорят о плавающей точке.

СТАНДАРТ ANSI/IEEE 754

Стоимость разработки программного обеспечения во много раз больше

Группа экспертов IEEE.

ЦЕНА ОШИБОК В ВЫЧИСЛЕНИЯХ

Американская зенитная ракета «Пэтриот».

В 1991 г. во время войны в Персидском заливе (США против Ирака) батарея американских зенитных ракет «Пэтриот» не смогла перехватить запущенную иракцами ракету «Скад» советского производства. Ракета попала в казарму американских солдат, при этом погибло 28 человек.

Причиной ошибки явилась погрешность в вычислениях времени. При вычислении нужно было умножать время, задаваемое тактовым генератором компьютера (оно измерялось в десятых долях секунды), на 1/10, но это десятичное число невозможно точно представить в двоичном виде. Для хранения данной константы использовался 24-разрядный регистр. Разница между точным значением 1/10 и её неточным двоичным представлением составляет в двоичном виде следующую величину: 0,0 000 000 000 000 000 000 011 001 100... или около 0,000 000 095 в десятичном виде. Компьютер работал около 100 ч, за это время накопилась ошибка в измерении времени, составившая 0,34 с. Скорость ракеты «Скад» составляла примерно 1700 м/с, т. е. за данное время она прошла более 500 м. Этого хватило для того, чтобы зенитные ракеты не смогли её перехватить.

А вот пример другой ошибки, не столь трагической, но очень дорогостоящей. В 1996 г. происходил первый пуск ракеты «Ариан-5» Европейского космического агентства. Через 40 с после старта она отклонилась от заданной траектории и взорвалась. Ракета создавалась в течение 10 лет, её разработка обошлась в 7,5 млрд долларов.

Это случилось из-за неточности в программном обеспечении инерциальной системы отсчёта. Горизонтальная скорость ракеты относительно пусковой площадки выражалась 64-разрядным числом с плавающей запятой. В программе оно преобразовалось в 16-разрядное число с фиксированной запятой. Оказалось, что это число превысило 32 768 (максимальное целое число со знаком, которое можно представить в 16 разрядах), преобразование дало ошибку, и ракета взорвалась.

Война в Персидском заливе. Беженцы. 1991 г.

стоимости разработки аппаратуры или, как говорят специалисты, «железа». Кроме того, написание программного обеспечения занимает огромное количество времени. Поэтому новую аппаратуру стараются проектировать под уже существующие программы. Программное обеспечение создаётся на языках программирования высокого уровня, чтобы его можно было с одинаковым результатом использовать для компьютеров разных типов (это называется *переносимостью программного обеспечения*).

Операции с числами с плавающей запятой мешали достижению переносимости, так как в прошлом их реализация сильно изменялась от одного семейства компьютеров к другому. Реализации отличались по количеству разрядов порядка и мантиссы, способу их кодирования, основанию системы счисления, способу округления, прерываниям и т. д.

Такое положение в области арифметики с плавающей запятой явилось причиной того, что в 1977 г. при Компьютерном обществе Института инженеров по электротехнике и радиоэлектронике (IEEE Computer Society) США была создана группа по разработке стандарта на двоичную арифметику с плавающей запятой. Членами этой группы стали представители фирм — производителей микропроцессоров и университетские учёные. Заседания группы проходили очень бурно, поэтому одно время говорили, что в США три достопримечательности, которые надо обязательно посетить, — Лас-Вегас, Большой каньон и заседания группы по разработке стандарта на двоичную арифметику с плавающей запятой.

Хотя стандарт был принят в 1985 г., уже к 1984 г. его проект был реализован ведущими фирмами (Intel, AMD, IBM и др.).

За историю своего существования стандарт стал фактически международным, и теперь практически все разработчики аппаратуры и программного обеспечения следуют его требованиям. Тем не менее в 2001 г. началась работа по пересмотру стандарта; это вызвано тем, что компьютерная арифметика за последние 20 лет добилась достижений, которые не отражены в стандарте.

МЕТОДЫ ВЫЧИСЛЕНИЙ

Расчёты требуется проводить во многих областях, и одну из ключевых ролей в этом играет вычислительная математика. Естественные науки, изучая то или иное явление, обычно дают его описание, формулируют законы. Описание носит либо качественный, либо количественный характер. Если требуется спрогнозировать какое-то явление или процесс в виде числа (например, предсказать температуру воздуха, атмосферное давление и т. п.), то необходимо количественное описание, или *математическая модель*.

ТОЧНО ИЛИ ПРИБЛИЖЁННО...

Леопольд Кронекер однажды сказал: «Целые числа создал Господь Бог, всё остальное — дело рук человеческих». С целыми числами работать просто и приятно. Вычисления с ними теоретически можно осуществлять с абсолютной точностью. Совсем другое дело — числа *вещественные* (математики предпочитают термин *«действительные»*), более знакомые нам как десятичные дроби. Последние бывают как конечные, так и бесконечные — периодические и непериодические. Например, из школьного курса математики известно, что число π не выражается конечной или бесконечной периодической десятичной дробью, т. е. является *иррациональным*.

Именно с действительными числами сталкиваются при решении физических задач. Масса тела, величина заряда, скорость, ускорение, сила тяжести — всё это примеры *физических величин*. Некоторые из них, такие, как масса или заряд, являются *скалярами* и характеризуются одним действительным числом. Другие же, *векторы,* задаются тремя действительными числами — координатами вектора.

Даже теоретически действительные числа не допускают ни точной записи, ни, уж подавно, точных вычислений. Но это ещё не всё. На практике решение физической задачи предполагает задание каких-то исходных данных или как минимум задание значений фундаментальных физических констант. И те и другие получают с помощью *измерений,* т. е. *приближённо*.

> К счастью, нам никогда и не требуется абсолютно точное решение. Допустимая погрешность, т. е. отклонение вычисленного значения искомой величины от её истинного значения, определяется из практических нужд. Например, нам могут быть известны размеры комнаты с точностью до сантиметра, но в обычной жизни нам вряд ли нужно знать её объём с точностью до кубического сантиметра.

ПРЕДСТАВЛЕНИЕ ЧИСЕЛ В МАШИНЕ

Физические величины принимают как очень большие, так и очень маленькие значения. Причём и те и другие могут встретиться в одной задаче. Например, скорость света в вакууме составляет около 300 000 000 м/с, а гравитационная постоянная примерно равна 0,000 000 000 066 7 Нм^2/кг^2. Для удобства работы с такими числами физики давно используют так называемую экспоненциальную форму записи. Вместо 300 000 000 пишут $3 \cdot 10^8$, а вместо 0,000 000 000 066 7 — $6{,}67 \cdot 10^{-11}$. Целое или дробное число, стоящее перед цифрой 10, называется мантиссой, а целое число, находящееся после 10, — порядком или экспонентой. Вся запись означает, что число равно мантиссе, умноженной на основание (в данном случае 10) в соответствующей степени.

Тот же принцип применяется для представления вещественных чисел в современных компьютерах, только вместо десятичной используется двоичная система счисления. Для хранения мантиссы и порядка отведено определённое для данного процессора количество разрядов. Позиция запятой, отделяющей целую

Леопольд Кронекер.

часть мантиссы от дробной, считается зафиксированной раз и навсегда (обычно перед первой двоичной цифрой мантиссы). Как и раньше, вся запись означает, что мантиссу нужно умножить на основание, возведённое в степень порядка. Только основание теперь равно 2.

Запись вещественного числа в экспоненциальной форме можно понимать ещё и так: порядок задаёт количество разрядов, на которое нужно передвинуть запятую в записи мантиссы, чтобы получить искомое число (влево или вправо, в зависимости от знака порядка). Поэтому такую форму представления чисел называют *форматом с плавающей запятой*, а сами вещественные числа — *числами с плавающей запятой*.

Количество разрядов в мантиссе определяет точность в представлении числа, а количество разрядов экспоненты — диапазон представления чисел. Многие компьютеры поддерживают два формата чисел с плавающей запятой — короткий и длинный. Например, короткий формат занимает 32 разряда, из них 8 разрядов отводится на порядок, 23 — на мантиссу и 1 разряд — на знак числа. Можно считать, что старший разряд мантиссы всегда равен 1. Тогда его хранить не надо, точность представления числа составит 24 двоичных разряда или примерно 7 десятичных. Порядок длиной 8 разрядов позволит представлять числа в диапазоне 2^{-128} — 2^{128} или примерно от 10^{-39} до 10^{39}.

В числах длинного формата (64 разряда) для мантиссы отведём, скажем, 52 разряда, а для порядка — 11. Это разрешило бы представлять числа с точностью до 16 десятичных разрядов в диапазоне примерно от 10^{-308} до 10^{308}.

Теперь понятно, что вещественные числа, которые могут быть представлены в машине, покрывают числовую ось своеобразной сеткой, причём чем ближе к нулю, тем меньше становятся её ячейки. Это значит, что для каждого машинного вещественного числа a существуют его ближайшие соседи a_0, a_1 ($a_0 < a < a_1$), а между a и его соседями никаких чисел нет.

Вообще представление вещественных чисел в данной машине удобно характеризовать двумя константами. Первая — *машинное* ε — это самое большое число, которое, будучи прибавлено к единице, не меняет её. Такое число равно $1/2^l$, где l — количество разрядов в мантиссе. Или $1/2^{24} \approx 10^{-7}$ для короткого формата и $1/2^{53} \approx 10^{-16}$ — для длинного. Вторая константа — самое маленькое ненулевое число, представимое в машине, или *машинное* θ, равно соответственно $1/2^{128} \approx 10^{-39}$ и $1/2^{1024} \approx 10^{-308}$. Зная эти константы, можно теоретически оценить погрешности при вычислениях.

РОЛЬ ОШИБОК ОКРУГЛЕНИЯ

При вычислении по формулам может катастрофически падать точность. Поэтому часто приходится искать обходной путь для получения правильной вычислительной схемы. В некоторых простых случаях решение физической задачи сводится к вычислению по формуле. Но даже здесь порой таятся подводные камни.

Пусть при решении можно воспользоваться разложением функции e^x в ряд Тейлора в окрестности точки $x = 0$:

$$e^x = \sum_{k=0}^{\infty} x^k/k!$$

Справедливости ради заметим, что существовали компьютеры с основанием в представлении действительных чисел, равным 8 и даже 16 — в любом случае степени двойки.

Этот ряд довольно быстро сходится, особенно при небольших x. На практике процесс вычислений обрывают, когда прибавление очередного слагаемого перестаёт изменять сумму.

При больших по абсолютной величине отрицательных x возникает проблема: значения слагаемых сначала очень быстро возрастают и лишь потом начинают уменьшаться. Некоторые члены ряда могут значительно превзойти ожидаемый результат.

Поскольку погрешность при вычислении каждого слагаемого имеет порядок самого слагаемого, умноженного на машинное ε, суммарная ошибка оказывается слишком большой. Чтобы обойти эту сложность, лучше воспользоваться свойствами функции e^x:

$$e^{-x} = 1/e^x.$$

Для больших положительных x сходимость ряда можно улучшить (а тем самым ускорить вычисления), применив тождество:

$$e^{N+x} = e^N \cdot e^x$$

или

$$e^{N \cdot x} = (e^x)^N,$$

где N — целое число.

НАХОЖДЕНИЕ КОРНЕЙ УРАВНЕНИЯ

Многие задачи сводятся к решению уравнения вида

$$f(x) = 0,$$

где $f(x)$ — некая функция, значения которой можно вычислить. Школьный пример — нахождение значения $\sqrt{a}$, решая уравнение

$$x^2 - a = 0.$$

Здесь $f(x) = x^2 - a$, где a — некоторое действительное число.

ИНТЕРВАЛЬНАЯ АРИФМЕТИКА

Один из подходов к оценке погрешности вычислений основан на использовании интервальной арифметики. Идея состоит в том, что приближённое значение вещественного числа заменяется парой чисел. Эта пара чисел задаёт интервал, внутри которого заведомо находится точное значение числа. Арифметические действия над числами заменяются на действия с интервалами. Это позволяет на конкретных данных увидеть, как растёт погрешность, и убедиться в необходимой точности вычислений.

Поскольку в компьютере нет встроенных команд интервальной арифметики, для реализации метода приходится использовать специальные подпрограммы. Ясно, что это значительно увеличивает время счёта. Поэтому описанный подход не применяется при проведении серийных вычислений. Но он бывает полезен при анализе роли погрешностей вычислений на отдельных примерах, а также при вычислении значений математических констант, когда точная оценка погрешности существенна для доказательства теорем.

Если известен отрезок, на концах которого функция $f(x)$ принимает разные знаки, и нас интересует один корень, причём любой, то для решения воспользуемся методом деления отрезка пополам. В нашем случае, если $a > 1$, это отрезок от 0 до a. Корень уравнения заведомо находится внутри отрезка. Вычислим значение функции $f(x)$ в середине отрезка и проверим знак. Выберем в качестве нового отрезка ту половину первоначального отрезка, на концах которой функция имеет разные знаки. Новый отрезок стал в два раза короче старого. Будем продолжать данный процесс до тех пор, пока длина отрезка не станет меньше допустимой погрешности.


```
алг Вычисление корня
нач вещ A,B,e,C
| e:=0.01
| вывод нс, «Введите A=»; ввод A
| вывод нс, «Введите B=»; ввод B
| нц пока B-A>2*e
| | C:=(A+B)/2
| | если f(A)*f(C)<=0
| | | то B:=C
| | | иначе A:=C
| | всё
| кц
| C:=(A+B)/2
| вывод нс, «Корень=», C,
|                «точность=», e
кон
алг вещ f(вещ X)
нач
| знач:=X**2-2
кон
```

Если длина отрезка в начале процесса была L, а требуется получить ответ с точностью ε, то надо будет сделать $\log_2(L/\varepsilon)$ шагов.

Метод деления пополам позволяет решить задачу отыскания корня одного уравнения, но он с трудом переносится на случай решения задачи для системы уравнений. Этого недостатка лишены *итерационные методы*.

Сначала задачу сводят к уравнению

$$G(x) = x.$$

Затем каким-либо способом определяют начальное приближение к корню x_0. Наконец, приближение уточняют по схеме $x_{k+1} = G(x_k)$. Если функция $G(x)$ и начальное приближение x_0 выбраны правильно, то этот процесс сходится, и мы в конце концов получим решение с желаемой точностью. Условие сходимости формулируется следующим образом: искомое решение и начальное приближение должны лежать на отрезке, на котором для двух любых точек x_1, x_2 справедливо неравенство

$$|G(x_2) - G(x_1)| < q|x_2 - x_1|,$$

где q — положительное число меньшее единицы, а $|x|$ — модуль вектора x.

Условие истинно, если G имеет модуль производной меньше единицы на отрезке от начального приближения до корня. Здесь в общем случае x уже не одно число, а набор чисел, например вектор. Соответственно $G(x)$ — это функция, преобразующая вектор в вектор.

Итерационным является также широко известный метод Зейделя для решения систем линейных уравнений. В отличие от точных методов, например метода Гаусса, метод Зейделя является приближённым: решение задачи проводится путём повторения шагов до тех пор, пока не будет достигнута нужная точность.

Пусть дана система

$$\sum_{j=1}^{n} a_{ij}x_j = b_i,\ i = 1, 2, \dots, n.$$

Задаётся начальное приближение $x_i^0 i$ ($i = 1, 2, \dots, n$). Далее проводятся шаги k по уточнению решения. На каждом шаге вычисляются компоненты нового приближения x_i^k. Формулы для этих вычислений получаются так. Перепишем i-е уравнение в виде

$$\sum_{j=1}^{i-1} a_{ij}x_j^k + a_{ii}x_i^k + \sum_{j=i+1}^{n} a_{ij}x_j^{k-1} = b_i.$$

Тогда получим формулу для вычисления нового (т. е. получаемого на k-м шаге) значения x_i:

$$x_i^k = (b_i - \sum_{j=1}^{i-1} a_{ij}x_j^k + \sum_{j=i+1}^{n} a_{ij}x^{k-1})/a_{ii}.$$

Метод Зейделя сходится не всегда. Закономерен вопрос, зачем нужны ите-

рационные методы для решения задачи, если для этого есть точные методы.

При решении задач часто встречаются особые матрицы — разреженные или матрицы, ненулевые элементы которых легко вычислить.

При преобразовании матрицы методом Гаусса число ненулевых элементов быстро растёт, и приходится хранить всю матрицу.

В методе Зейделя вычисляются только суммы, поэтому в случае разреженной матрицы можно значительно сэкономить память, если хранить не всю матрицу, а лишь ненулевые элементы и их индексы.

РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Приближение элементарных функций, решение уравнений — всё это типично математические задачи, встречающиеся, впрочем, на каждом шагу. Непосредственно же рассмотрение физической задачи приводит, как правило, к дифференциальному уравнению той или иной степени сложности или к системе таких уравнений.

Рассмотрим тело, подвешенное на пружине. Расположим ось X вдоль линии натяжения пружин. Если отклонить тело вдоль оси X, то на него будет действовать сила, стремящаяся вернуть тело в первоначальное состояние. По закону Гука эта сила равна

$$F = -kx.$$

Воспользовавшись вторым законом Ньютона, получаем

$$ma = -kx.$$

Вспомнив, что ускорение есть производная скорости ($a = v'$), а скорость — производная координаты ($v = x'$), получаем систему дифференциальных уравнений

$$v' = -kx/m,$$
$$x' = v.$$

Теперь можно построить простой метод для решения предложенной системы. Если промежуток времени Δt мал, производные $x'(t)$ и $v'(t)$ можно приближённо вычислить по формулам

$$x'(t) \approx (x(t+\Delta t) - x(t))/\Delta t,$$
$$v'(t) \approx (v(t+\Delta t) - v(t))/\Delta t.$$

Разобьём интересующий нас временной интервал от $t = 0$ до $t = T$ на небольшие отрезки длиной Δt: $t_0 = 0$, $t_1 = \Delta t, \ldots, t_i = t_{i-1} + \Delta t, \ldots, t_N = T$. Теперь, задав координату x_0 и скорость v_0 тела в начальный момент времени t_0, можно последовательно вычислить их значения в моменты времени $t_1, \ldots, t_N$, используя формулы

$$x_{i+1} = x_i + v_i \Delta t,$$
$$v_{i+1} = v_i - (kx/m)\Delta t.$$

Таким образом задача отыскания функции на отрезке, т. е. объекта, задаваемого бесконечным количеством чисел, свелась к нахождению таблицы значений функций x и v в отдельных точках этого отрезка. Такой процесс сведения «бесконечной» задачи к «конечной» называется *дискретизацией*. Вопросами же оценки точности полученного решения и её зависимости от N занимается теория приближений.

* * *

Существует ещё одна проблема, актуальная для любого численного метода, — проблема эффективности. Обычно в докомпьютерные времена для оценки сложности работы программы подсчитывали количество умножений. Однако в реальности доля умножений в вычислениях отнюдь не самая большая. Во-первых, в современных компьютерах умножение и сложение вещественных чисел требуют одинакового или почти одинакового времени. Но главное — время доступа к оперативной памяти во много раз больше, чем время выполнения любой арифметической операции. Кроме того, на практике при выборе того или иного способа решения задачи на компьютере важную роль играет возможность приспособить алгоритм к особенностям аппаратуры.

Карл Фридрих Гаусс.

Разреженной является матрица, у которой большинство элементов равно нулю.

МЕТОД НЬЮТОНА

Общий случай	**Пример Ньютона**
Пусть задана функция $y = f(x)$	Ньютон выбирает функцию $y = x^3 - 2x - 5$
Требуется решить приближённо уравнение $f(x) = 0$	Ньютон собирается решить приближённо уравнение $x^3 - 2x - 5 = 0$
Предположим, что нам известно значение X^0, которое мало отличается от корня	Ньютону известно, что значение $x = 2$ мало отличается от корня
Сделаем замену $x = x_0 + p$	Ньютон делает замену $x = 2 + p$
Подставим $x_0 + p$ вместо x в уравнение $f(x) = 0$ и получим новое уравнение $g(p) = 0$	Ньютон получает новое уравнение $p^3 + 6p^2 + 10p - 1 = 0$
Теперь отбросим в уравнении все члены, степень которых выше первой, оставив только свободный член и член первой степени по p: $Bp + A = 0$	Ньютон отбрасывает в уравнении члены $p^3 + 6p^2$ и оставляет уравнение $10p - 1 = 0$
Найдём приближённое значение корня p^0 из уравнения $Bp + A = 0$	Ньютон находит $p = 0{,}1$ из уравнения $10p - 1 = 0$
Возьмём в качестве приближённого значения корня исходного уравнения $x^1 = x_0 + p_0$	Ньютон откладывает этот шаг до конца вычисления

ТОЧНОСТЬ МЕТОДА НЬЮТОНА

Если один шаг метода Ньютона не даёт требуемой точности, то можно сделать ещё один шаг:

$$x_2 = x_1 - f(x_1)/f'(x_1)$$

и т. д. Метод Ньютона замечателен тем, что *если исходное приближённое значение корня достаточно близко к корню, то при каждом шаге Ньютона число верных десятичных знаков ответа будет как минимум удваиваться*. Если попробовать решить методом Ньютона уравнение, в котором заранее известен точный корень: $x^2 - 1 = 0$, то в качестве приближённого значения корня берётся $x_0 = 1{,}1$. Это приближение имеет один верный знак. Можно предположить, что после одного шага Ньютона получится не меньше двух верных знаков, а после двух шагов — не меньше четырёх. Поскольку $(x^2 - 1)' = 2x$, то:

$$x_1 = x_0 - f(x_0)/f'(x_0) = 1{,}1 - 0{,}21/2{,}2 \cdot 1{,}1 - 0{,}095 = 1{,}005,$$

$$x_2 = x_1 - f(x_1)/f'(x_1) = 1{,}005 - (1{,}0052 - 1)/2{,}010 \times 1{,}005 - 0{,}004985 = 1{,}00001.$$

Результат превзошёл ожидания: после одного шага три верных знака, а после двух шагов — пять верных знаков.

Коэффициенты A и B в уравнении для p можно легко вычислить, не делая подстановку $x = x_0 + p$ и не выписывая уравнение для p. Поскольку $f(x_0 + p) = g(p) = \ldots + Bp + A$, то подставив $p = 0$, убедимся, что $A = f(x_0)$. Аналогично, продифференцировав по переменной p обе части тождества $f(x_0 + p) = \ldots + Bp + A$ и подставив $p = 0$, получим $B = f'(x_0)$. Таким образом, вместо уравнения $f(x_0 + p) = 0$ мы решаем уравнение $f'(x_0)p + f(x_0) = 0$. Корень этого уравнения даётся формулой $p_0 = -f(x_0)/f'(x_0)$, следовательно, в качестве нового приближённого значения корня уравнения $f(x) = 0$ можно взять $x_1 = x_0 - f(x_0)/f'(x_0)$.

Переход от приближённого значения корня x_0 к более точному приближённому значению x_1 называется *шагом метода Ньютона*.

Метод Ньютона не работает, если в уравнении $f'(x_0)p + f(x_0) = 0$ коэффициент $f'(x_0)$ равен нулю или близок к нулю. То же случается, когда рядом с приближённым значением x_0 два кор-

ня уравнения. В этом случае Ньютон предлагает оставить в уравнении для p не только свободный и линейный члены, но и квадратичный член:

$$f(x_0 + p) = g(p) = \ldots + Cp^2 + Bp + A$$

(можно доказать, что $C = f''(x_0)/2$).

Отбросив члены порядка 3 и выше, получим квадратное уравнение. Из него находятся две поправки к x_0 и два новых приближённых значения корня, к каждому из которых можно применить обычный метод Ньютона.

Ньютон был одним из создателей дифференциального исчисления. Основная идея исчисления состоит в том, что график любой функции, заданной формулой или бесконечным рядом, при рассматривании в микроскоп будет почти неотличим от некоторой прямой. Эта прямая называется касательной к графику функции, а её угловой коэффициент — производной данной функции. Геометрический смысл метода Ньютона: для решения уравнения $f(x) = 0$ вблизи приближённого корня x_0 нужно график функции заменить касательной прямой к этому графику в точке x_0 и найти пересечение касательной с осью X. Поэтому метод Ньютона иногда называют *методом касательных*. Метод Ньютона может быть с успехом применён не только к одному уравнению, но и к системе нескольких уравнений с несколькими неизвестными.

Суть метода остаётся прежней. Например, если для системы двух уравнений с двумя незвестными

$$f(x, y) = 0,$$
$$g(x, y) = 0$$

мы знаем приближённое решение x_0 и y_0, то более точное решение нужно искать, сделав замену

$$x = x_0 + p,$$
$$y = y_0 + q$$

и записав

$$f(x_0 + p, y_0 + q) = \ldots + Ap + Bq + C = 0,$$
$$g(x_0 + p, y_0 + q) = \ldots + Dp + Eq + F = 0.$$

Если в последней системе уравнений отбросить все члены, кроме уже выписанных, то поправки p и q

«...Уравнения высших степеней решаются совершенно так же ... доказательство метода явствует из самого способа действий, на основании чего его легко в случае необходимости вспомнить».

Исаак Ньютон

ЧТО ПИШЕТ НЬЮТОН

В работе «Анализ с помощью уравнений с бесконечным числом членов» (1664—1671 гг.) Ньютон излагает способ числового (приближённого) решения уравнения следующим образом.

«Пусть требуется решить уравнение $x^3 - 2x - 5 = 0$ и 2 представляет то число, которое отличается от искомого корня меньше, чем на свою десятую часть (т. е. Ньютон каким-то образом нашёл число, мало отличающееся от искомого корня. — *Прим. ред.*). Тогда я полагаю $2 + p = x$ и подставляю это выражение в уравнение, причём получается новое:

$$p^3 + 6p^2 + 10p - 1 = 0,$$

у которого следует определить корень p, чтобы прибавить к первому результату (корень x исходного уравнения будет равен $2 + p$, где p — корень нового уравнения. — *Прим. ред.*). Отсюда (пренебрегая $p^3 + 6p^2$ по малости. — *Прим. ред.*) имеем приблизительно

$$10p - 1 = 0 \text{ или } p = 0{,}1.$$

Поэтому я пишу в результате 0,1 и полагаю $0{,}1 + q = p$; это выражение я подставляю, как и раньше, причём получается

$$q^3 + 6{,}3q^2 + 11{,}23q + 0{,}061 = 0,$$

а так как уравнение

$$11{,}23q + 0{,}061 = 0$$

почти соответствует истине, то q почти равно –0,0054.

Полагая $-0{,}0054 + r = q$, я это выражение подставляю, как раньше, и продолжаю эти операции сколько угодно раз... и получаю приблизительно $r = -0{,}00004853$ и имею искомый результат:

+2,000 000 00
+0,100 000 00
–0,005 400 00
–0,000 048 53
———————
+2,09455147 = x».

Исаак Ньютон.

можно будет найти, решив систему двух линейных уравнений с двумя неизвестными.

Геометрически этот метод состоит в замене графиков функций f и g их касательными плоскостями.

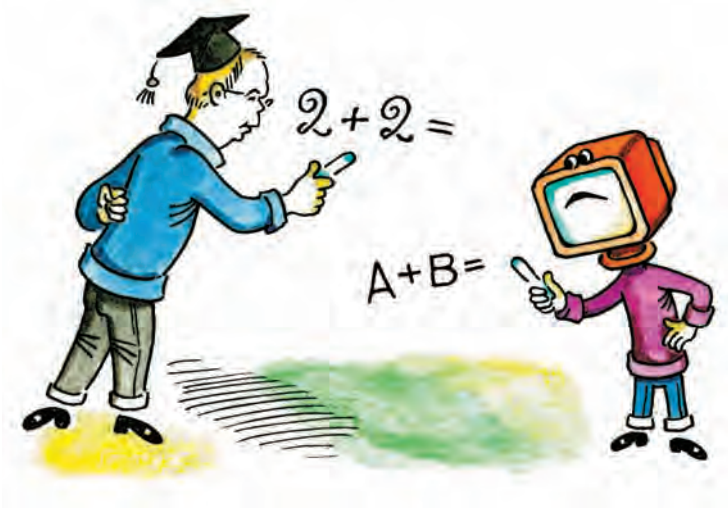
КОМПЬЮТЕРНАЯ АЛГЕБРА

ЧТО ТАКОЕ КОМПЬЮТЕРНАЯ АЛГЕБРА

Компьютерная алгебра — область математики, лежащая на стыке алгебры и вычислительных методов. Поэтому для неё трудно определить чёткие границы. Часто к компьютерной алгебре относят вопросы либо слишком алгебраические, чтобы содержаться в учебниках по вычислительной математике, либо слишком вычислительные, чтобы содержаться в учебниках по алгебре. Ответ на вопрос о том, относится ли конкретная задача к компьютерной алгебре, часто зависит от склонностей специалиста.

Термин «компьютерная алгебра» возник как синоним терминов «символьные вычисления», «аналитические вычисления», «аналитические преобразования» и т. д. Во французском языке он дословно означает «формальные вычисления».

Когда говорят о вычислительных методах, то считают, что все вычисления выполняются с вещественными или комплексными числами. В действительности же всякая программа для ЭВМ имеет дело только с конечным набором рациональных чисел, поскольку лишь такие числа представляются в компьютере. Для записи целого числа отводится обычно 16 или 32 бит, для вещественного — 32 или 64 бит. Это множество не замкнуто относительно арифметических операций, что может выражаться в различных переполнениях, например при умножении достаточно больших чисел или при делении на маленькое число. Ещё более существенной особенностью вычислительной математики является то, что арифметические операции над целыми и вещественными числами, выполняемые компьютером, отличаются от точных арифметических операций с рациональными числами. Более того, для компьютерных операций не действуют основные законы арифметических операций (законы ассоциативности, дистрибутивности).



Набор объектов в символьных вычислениях весьма разнообразен, в частности, понятие числа в них значительно шире, чем в вычислительной математике. Оно включает в себя целые числа, рациональные числа, алгебраические числа, кольца вычетов, конечные поля и др. Величина допустимых целых чисел практически неограничена: любая система компьютерной алгебры умеет работать с такими числами, как 1000! (факториал). Вещественные числа реже применяются в компьютерной алгебре, они задаются с произвольной точностью, например тысяча десятичных значащих цифр. Гораздо чаще употребляется множество рациональных чисел, которое всё равно остаётся конечным (поскольку оперативная память ЭВМ конечна), но пользователь никогда

не замечает этого ограничения. Он может выполнять операции практически с любыми рациональными числами, если это приемлемо по времени. Компьютерные операции над рациональными числами совпадают с соответствующими математическими операциями. Таким образом, снимается одна из основных проблем вычислительных методов — оценка погрешности вычислений.

В настоящее время широкое применение получили вычисления в кольцах вычетов. Кольцо вычетов по модулю натурального числа n можно представлять себе как множество целых чисел от 0 до $n-1$ с операциями сложения, вычитания и умножения. Результатом такой операции является остаток от деления на n результата соответствующей операции, выполняемой над целыми числами (если результат вычитания отрицателен, то прибавляем n). В реальных криптографических системах используются вычеты по модулю n, где n — натуральное число, десятичная запись которого содержит несколько сотен цифр. Особенно важное значение имеют кольца вычетов, где n — простое число. В таких кольцах существует однозначно определённая операция деления на любое ненулевое число (они называются конечными полями).

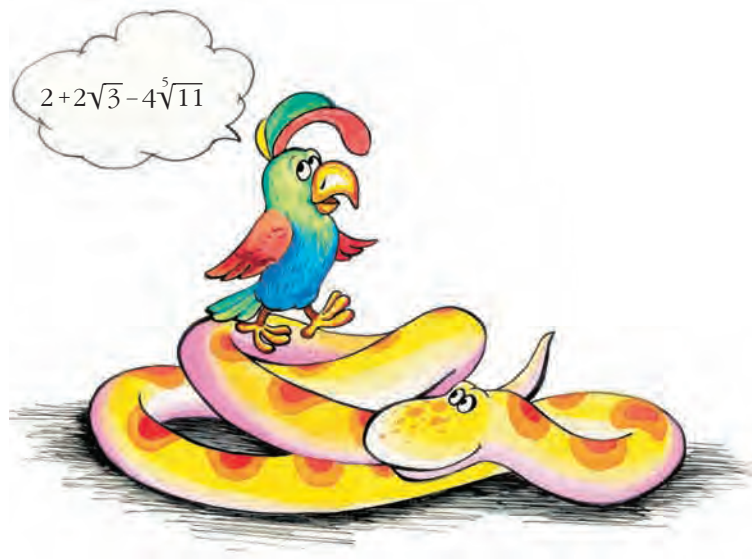
При решении нелинейных алгебраических (в частности, квадратных) уравнений сталкиваются с алгебраическими числами. Алгебраическими чиселами могут быть простые радикальные выражения, например:

$$2+2\sqrt{3}-4\sqrt[5]{11},$$

вложенные радикалы, например:

$$\sqrt[3]{2+34\sqrt[4]{2}}.$$

Уравнения степени 5 и выше в общем виде «неразрешимы в радикалах», поэтому для задания корня такого уравнения нужно задать минимальный многочлен, корнем которого является данное число, а иногда требуется также указать интервал на прямой или область в комплексной плоскости, где содержится единственный корень этого многочлена.



На полиномиальной арифметике основаны теоретические методы аналитической механики, реализованные во многих областях математики, физики и других наук. Кроме того, в компьютерной алгебре рассматриваются дифференциальные поля (функциональные поля), допускающие показательные, логарифмические, тригонометрические функции, матричные кольца (элементы матрицы принадлежат кольцам достаточно общего вида) и др. Для записи промежуточных результатов вычислений требуется значительный объём памяти ЭВМ, так как количество информации непомерно увеличивается.

АЛГОРИТМЫ КОМПЬЮТЕРНОЙ АЛГЕБРЫ

Компьютерная алгебра имеет дело с алгоритмами, которые существенно отличаются от алгоритмов вычислительной математики. К алгоритмам вычислительной математики предъявляются следующие требования: они должны решать задачу с необходимой точностью при заданных вычислительных ресурсах. В компьютерной алгебре вычисления обычно производятся без округления, анализу сходимости уделяется меньше внимания, зато набор математических, в первую очередь алгебраических,

В научных исследованиях и технических расчётах специалистам приходится гораздо больше заниматься преобразованиями формул, чем собственно численным счётом. Однако с появлением ЭВМ основное внимание уделялось автоматизации последнего, хотя более 40 лет назад ЭВМ применяли для решения таких задач символьных вычислений, как символьное дифференцирование. Активная разработка систем компьютерной алгебры началась в конце 60-х гг. С тех пор создано значительное количество систем, получивших различную степень распространения. Некоторые системы продолжают развиваться, другие отмирают, постоянно появляются новые.

объектов сложной структуры значительно шире. Поэтому ограничения по времени счёта и по используемой памяти в символьных вычислениях существенно более обременительны, чем в вычислительных методах.

Использование компьютеров в алгебраических исследованиях не только поставило перед специалистами ряд новых задач, но и заставило пересмотреть прежние задачи, в частности задачи, решённые «за конечное число шагов». При этом методы решения отличались большим разнообразием и универсальные методы для конкретных вычислений практически не применялись. С увеличением размера задачи резко возрастали время счёта и необходимая память компьютера.

В качестве примера рассмотрим задачу перемножения двух двузначных чисел в некоторой позиционной системе счисления, например в десятичной:

$$(10a+b)(10c+d) = \\ =100ac+10(ad+bc)+bd.$$

Имеем четыре операции умножения однозначных чисел, три операции сложения и две операции умножения числа на степень 10. Будем считать, что умножение на степень 10 выполняется очень быстро, сложение — медленнее, а умножение двух чисел — существенно медленнее (весьма естественное предположение).

Можно ли уменьшить количество умножений, увеличив, если необходимо, количество сложений? Оказывается, да. Впервые такой алгоритм, использующий только три умножения, предложил в 1956 г. А. А. Карацуба.

Перемножив $a-b$ и $d-c$, получим $ad+bc-bd-ac$; два других умножения нужны, чтобы вычислить ac и bd. Комбинируя эти три значения, используя лишь операции сложения и сдвига (умножения на степень двойки), легко получить искомую формулу, так как

$$ad+bc=(a-b)(d-c)+ac+bd.$$

Задача допускает различные обобщения, например: перемножить два комплексных числа, используя только три вещественных умножения; перемножить две матрицы второго порядка, используя только семь скалярных умножений.

Основой любой системы компьютерной алгебры являются вычисления с многочленами. Вот две задачи, активно исследуемые в последние десятилетия. Первая из них — задача факторизации, т. е. разложения многочлена на неприводимые множители. Пусть дан многочлен $f(x)$ от одной переменной с целыми коэффициентами, требуется разложить его на неприводимые множители. С точки зрения «чистой» математики эта задача давно решена полностью и окончательно: есть алгоритм, позволяющий находить требуемое разложение «за конечное число шагов». Один из таких алгоритмов получен в 1882 г. Леопольдом Кронекером, чьим именем он и называется в настоящее время, хотя ещё австрийский астроном Шу-

берт знал его за 100 лет до Кронекера. Алгоритм Кронекера основан на следующих фактах:

• если многочлен степени n разлагается на множители нетривиальным образом, то степень хотя бы одного из них не превосходит $n/2$;

• коэффициенты многочлена степени m однозначно определяются его значениями в $m + 1$-точке;

• если $g(x)$ делит $f(x)$, то $g(j)$ делит $f(j)$ для любого целого числа j;

• у любого ненулевого целого числа имеется только конечное число делителей.

К сожалению, этот метод требует перебора очень большого числа вариантов и не может применяться для решения реальных задач.

Следующий шаг в исследовании алгоритмов факторизации был сделан лишь в 60-х гг. XX столетия, когда нашли достаточно эффективный алгоритм для разложения на множители многочленов с коэффициентами из конечного поля. Этот метод в сочетании с некоторыми классическими и новыми результатами позволил разлагать на множители многочлены сотой степени (и даже выше).

Другая задача связана с упрощением систем алгебраических уравнений (т. е. многочленов) путём приведения системы к некоторой стандартной форме. Если система зависит от многих неизвестных, но является линейной, то её можно привести к диагональному виду («решить»). Стандартная форма нелинейных систем, зависящих от многих неизвестных, имеет более сложный вид (называемый базисом Гребнера). Она применяется во многих областях науки, поэтому эффективным методам её построения посвящено много работ.

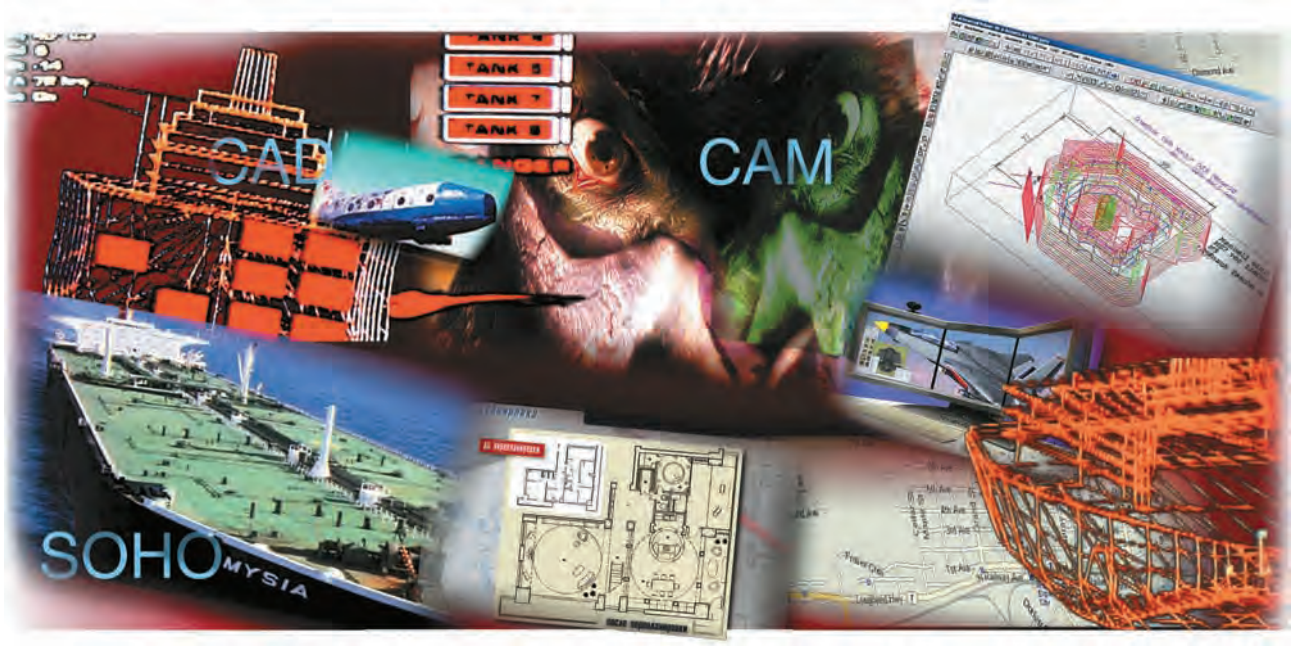
Компьютерная алгебра, как правило, имеет дело не с численными значениями функций, а с их представлением в виде формул достаточно общего вида. В частности, класс элементарных функций разрешает использовать в формулах наряду с арифметическими операциями операции извлечения корня, тригонометрические, логарифмические, показательные функции и ряд других. Основу дифференциального и интегрального исчислений составляют операции дифференцирования и интегрирования.

Операция дифференцирования легко описывается алгебраически, её реализация не представляет трудностей. Гораздо сложнее обстоит дело с обратной операцией — интегрированием. Вычислительная математика связана обычно с определёнными интегралами; результатом вычислений при этом является некоторое число. Компьютерная алгебра вычисляет неопределённые интегралы, являющиеся функциями. До недавнего времени считалось, что алгоритмов нахождения неопределённых интегралов для достаточно широкого класса функций не существует. Активно такие алгоритмы начали разрабатывать только в конце XX в.

Аналогичное различие наблюдается в подходе вычислительной математики и компьютерной алгебры к решению дифференциальных уравнений. Вычислительная математика понимает решение как функцию, значения которой вычислены в некоторых точках. Компьютерная алгебра стремится найти решение в виде формулы (что возможно далеко не всегда). Существенным вкладом в эту область является алгоритм решения линейных дифференциальных уравнений второго порядка, полученный в 70-х гг.

Компьютерная алгебра используется также при нахождении решений дифференциальных уравнений в виде степенных рядов, при анализе устойчивости решений, поиске периодических решений и в других задачах теории дифференциальных уравнений. Широко распространена компьютерная алгебра при построении разностных схем для численного решения дифференциальных уравнений.

Леопольд Кронекер.



АВТОМАТИЗАЦИЯ ИНТЕЛЛЕКТУАЛЬНОЙ ДЕЯТЕЛЬНОСТИ

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

CAD

С помощью ЭВМ легко рисовать прямоугольники, эллипсы, параллельные линии и фигуры различной сложности — здесь не нужна твёрдая рука чертёжника. На экране компьютера проще готовить и сам чертёж. В распоряжении конструктора — заготовки часто использующихся элементов, шрифты различных размеров и стилей. Можно взять фрагменты готовых чертежей (даже ввести со сканера). Нарисованную фигуру можно повернуть на произвольный угол или симметрично отобразить её, увеличить или уменьшить размер. ЭВМ позволяет однотипные действия записать в единую макрокоманду, создавать собственные команды, расширять функции системы, дописывая небольшие программы на специальном языке. Раньше надо было переделывать чертёж полностью иногда даже при небольших изменениях модели, в компьютер такая корректура вносится быстро, и сразу можно получить результаты изменений на бумаге. Для решения таких задач и разработаны системы CAD (Computer Aided Design), или, в русском варианте, САПР (система автоматизированного проектирования).

Первым применил компьютеры в системах автоматизированного проектирования и производства Патрик Хэнрети. В 1957 г. он разработал первую коммерческую систему числового программного управления — PRONTO. Затем, в 1960 г., Айвен Сазерленд на

компьютере TX-2 в лаборатории Линкольна Массачусетсского технологического института создал пакет SKETCHPAD, который считается первым шагом в индустрии CAD-систем.

В том же году была основана компания McAuto (McDonnell Douglas Automation Company). Она сыграла значительную роль в исследованиях в области автоматизированного проектирования. Первые CAD-программы использовали простейшие алгоритмы для прорисовки шаблонов, состоящих из отрезков и линий, сначала только в одной плоскости, а затем и в 3D-пространстве. В течение последующих лет CAD-системы непрерывно развивались. Уже в конце 70-х гг. типичная CAD-система базировалась на 16-битном миникомпьютере с максимумом памяти в 512 кбайт, объём дисковой памяти для хранения информации достигал 300 Мбайт, и стоила такая система около 125 тыс. долларов.

Инженеры сразу же оценили преимущества автоматизации построения геометрических элементов: автоматическую штриховку и нанесение размеров, точность чертежа и т. п. Обычный чертёж проектируемого объекта представляет собой традиционный способ плоского геометрического моделирования. Его выполняет человек с помощью линейки, циркуля и транспортира на листе бумаги, кальке и т. п. На этот лист шаг за шагом наносят всю заданную и вновь появляющуюся информацию. Нужно обеспечить определённую точность, сделать построения в максимально возможном масштабе. Однако на чертеже, созданном за кульманом, погрешность построений редко бывает меньше десятых долей миллиметра. CAD-системы дают возможность не только более точно выполнить построения и автоматически провести нужные вычисления, но и рассчитать параметры проектируемого объекта (например, центр тяжести, моменты инерции, вес, различные нагрузки), получить полный набор документов, включая расчёты модели и чертежи. Построенную на экране компьютера 3D-модель используют при изучении потенциального спроса, если изделие не спроектировано на заказ и его собираются самостоятельно продавать (допустим, когда хотят выпускать автомобиль). Прототип на экране часто выглядит значительно естественнее, чем модель, изготовленная из любого материала.

В CAD-системах обычно распространены два типа плоского моделирования. В первом случае основными элементами являются отрезки, дуги, линии, кривые, а в качестве базовых операций моделирования используют продление, обрезку и соединение элементов. К этому типу относится одна из популярных САПР — система AutoCAD.

В другом случае элементы — замкнутые контуры, а операции — объединение, дополнение, пересечение.

Компьютерное моделирование используют во многих специальностях; при этом применяют различные методы проектирования: геометрические, технологические, эргономические и т. п. Это привело к более глубокой специализации систем CAD/CAM/CAE (Computer Aided Engineering). Инструменты проектирования также различаются в зависимости от области применения. Например, в 1991 г. была создана система ArhiCAD — инструмент для архитекторов и дизайнеров.

Американская корпорация «Боинг» в 1978 г. начала одновременное проектирование новых (в то время) моделей самолётов 757 и 767. Несмотря на то что машины сильно отличались, у них было много унифицированных узлов. При разработке активно использовались CAD-системы. Это позволило сократить как время разработки, так и затраченные средства.

Впервые этот пакет САПР был продемонстрирован в Лас-Вегасе на выставке COMDEX в ноябре 1982 г. В 2002 г. выпущена уже семнадцатая версия пакета AutoCAD. Мировой лидер в разработке боевых самолётов ОКБ Сухого также использует AutoCAD для подготовки документации.

Аналогично разделение и при объёмном моделировании. Так, основные объекты здесь тела, ограниченные различными поверхностями, а изображение изделия получают при помощи операций объединения, пересечения и т. д. Данный метод часто применяют в машиностроении. Объёмное моделирование подчас удобнее чертежей: например, в сборочном чертеже болт может быть представлен несколькими видами, а в объёмной сборке — одним (моделью болта).

Однако далеко не всё можно получить путём геометрических построений. А если необходимо создать полную модель объекта, например самолёта, с абсолютной точностью? Для этого используют многоуровневую декомпозицию объектов. Изделие представляется как система агрегатов, которые, в свою очередь, состоят из узлов, а те — из деталей. Тот же метод применим и для объёмного моделирования. При этом в процессе проектирования электромотора нет необходимости иметь точную модель роторного шарикоподшипника. Достаточно его габаритной модели с указанием посадочных мест.

Часто хочется, опираясь на методы математического моделирования, создавать обобщённые модели изделий, например сконструировать общую модель поршневого двигателя внутреннего сгорания. (Все моторы этого вида основаны на одних и тех же принципах и состоят из однотипных деталей: цилиндров, поршней, шатунов и т. п.) Сегодня существуют методики расчётов деталей двигателя по таким исходным данным, как мощность и число оборотов. Однако конструктор, создавая двигатель, учитывает массу критериев: прочность, технологичность, ремонтопригодность и т. д. Сложность формализации подобных характеристик ограничивает построение обобщённых моделей, ведь, как правило, полной преемственности в конструкции не бывает.

CAM

Многие предприятия на этапе автоматизации обнаруживают, что ускорение черчения за компьютером часто не приводит к сокращению общих сроков выпуска изделия, а возросшее качество чертежей в результате не сильно влияет на качество самого изделия. Иногда даже говорят о неэффективности систем автоматизированного проектирования при решении производственных задач.

Но сложные станки с числовым программным управлением (ЧПУ) практически нереально использовать без предварительного геометрического моделирования. Так, крайне сложно не только вручную отфрезеровать две одинаковые пресс-формы автомобильного крыла, но и запрограммировать его обработку на стойке станка с ЧПУ: программа будет содержать десятки тысяч команд.

Наиболее эффективно геометрическое моделирование в случае, когда система включает в себя и конструкторское, и технологическое моделирование. Таким образом, есть смысл говорить об интегрированных CAD/CAM-системах, позволяющих и создавать модели, и готовить программы для станков с ЧПУ.

Система CAM (Computer Aided Manufacture) позволяет автоматизировать решение геометрических задач в технологии. Часто она выполняет расчёт траектории движения режущего инструмента. Результатом геометрического моделирования будет и набор чертежей (т. е. конструкторская документация), и программа в специальном формате для управления станком с ЧПУ. На этом этапе полученную программу можно отредактировать, а затем снова вернуть в CAD для проверки внесённых изменений.

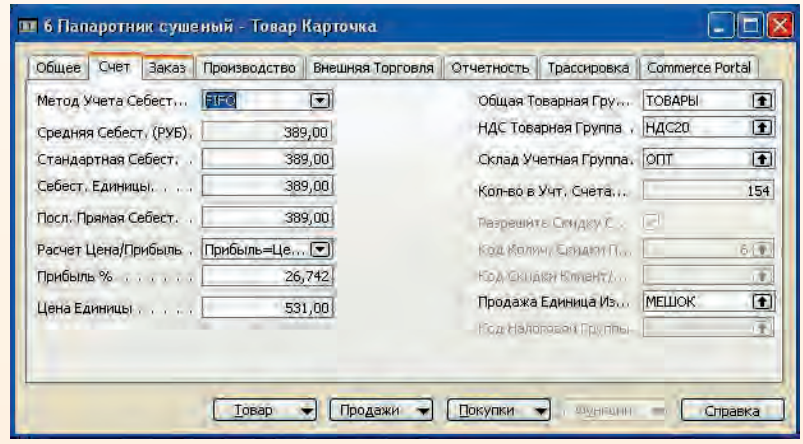
Первые станки с ЧПУ имели контроллеры для управления инструментом и моторами станков в процессе работы. Программа в такие системы вводилась с перфоленты, а само программирование сводилось к добавлению и удалению команд или изменению их параметров. У каждой подобной системы был собственный уникальный управляющий язык. Команды, представлявшие собой набор букв и чисел, задавали режим работы станка, перемещение инструмента, глубину резки, скорость движения и пр. В дальнейшем контроллер заменили компьютером, оборудованным дисплеем и клавиатурой, со специальным программным обеспечением для управления станком.

Внедрение CAD/CAM-систем позволяло не только автоматически сгенерировать программу по созданной 3D-модели изделия, но и увидеть результаты работы на экране, смоделировав выполнение программы на станке, обнаружить ошибки до производственного процесса. CAD/CAM-системы имели модульное построение, таким образом можно было по поставленной задаче спроектировать элементы производственного процесса, «подобрав» станки с ЧПУ для выполнения полного цикла производства. Дальнейшее расширение этой идеи привело к созданию CIM (Computer Integrated Manufacturing), гибких автоматизированных производств (ГАП). Результаты работы CAD используют и для подготовки программ для станков с ЧПУ, и для построения производственного процесса. Такие системы содержат подробную базу данных для настройки ГАП, включая все технологические процессы на всех станках. CIM выгодны, когда нужно наладить производство изделий малыми сериями. Так, мини-фабрики по производству интегральных схем используют CIM из-за особых качеств: быстроты и удобства настройки, сквозного обеспечения контроля качества.

Станок с ЧПУ.

Уже в ближайшее время индустриальные технологии могут развиться настолько, что отпадёт необходимость в массовом производстве. Изготовление единичных товаров не будет дороже серийного производства, а автоматические мини-заводы будут работать повсюду. Например, покупатель сможет заказать многие параметры сотового телефона: цвет, корпус, функции. Поистине научно-техническая революция!

ПОСТРОЕНИЕ ИНФОРМАЦИОННЫХ СИСТЕМ



Управление современным предприятием, даже если оно не очень большое, требует не только определённых навыков от руководства, но и полной и адекватной информации о состоянии дел. Бухгалтерия издавна накапливала и предоставляла информацию о движениях финансов. Иногда ошибочно полагают, что бухгалтерский учёт нужен для фискальных органов, т. е. для правильного исчисления налогов. На самом деле на основании информации, которая поступает от бухгалтерии, руководством фирмы принимаются управляющие решения. Даже проведя элементарный анализ, можно сделать выводы о состоянии дел в фирме: есть прибыль или убытки, на какую сумму осталось нереализованной продукции на складах и многое другое. Однако из одного лишь баланса предприятия многого не видно, например номенклатуры залежавшихся товаров на складах. Более того, причины затоваривания складов тоже не лежат на поверхности. Для этого нужен детальный анализ финансов, производственного процесса, состояния рынков сбыта. Причём на это потребуется значительное время, а пока ищут ответы на поставленные вопросы, могут быть допущены очередные ошибки в планировании, производстве, финансах.

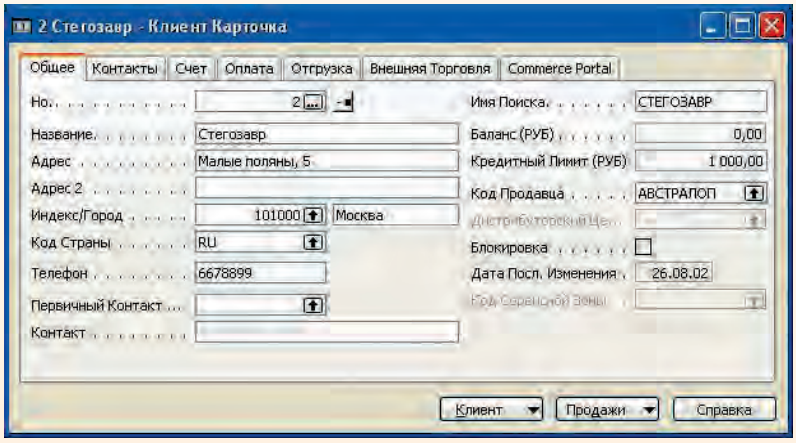
Поэтому в компьютеризованный век не обойтись без информационных систем, которые не только учитывают финансы, но и оперативно предоставляют руководителю необходимую информацию для анализа и принятия решений и даже являются в некотором смысле автоматизированными системами управления. С точки зрения руководителя, такая система должна быть сильно интегрирована, попросту иметь одну кнопку, нажав на которую директор фирмы получит исчерпывающую информацию и все возможные пути развития. Однако столь идеальных систем пока не существует. Поэтому информационные системы часто разрабатывают по заказу предприятий «с нуля», взяв за основу одну из современных баз данных. Но база данных, как и бухгалтерия, — это ещё не вся система, и индивидуальные разработки программного обеспечения оправданы по затратам и качеству скорее для очень специфичных задач учёта и управления, например учёт в реестре неких уникальных природных ресурсов. Предприятий, осуществляющих розничную торговлю, т. е. попросту магазинов, или производящих гарантийное обслуживание техники — тысячи, и процессы, протекающие в них, хорошо изучены и описаны. Так, фискальная часть финансового блока небольших коммерческих фирм одинакова практически у всех предприятий: все сдают в налоговую инспекцию одинаковые отчёты, в них совпадают планы счетов и т. п. Или, если фирма имеет склад, то учёт товаров, отгрузка и поставка, т. е. все процессы, одинаковы практически для всех складов. Поступивший товар должен быть учтён, в любой момент надо иметь возможность получить сведения о состоянии склада, об объёмах произведённых отгрузок.

Удобно использовать некоторые универсальные информационные системы для предприятий, которые можно настроить на конкретный тип деятельности. Если предприятие состоит из двух специалистов и занимается, например, консультациями с выездом к заказчику, то такой фирме, вероятно, достаточно только вести финансовой учёт в информационной системе. То есть иметь представление о текущем финансовом состоянии, получать информацию о текущих прибылях и убытках, сравнивать их с предыдущими периодами, планировать прибыль и многое другое.

Однако для учёта своих клиентов удобно вести базу по всем настоящим и, возможно, потенциальным заказчикам. В эту базу желательно включать не только сведения о клиентах, но и счета за работы, контролировать их своевременную оплату, автоматически направлять запросы должникам. То есть такие функции должны реализовываться не самостоятельно, а на основе информации из финансового блока, например, надо знать, что этот заказчик «хороший», потому что он оплатил счёт вовремя.

Ту же информацию нужно хранить, если фирма непосредственно занимается сбытом товаров. Но при сбыте требуется ещё иметь модуль, контролирующий закупки от поставщиков, связанный с финансами и складом фирмы.

Все проведённые предприятием работы также надо учитывать, особенно если они основываются на длительных договорах. Тут необходимо отслеживать прибыльность каждой такой группы работ, учитывать и планировать ресурсы, которые затрачены или будут затрачены. К ресурсам относятся не только материалы (можно рассмотреть в качестве примера строительную фирму), но и работники (бригады строителей). Если при начале работы будут просчитаны её



Если вы получили счёт в сумме 0 р. 00 к. за коммунальные услуги, то поспешите позвонить в ДЭЗ. Возможно, за неоплату этого счёта автоматическая информационная система отключит вам горячую воду.

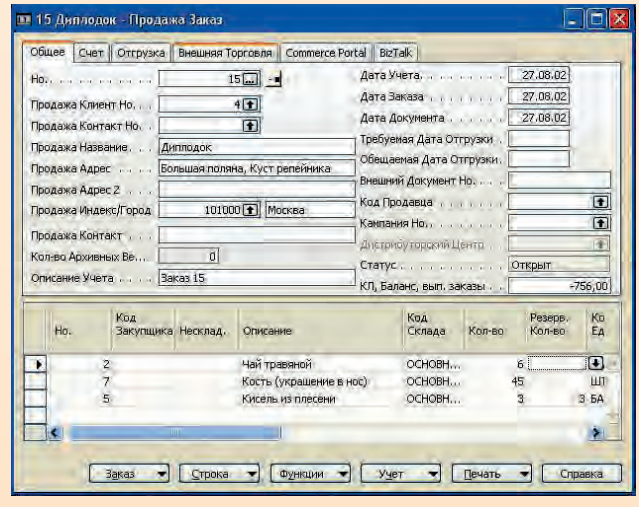
бюджет и планируемая прибыль, а в процессе осуществления руководитель сможет в любой момент посмотреть текущее состояние выполнения работы, затраты на неё и финансовое состояние, то это может оказаться решающим фактором, например при отказе от контракта на работу (вероятно, он будет просто невыгодным). Или планируемые ресурсы могут быть уже заняты (на другом строительстве). На все эти вопросы даст ответы проектируемая система.

Для производственных фирм естественной потребностью являются система управления производством, контроль качества, учёт, распределение и планирование производственных мощностей. Не надо забывать, что производство не должно быть убыточным. Фирма, выполняющая ремонт, т. е. сервисное обслуживание, естественно, должна иметь блок, управляющий и учитывающий этот часто циклический процесс. Так, на автосервисе важно регулярно приглашать клиентов на техобслуживание.

Одной из современных частей таких систем стали маркетинговые модули, которые позволяют учитывать и оценивать эффективность тех или иных контактов и презентаций. Подобные системы собираются из кирпичиков-модулей и затем настраиваются на конкретное предприятие.

Процесс конструирования и внедрения таких систем непростой, он проводится в несколько этапов. Как правило, в любом внедрении принимают участие аналитики, специалисты, хорошо представляющие область работы фирмы-заказчика, и программисты, которые осуществляют доводку и настройку системы на эту фирму. При работе аналитики часто используют специальные методы построения модели и процессов, происходящих на предприятии. Эти технологии хорошо проработаны, и, в идеале, результатом такого проектирования будет автоматическая настройка системы на конечную фирму. Необходимо всё-таки заметить, что чудес не бывает, и в каждом конкретном случае проявляется своя специфика. Поэтому параметрической настройкой не обойтись. Процессы, протекающие на предприятии, часто необходимо подвести под некоторый стандарт, принятый в настраиваемой системе, а часто, наоборот, перестроить систему, чтобы не разрушать привычные и отлаженные производственные процессы. То есть тут осуществляется движение навстречу друг другу: предприятия к системе, а системы к предприятию.

Эффективность внедрения системы зависит от самого ответственного этапа, когда осуществляются доводка и обучение персонала и от опытной эксплуатации переходят к промышленной, ведь всего 60 из 100 внедрений оказываются по-настоящему успешными. Важно, чтобы процесс проектирования информационных систем происходил по заранее разработанной технологии, чтобы внедрение было производственным процессом, а не отдавало программистским шаманством.



ЭКСПЕРТНЫЕ СИСТЕМЫ

У человека, выполняющего некую работу, часто возникает необходимость посоветоваться с кем-нибудь, достигшим в этом деле высот мастерства. Не зря поговорка гласит: «Ум хорошо, а два — лучше». В сложных случаях не помешала бы консультация и нескольких знатоков. Однако многие попадали и в такую ситуацию: заглохший автомобиль стоит у обочины, а проезжающие мимо дают «единственно верный» совет по устранению неисправности, причём все советы противоречат друг другу! Но это ещё не худший вариант... Похожее положение: врач в отдалённой сельской больнице или на судне в открытом море, столкнувшийся с симптомами незнакомого ему заболевания. Действовать надо срочно, ведь речь идёт о спасении жизни больного, а вот что делать — неясно.

В этих и десятках других ситуаций помочь могут так называемые *экспертные системы* (ЭС) — класс интеллектуальных программных систем, способных решать задачи, для которых отсутствуют алгоритмы или же недоступна вся информация, необходимая, чтобы применить известные алгоритмы и методы. Любая экспертная система предназначена для решения задач в конкретном виде профессиональной деятельности человека (лечение больных, анализ ситуации на биржевом рынке, определение месторождения полезных ископаемых по результатам геологической разведки, игра в шахматы и многое-многое другое). Понятно, что универсальная ЭС может оказаться слишком сложной, да и целесообразность её построения сомнительна.

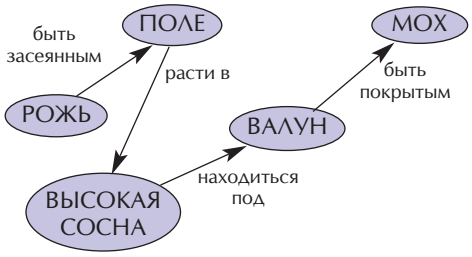
Для успеха любому профессионалу совершенно необходимо обладать особым набором знаний. В шахматах — не только научиться передвигать фигуры по доске, но и знать дебюты, уметь адекватно оценивать позицию, помнить сотни партий, сыгранных великими мастерами. В авторемонте — изучить устройство автомобиля и основные неисправности, вовремя обнаруживать и устранять их. Врачу надо блестяще освоить анатомию и физиологию человека; ему должны быть известны наиболее распространённые болезни, их симптомы, методы лечения. Совокупность знаний (надо чётко различать знания и данные), обязательных и достаточных для работы ЭС при решении задач какого-либо класса, называется *предметной областью*. Знания из одной предметной области связаны между собой; это могут быть описания различных объектов, процессов, явлений, а также отношений между ними. Все эти сведения хранят в *базе данных* (БД). Структура данных обычно достаточно проста (чаще всего это таблица). В отличие от данных знания имеют более сложную структуру и хранятся в специальном виде.

Знания можно разделить на *процедурные* и *декларативные*. Процедурные знания описывают последовательность действий — это алгоритмы, программы, служебные инструкции на случай чрезвычайной ситуации и т. д. Все остальные знания считаются декларативными. К ним относятся, например, и текст этой статьи, и формулировка аксиом Евклида, и итоговая таблица чемпионата мира по футболу...

При создании ЭС, работающей в некоторой предметной области, необходимо решить три основные проблемы.

Прежде всего требуется формализовать знания о предметной области, т. е. выбрать подходящую *модель представления знаний*. Существует несколько моделей.

В *сетевой* модели предполагается, что знания можно представить как совокупность объектов (понятий) и связей (отношений) между ними. Она допускает наглядную графическую интерпретацию в виде *семантической сети* (понятия и объекты — вершины сети, а связи и отношения между ними — соединяющие их линии). Например, текст «В поле, засеянном рожью, растёт высокая сосна. Под ней лежит покрытый мхом валун» будет выглядеть так:



Недостатком сетевой модели является то, что знания могут быть представлены в ней разными способами, а связи чаще всего неоднородны. Это весьма затрудняет и усложняет анализ знаний.

Стремление упростить строение сети, придать ей регулярный характер привело к появлению понятия *фрейм,* предложенного известным американским специалистом в области искусственного интеллекта Марвином Минским в 1975 г. Фрейм — это минимальная структура, описывающая объект, понятие, событие, процесс и т. п. Он имеет имя и состоит из последовательности *слотов*. Например, фрейм с именем «Матч» можно определить так:

```
Матч = <Игра> <Команда 1>
  <Команда 2> <Счёт>.
```

Такой фрейм называется *прототипом*, в нём слоты имеют переменные значения. Когда слоты получают некоторое значение, образуется *экземпляр* фрейма:

```
Матч = <Футбол> <Реал>
  <Барселона> <2:2>.
```

Легко заметить, что при исключении любого слота фрейм теряет полноту, а иногда и смысл.

Фреймы используют также для описания связей (отношений) между вершинами сети. Крайне важно, что значением слота, в свою очередь, может являться другой фрейм. Это позволяет строить сложно организованные иерархические структуры знаний.

В отличие от сетевой модели *продукционная* модель представления знаний (её идея была предложена американским математиком Эмилем Постом в 1943 г.) основана, как следует из названия, на системах *продукций*. Самая простая продукция состоит из имени и *ядра:*

```
<Имя продукции> : <Ядро>.
```

Ядро в виде «Если *A,* то *B*» (часть *A* называют *посылкой,* а часть *B* — *следствием*) описывает связь явлений, действий, фактов *A* и *B* («Если рубят лес, то летят щепки», «Если опустился туман, то зажечь фары»). Иногда ядру предшествует условие; тогда правая часть продукции принимает вид: «При выполнении условия *C*: если *A*, то *B*». Иными словами, условие задаёт ситуацию, когда ядро продукции может быть применено. (Находясь в квартире, зажечь фары невозможно, поэтому последний пример целесообразно дополнить условием: «Когда едешь в автомобиле, если опустился туман, то зажечь фары».) Продукция может содержать и некоторые другие элементы, при этом ядро — обязательная её часть наряду с именем.

Благодаря своим достоинствам продукционная модель получила широкое распространение. Её универсальность — в возможности представления как декларативных, так и процедурных знаний, простоте добавления новых знаний, а также возможности параллельной обработки системы продукций. Основной недостаток заключается в том, что при добавлении новых знаний сложно проверить их согласованность.

Слот — это часть фрейма. Он содержит название и значение, им может быть другой фрейм. Фрейм, из которого исключён любой слот, теряет свою полноту и смысл.

На основе различных моделей представления знаний строятся языки программирования, применяемые в ЭС. Например, известный язык PROLOG использует продукционную модель, а язык FRL (Frame Representation Language) — фреймовую.

Другая важнейшая проблема, связанная с разработкой ЭС, — создание *базы знаний* (БЗ), т. е. получение и отбор знаний, необходимых и достаточных для функционирования ЭС. Система знаний в ЭС должна быть полной и непротиворечивой. Базы знаний можно разделить на *формализованные* и *неформализованные*. Первые либо сведены в справочные руководства, либо записаны в виде инструкций, таблиц, формул, алгоритмов, которые легко вводятся в БЗ. Но этого слишком мало для успешной профессиональной деятельности. Например, чтобы хорошо играть в шахматы, совершенно недостаточно просто знать правила игры!

Второй вид знаний нельзя получить в готовом виде — их описаний нет в книгах, отсутствуют соответствующие формулы и алгоритмы, но зато эти знания можно сформулировать в виде набора некоторых правил, основанных на практике, обобщающих опыт, умение, интуицию специалистов. ЭС ориентированы именно на решение задач с использованием неформализованных знаний.

Важнейшим этапом в создании ЭС является получение необходимой информации от экспертов (специалисты, обладающие глубокими познаниями в какой-либо предметной области).

Именно они определяют основные знания, которыми будет обладать ЭС.

Вопрос общения с экспертом сам по себе непрост. Помимо проблем психологического характера (как установить контакт, как правильно сформулировать, что же надо узнать, наконец, как оценить достоверность сообщённых сведений) существует ещё одна. Зачастую специалист сам не в состоянии объяснить, почему он принимает то или иное решение. Например, гроссмейстер после выигранной партии долго разбирает позицию, в которой он сделал победный ход. Анализ убеждает в том, что ход действительно был лучшим. Очевидно и то, что во время партии такой анализ невозможен — дорога каждая секунда. Значит, гроссмейстер действовал интуитивно, руководствуясь общим впечатлением от позиции, сложившимся на основании эрудиции и огромного опыта. Точно так же полководец, во время сражения бросивший свои последние резервы на левый фланг, вряд ли объяснит, почему он сделал именно это и именно тогда, а не часом раньше. Для внесения подобных знаний в БЗ надо уметь формулировать основания для принятия таких решений.

С помощью базы данных можно получить информацию о конкретном предмете, процессе или ситуации, причём лишь ту, которая была ранее туда помещена. При работе компьютерных программ данные выступают в роли объекта обработки, они пассивны. Программа находит их, изменяет и снова записывает в определённое место памяти. Базы знаний отличаются от баз данных наличием встроенных механизмов логического вывода. Благодаря этому знания во время работы ЭС ведут себя активно: изменение тех или иных знаний может вызвать обращение к программам их обработки. Встроенные процедуры позволяют не только делать иными имеющиеся знания, но и получать из них новые. Таким образом, база знаний пополняется как знаниями, полученными от экспертов, так и знаниями, вырабатываемыми самой экспертной системой в процессе работы.

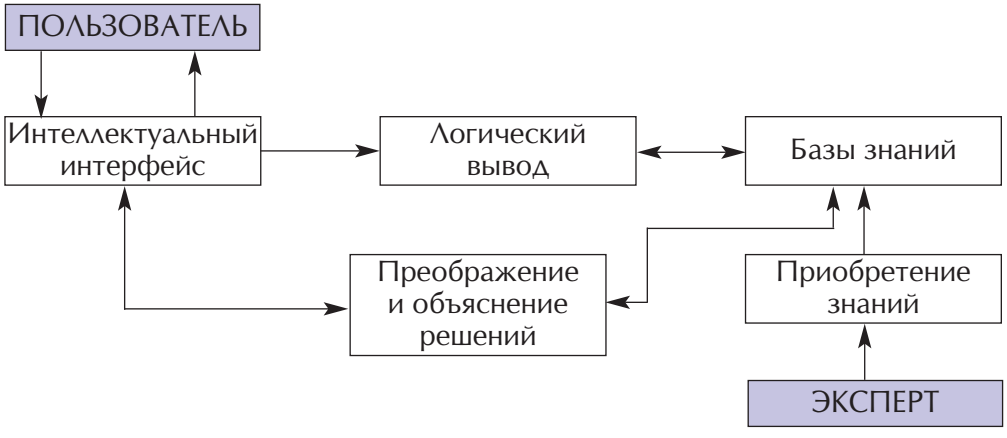
Очень важно, что в ЭС реализуются не одни механизмы логического

вывода, опирающиеся на классическую логику (дедуктивные). В них используются и так называемые правдоподобные рассуждения (нечёткий вывод, основанный на нечёткой логике), и элементы рассуждений по аналогии и по ассоциации, которые человек обычно применяет при решении задач в условиях неопределённости.

В совокупности средства, решающие эти и некоторые вспомогательные задачи, и образуют экспертную систему. Её обычная схема представлена на рисунке.

Блок логического вывода вырабатывает одно или несколько решений. ЭС должна не навязывать их пользователю, а предлагать возможные варианты решения в порядке их предпочтительности или вероятности. Работа ведётся в режиме диалога (интерактивный режим), поэтому желательно, чтобы ЭС могла «объяснить» свои действия и выводы. Ведь иногда пользователь считает полученный вариант неподходящим или не лучшим. Причина этого кроется не столько в самом решении, сколько в недостаточной его обоснованности.

В настоящее время существуют ЭС, помогающие человеку в различных областях — от составления прогноза погоды до синтеза новых органических веществ с заданными свойствами. Нет сомнения, что сфера их применения будет постоянно расширяться, а возможности возрастут.



РАСПОЗНАВАНИЕ ОБРАЗОВ

Под термином «распознавание образов» обычно подразумевается направление научно-исследовательских работ, относящихся к области прикладной математики. Специалисты по распознаванию образов с помощью средств вычислительной техники решают ряд задач, связанных исключительно с интеллектуальной деятельностью человека. Основной метод состоит в построении математических моделей процесса

«МЫСЛИТЕЛЬНАЯ МАШИНА»

Потребность проверить собственные умозаключения и выводы, установить их правильность всегда была свойственна людям.

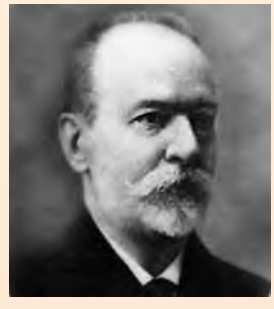
Французский философ Рене Декарт (1596—1650) полагал, что знания, которые приобретает человек, могут быть выведены столь же строго, как и теоремы в математике. Нельзя ли представить рассуждения в виде вычислений? Над этой проблемой размышлял Готфрид Вильгельм Лейбниц. Правда, спустя ещё полтора столетия, в середине XIX в., когда Чарлз Бэббидж предложил проект своей аналитической машины, его помощница леди Лавлейс сомневалась в возможности этого. Она писала, что аналитическая машина не претендует на то, чтобы создавать что-то новое, и не может предугадать аналитические зависимости или истины. Впрочем, вполне вероятно, что её мнение было основано на весьма ограниченных (на современный взгляд) характеристиках аналитической машины.

Тем не менее ещё при жизни Бэббиджа, в 1870 г., английский логик Уильям Стенли Джевонс (1835—1882) создал первую в мире машину, механизировавшую простейшие логические выводы (фактически, она могла вычислять функции алгебры логики). Впоследствии её воспроизвёл российский учёный, профессор П. Д. Хрущёв (1849—1909), а профессор А. Н. Щукарёв (1864—1936) усовершенствовал конструкцию. В 1914 г. в журнале «Вокруг света» А. Н. Соков писал: «Если мы имеем арифмометры, складывающие, вычитающие, умножающие миллионные цифры поворотом рычага, то, очевидно, время требует иметь логическую машину, способную делать безошибочные выводы и умозаключения одним нажатием соответствующих клавиш. Это сохранит массу времени, оставив человеку область творчества, гипотез, фантазии, вдохновения — душу жизни».

Однако говорить о создании настоящей «мыслительной машины» стало возможным только после изобретения такого устройства, как компьютер.

Рене Декарт.

Уильям Стенли Джевонс.



П. Д. Хрущёв.

А. Н. Щукарёв.

распознавания, создаваемых путём предположений-гипотез о том, как задачи подобного рода решаются людьми. Примерами служат задачи идентификации (распознавания) человека по его фотографии, узнавания мелодии, диагностики болезни и т. п.

Обычно объекты распознавания (фотоизображение, мелодия, болезнь) называют общим словом «образ». Процесс распознавания образа человеком обычно трактуется в теории распознавания образов (или просто теории распознавания) как отнесение объекта к той или иной совокупности в чём-то близких объектов. Так, узнавание человека по фотографии можно рассматривать как отнесение объекта (фотографии конкретного человека) к совокупности всевозможных фотоизображений этого человека. Узнавание мелодии — как отнесение объекта (мелодии) к совокупности всевозможных исполнений данной мелодии. Диагностирование болезни — как отнесение конкретного набора наблюдаемых у больного симптомов к совокупности всевозможных проявлений симптоматики одной конкретной болезни.

Группа объектов, обладающих общими признаками, называется *классом*. То есть теория распознавания занимается задачами определения истинности или ложности утверждения о том, что какой-то объект принадлежит к некоторому классу объектов. В целом теория распознавания образов изучает плохо формализуемые задачи, когда человек не в состоянии сформулировать однозначные правила, позволяющие ему решать задачи распознавания быстро и эффективно.

В теории распознавания предполагается, что любой объект обладает конечным набором характеристик, значения которых его и определяют. Другими словами, эти характеристики описывают объекты. Поэтому характеристики — это признаки, совокупность всех признаков — признаковое пространство, а количество признаков — размерность признакового пространства. Так, для фотоизображений признаковым пространством считается совокупность строк цифровой развёртки фотоизображе-

ния, где каждая строка является признаком. Для мелодии в качестве признакового пространства рассматривается совокупность коэффициентов разложения Фурье для акустических колебаний, каждый из коэффициентов Фурье — признак. Очевидно, что размерность признаковых пространств для приведённых выше примеров может быть весьма велика, так как в задачах не удаётся выделить небольшое количество признаков, содержащих исчерпывающую информацию для отнесения объектов к какому-либо классу.

Основой всех моделей, используемых в теории распознавания, является так называемая *гипотеза компактности*. Согласно ей, предполагается такое устройство признакового пространства, что в нём похожие объекты находятся близко, а непохожие — далеко. Поскольку все дальнейшие построения математических моделей распознавания образов явно или неявно опираются на данную гипотезу, успех решения конкретной задачи идентификации определяется удачным выбором признакового пространства. Классы объектов в признаковом пространстве в соответствии с гипотезой компактности расположены так, что все объекты из одного и того же класса близки друг к другу и далеки от объектов из чужих классов.

Человек, прежде чем решать любую задачу распознавания, обязательно проходит стадию обучения. Ему показывают на примерах, какие объекты к каким классам относятся, перечисляют все классы, проверяют, хорошо ли он усвоил данные сведения. В моделях применяется такой же подход.

Другими словами, прежде чем приступить к распознаванию, модель необходимо научить делать разбиение признакового пространства. В теории распознавания образов существует два способа обучения: обучение с учителем и обучение без учителя.

При обучении с учителем априорно задаётся совокупность объектов с известной принадлежностью к тому или иному классу. Эта информация используется для разбиения признакового пространства на области, ассоциированные с соответствующим классом. Качество обучения, т. е. качество распознавания, оценивается по работе распознающей модели на контрольной совокупности объектов, для которых известна принадлежность к конкретному классу, но эта информация никак не используется обученной моделью.

При втором способе также задаётся совокупность объектов, однако их принадлежность к классам не определена. Вообще говоря, даже число классов может быть неизвестно. В этом случае признаковое пространство разбивается на конечное число областей в соответствии с привносимым извне критерием, условно отражающим гипотезу компактности и специфику задачи.

Существующие методы решения задач распознавания образов весьма сильно опираются на специфику исходной информации. Поэтому к настоящему времени сформировались широкие научные направления, специализирующиеся на распознавании визуальных, музыкальных образов,

на медицинской диагностике и т. п. Одновременно развивается и общая теория, позволяющая строить универсальные системы, не ориентированные на специфику распознаваемых образов и дающие возможность добиваться оптимального качества распознавания.

В качестве иллюстрации рассмотрим задачу автоматического ввода машинописных текстов в ЭВМ как задачу распознавания при обучении с учителем.

Пусть (для определённости) в ЭВМ вводится русский машинописный текст. Под автоматическим вводом будем подразумевать умение ЭВМ обработать текст так, что в её памяти размещается последовательность цифровых кодов, каждый из которых отождествлён с определённой буквой, а порядок следования букв повторяет порядок букв в тексте.

Прежде всего нужно определить, что рассматривается в качестве объектов и признакового пространства. Для этого необходимо представлять, в каком виде машинописные символы могут быть введены в ЭВМ и какие из характеристик требуются для дальнейшей обработки. Обычным способом введения текстов является их постраничное сканирование. Вместо текста в ЭВМ возникает числовая матрица, её размер равен площади отсканированного материала. Элементы матрицы представляют собой числа, пропорциональные степени яркости в соответствующей точке сканирования. Другими словами, в ЭВМ при сканировании страницы машинописного текста образуется картинка из точек разной яркости, которые в совокупности дают изображение, повторяющее оригинал.

Существенное значение имеет то, что в машинописном тексте все буквы разделены между собой пробелами (промежутками). Это позволяет ЭВМ с помощью специальной программы разбить всё изображение отсканированной страницы на небольшие куски, где каждый соответствует одной и только одной букве. Программа располагает частицы исходной матрицы в порядке следования букв текста: слева направо и сверху вниз.

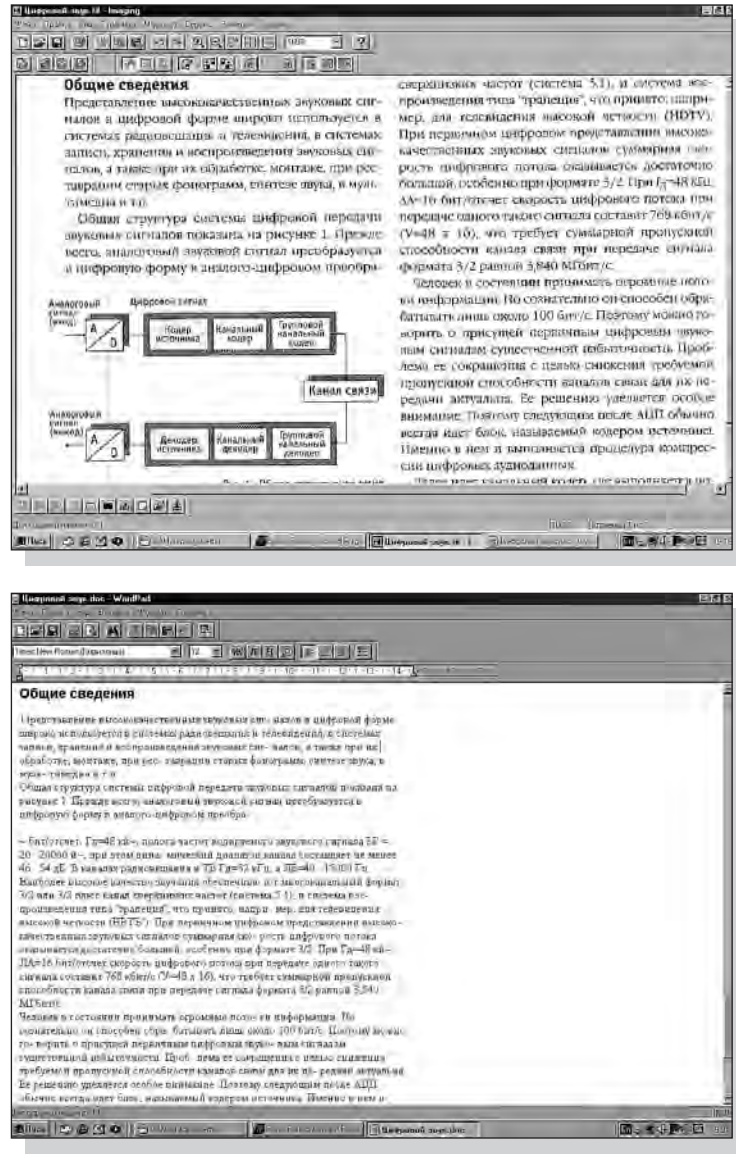
Бесспорно, в качестве объектов естественно рассматривать полученные небольшие фрагменты исходной числовой матрицы. В роли характеристик таких объектов-матриц выступают различные величины. Например, значения статистических моментов, моментов ортогональных разложений и прочие вычисляемые величины. Конечная совокупность этих характеристик объектов и образует признаковое пространство. Каждому объекту-матрице присущи конкретные значения характеристик-признаков. Таким образом, объектам отвечают точки в признаковом пространстве, размерность которого определяется количеством вычисляемых для каждого объекта харак-

теристик. Необходимо отметить, что в силу погрешностей печати и сканирования одним и тем же буквам соответствуют разные точки в признаковом пространстве.

Можно предположить, что каждой букве алфавита соответствует своя область в признаковом пространстве и данные области не сильно перекрываются. Значит, в признаковом пространстве существуют 33 (по числу букв русского алфавита) различных класса. Тогда задача автоматического ввода русского машинописного текста может быть сформулирована как задача отнесения каждой введённой в ЭВМ буквы текста к одному из 33 классов. До начала автоматического распознавания букв (отнесения к 33 классам) вводимого текста необходимо ввести в ЭВМ эталонные изображения всех букв с точным указанием классов, к которым они относятся. Такая подготовка позволит ЭВМ построить 33 области в признаковом пространстве, с каждой из которых будет связана одна и только одна из букв алфавита.

Процедура построения областей и является обучением ЭВМ с помощью учителя. Такое обучение проводят различными методами. Вот пример задачи, которую можно решить путём разбиения признакового пространства на 33 непересекающиеся области с минимальным количеством гиперплоскостей. Качество выбранного метода может быть всегда оценено стандартным образом — подсчётом доли ошибок при распознавании контрольной последовательности букв.

Итак, решение задач, сформулированных как задачи распознавания образов, базируется на априорных (эвристических) предположениях о наличии признакового пространства



Распознавание текста: вверху — сканированный текст; внизу — результат распознавания текста.

со свойствами, позволяющими построить его разбиение так, чтобы в контрольном примере электронно-вычислительная машина допускала минимальное колличество ошибок.

ДОМАШНИЙ ОФИС

В ряду персональных компьютеров 80-х гг., предназначенных для различных организаций и учреждений, для учёбы, для дома, последние обычно обладали самыми скромными возможностями и минимальным набором программ. Однако именно такими машинами пользовались в малом

бизнесе. Ведь систематическое обслуживание персональных компьютеров, которое предполагает не только их установку и ремонт, но и консультации, облегчающие работу с программным обеспечением, стоит дорого и доступно лишь крупным фирмам. Частным лицам и мелким предпринимателям, с трудом позволившим себе покупку ЭВМ с набором программ, её профессиональное сопровождение оказывалось не по карману. Поэтому изначально от компьютеров такие клиенты требовали высокой надёжности, а от программ — простоты установки и использования. Стоимость набора программ, эксплуатируемых в домашнем офисе, часто в несколько раз превышает и так немалую стоимость компьютера. Поэтому суммарные затраты на полноценное оборудование компьютеризованного офиса часто велики как для дома, так и для малых фирм.

В 90-х гг. ситуация изменилась. Многие домашние машины уже не уступали по своим данным компьютерам, установленным в солидных компаниях. Это стало возможным благодаря массовому производству персональных ЭВМ и снижению их стоимости.

В то же самое время в продажу поступили ноутбуки (от *англ.* notebook — «блокнот») — переносные компьютеры с плоскими жидкокристаллическими экранами. Такие машины удобны для малого бизнеса, поскольку владелец может всю документацию носить с собой и даже продолжать работу в другом офисе, гостинице, дома и т. д. Пусть ноутбук более дорогой, чем настольный персональный компьютер, зато он незаменим при частых переездах.

Для новых мощных машин требовалось и соответствующее программное обеспечение. Ни частных пользователей, ни бизнесменов не устраивал набор спецвыпусков программ с сильно ограниченными возможностями. Однако покупать нужные программы им по-прежнему было тяжело. С учётом этого некоторые фирмы-разработчики программ стали предлагать так называемые «лимитед эдишн» (*англ.* limited edition — здесь «ограниченные серии»), практически вполне полноценные и ориентированные на работу дома и на малом предприятии.

В конце XX в. появился термин SOHO (*англ.* Small Office / Home Office — «Малый офис / домашний офис»), обозначающий тип компьютера для дома и малого бизнеса с соответствующей компоновкой и потребностями в программах и оборудовании.

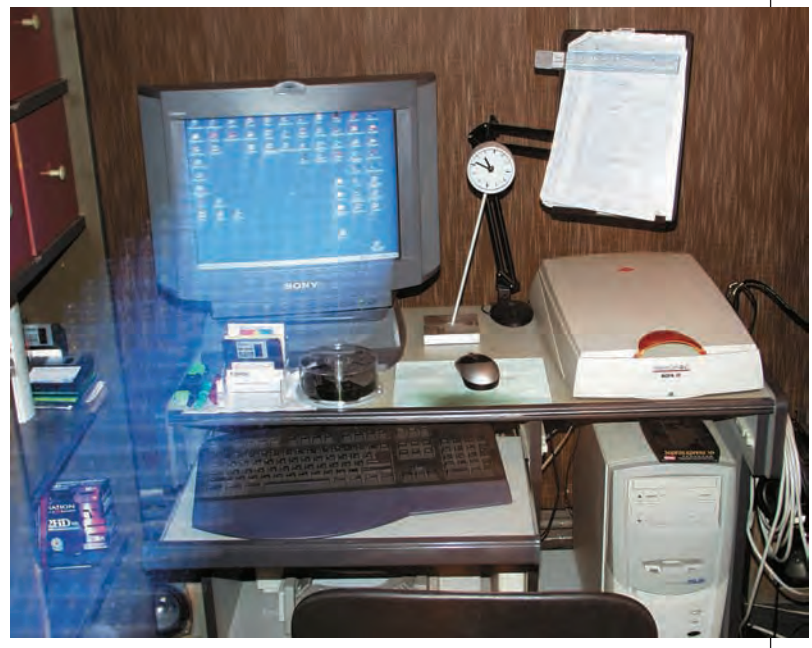
Современный SOHO базируется на оконной системе типа Microsoft Windows; его офисный пакет состоит из развитого текстового редактора, электронных таблиц, базы данных и почтовой программы, совмещающей

функции простейшего планировщика, записной книжки и календаря. (Естественно, что такой офисный пакет фирмы Microsoft носит название Office и включает все перечисленные компоненты: редактор Word, электронные таблицы Excel, базу данных Access и пакет Outlook.)

Системы, подобные Microsoft Outlook, фактически служат органайзером. Как правило, с их помощью можно получать и отправлять почту через Интернет, вести календарь и предстоящие дела, создавать персональную записную книжку с координатами партнёров, составлять перечень текущих задач и контролировать их выполнение, заносить разнообразные записи, делать заметки и т. д.

Пакет для подготовки презентаций позволяет комбинировать текст, рисунки и фотографии, а затем просмотреть полученные для демонстрации перед аудиторией кадры и выстроить их в желаемой последовательности — как на экране компьютера, так и на обычном экране при помощи проектора, подключённого, например, к ноутбуку.

Набор программ для дома и офиса, как правило, включает и программу для подготовки собственных web-страниц, где на всеобщее обозрение можно представить свой бизнес или сообщить информацию о себе лично.



Если в фирме более двух сотрудников, может потребоваться система управления с бо́льшими возможностями, чем предусмотренные стандартным офисным пакетом. Такие системы нечасто используются дома. Кроме того, нередко интересы предпринимателя и особенности бизнеса определяют необходимость в дополнительном программном обеспечении или оборудовании, в том числе весьма дорогом.

ИГРЫ

КОМПЬЮТЕРНЫЕ ИГРЫ

Появление компьютерных игр можно отнести к моменту, когда компьютеры из сферы экспериментальной и почти секретной (ведь на них должны были рассчитываться траектории снарядов и ракет во время военных действий) начали переходить в мир научный и практический. Это произошло в конце 60-х гг. XX в.

Компьютер стал обладать неким более или менее дружественным пользователю интерфейсом — вместо лампочек и загадочных индикаторов появились алфавитно-цифровые дисплеи. Конечно, ни о какой графике не могло идти и речи... Но за компьютерами работают люди, и ничто человеческое им не чуждо. И вот в один прекрасный вечер после тяжёлого трудового дня молодой программист (а кому ещё могла прийти в голову идея использовать компьютер не по прямому назначению) решил написать небольшую программу, которая играла бы с ним в какую-нибудь не очень сложную игру, например «Быки и коровы»... И конечно же такая идея пришла в голову не только ему одному... Вскоре программы для развлечения начали появляться всё чаще и чаще и даже стали входить в состав пакетов программ, поставляемых вместе с компьютерами.

С созданием компьютерной графики и появлением настоящих домашних компьютеров игровая индустрия стремительно выросла. Игры выпускались тысячными тиражами, не считая пиратских копий. Примерно за десять лет для домашнего компьютера ZX-Spectrum фирмы Sinclair Research было выпущено более 6 тыс. (!) наименований игр.

Сейчас игровая индустрия является одной из точек опоры, на которых стоит индустрия персональных ЭВМ, да и для чего нужен дома компьютер, как не для игр?

В каждой шутке есть доля правды, и мир, наверное, не увидел бы «компьютерной революции», если бы не компьютерные игры.

В игровой индустрии выделяют:

• *Игры текстовые.* К ним относятся «Star Trek», «Adventure», «Rogue». Задача игрока: выполнить некоторые действия, чтобы достигнуть цели (завоевать галактику, найти сокровище, победить дракона и т. п.). Для таких игр достаточно лишь алфавитно-цифрового дисплея. Управление осуществляется через ввод простых команд вроде «передвинуться на десять шагов вперёд». На это компьютер отвечает: «Вы оказались в тёмном подземелье, и рядом с вами стоит большое чудовище с недобрыми намерениями. Ваши действия?». Чтобы игра продолжалась, игрок должен вводить новые команды.

• *Игровые приставки.* В 1980 г. кто-то из менеджеров фирмы ATARI вдруг осознал, что обычными игрушками никого не удивишь и пора осваивать новые направления. Электроника в те времена уже позволяла делать многое. Так появились игровые приставки — небольшие коробочки, подключаемые к обычному телевизору и позволяющие играть в простенькие на современный взгляд игры. Компьютерная графика была новшеством, и схематичные изображения персонажей на телевизионном экране просто поражали воображение покупателей. В дальнейшем мощность приставок росла, сами игры (программы) перекочевали в так называемые *картриджи,* а позже и на компакт-диски. Сегодня игровые приставки порой обгоняют по возможностям самые «навороченные» домашние компьютеры. В них используются современные графические процессоры, специально оптимизированные для быстрой обработки графики. Помимо этого, приставки достаточно дёшевы по сравнению с персональными компьютерами с аналогичными возможностями, и покупатель может позволить себе купить новый аппарат взамен вышедшего из моды.

• *Домашние компьютеры.* Игры на приставках были очень популярны, но пользователь хотел не только играть в игры, но и решать какие-то простые задачи: писать программы, вести каталог домашней литературы и т. д. Домашние компьютеры подключались к обычному телевизору, обладали небольшим объёмом памяти, а в качестве внешнего устройства хранения данных использовали кассетный магнитофон. Наиболее известные из них — ZX-Spectrum (Великобритания), Commodore 64 (США и Западная Европа), MSX (Япония).

Сейчас игры выпускаются в основном под платформу PC и Macintosh. С каждым годом в продаже появляется всё больше новых игр на любой вкус. Все последние достижения в области трёхмерной графики, звука, искусственного интеллекта нашли своё место в этих новинках.

Вы родились слишком рано, если ваших личных дисков с играми значительно больше, чем у ваших детей.

Компьютерные игры можно разделить по жанрам.

• *Аркадные игры.* Уничтожить пришельцев, собрать все клады или убежать от каких-либо монстров — вот цель в этих играх. Они отличаются большим разнообразием графики, звуковых эффектов. При этом, если игрок не слишком уравновешен, не исключены последствия в виде разбитых клавиатур и сломанных *джойстиков* — специальных устройств для управления движением.

• *Стратегия.* Жанр не менее азартный, чем предыдущий. Обычно игрок — военачальник, под командованием которого находятся какие-либо войска. Тут, чтобы разгромить вражескую армию, помимо быстроты реакции необходимо проявить умственные способности.

• *Симуляторы.* Здесь пользователь может попробовать себя в роли пилота или капитана космического корабля. Современные авиасимуляторы настолько сложны, что порой не хватает клавиш на клавиатуре, чтобы воспроизвести все детали управления летательным аппаратом.

• *Логические игры.* Игры-головоломки, карточные игры, шахматы и шашки. Здесь игроку есть над чем подумать.

• *Ролевые игры.* Они обязаны своим появлением настольной игре «Подземелье и драконы», по правилам которой каждый игрок имеет какие-либо отличительные характеристики (сила удара, сила магии, жизненная сила и т. д.). От баланса этих показателей зависит роль игрока в группе. Так, например, маг обладает наибольшим интеллектом, боец — силой, а вор — ловкостью. Группа должна выполнить некое задание: победить дракона, найти сокровище или спасти принцессу. Каждое действие обыгрывается, ведёт игру «мастер подземелий», отслеживающий каждое движение игроков и объявляющий

ФЕНОМЕН «ТЕТРИСА»

Российский программист Алексей Пажитнов создал чудовище. И имя этому монстру — «тетрис». Знаменитая игра покорила все платформы, поселилась в карманных устройствах и сотовых телефонах. Правила её просты, как всё гениальное. Сверху падают геометрические фигуры, которые можно вращать и укладывать так, чтобы между ними не оставалось свободного места. Мегаватты энергии были «съедены» компьютерами по всему миру, врачи-окулисты получили сверхприбыли от притока новых пациентов, а «тетрис» продолжает шествие по экранам наших мониторов. Пожалуй, это лучшая компьютерная игра из всех созданных когда-либо.

о новых событиях. В электронной версии «мастером», как правило, выступает компьютер. По времени игра разделена на раунды. За каждый раунд группа может произвести определённое количество действий. Это очень похоже на настольные игры, где продвижение по игровой доске зависит от выпавших на кубике очков. С развитием Интернета ролевые игры обрели ещё большую популярность, так как теперь стало неважно, где находится партнёр по игре — в Китае или Норвегии.

Влияние игр на неокрепшее сознание детей и подростков неоднозначно. Кто-то развивает логическое мышление, кто-то, наоборот, слишком увлекается стрельбой по монстрам и забывает про окружающий реальный мир. Бесспорно одно: зарождается интерес к компьютеру.

Игры очень важный способ обучения. Довольно непростую для восприятия информацию можно передавать в игровой форме, причём это актуально как для пятилетних детей (обучение счёту или письму), так и для вполне взрослых людей (деловые игры или тесты в игровой форме).

Если говорить о компьютерной промышленности, то игры, пожалуй, один из сильнейших стимулов для выпуска нового «железа», разработок в области трёхмерной графики и объёмного звука.

Игры — часть истории компьютера. История развивается по спирали. И впереди новый бум приставок, разделение компьютеров на домашние и «не очень», огромное количество новых игр и связанных с ними технологий и аппаратных средств.

ТРЕНАЖЁРЫ

Компьютерные технологии настойчиво проникают в повседневную жизнь. Компьютеры применяются даже в тех областях, которые на первый взгляд не имеют ничего общего с электроникой и информатикой.

Казалось бы, какая связь между современными компьютерами и спортом, где всё решают реакция, сила и выносливость спортсменов? Самая непосредственная. На компьютере можно не только рассчитать оптимальную диету для спортсмена или выбрать наиболее выгодную комбинацию обхода противника на футбольном поле, но и смоделировать движение по трассе для гонок «Формула-1» в реальном времени.

За последние десять лет использование компьютеров для тренировки

спортсменов стало привычным. Например, Хуан Пабло Монтойа, готовясь к дебюту в «Формуле-1», знакомился с трассой гонок по игре «Grand Prix».

Такое распространение компьютеров в качестве тренажёров обусловлено развитием технологии трёхмерной графики, увеличением производительности компьютеров, снижением затрат на создание программ за счёт появления мощных средств разработки программного обеспечения.

Главная задача тренажёра — *симуляция,* т. е. имитация ситуаций реальной жизни, будь то движение по гоночной трассе, спуск с крутой вершины на лыжах или движение на яхте по морю.

Программа, которая отвечает за прорисовывание ландшафта или дороги, должна максимально точно передавать реалии той или иной географической точки, учитывая и такие факторы, как скорость ветра, освещённость, звуковые воздействия. Также неплохо бы показать тренирующемуся две-три нештатные ситуации, из тех, что случаются на соревнованиях, чтобы оценить его реакцию и впоследствии рекомендовать нужные модели поведения.

Конечно же только программой, какой бы сложной и универсальной она ни была, не обойтись. Для более полной симуляции окружающего мира требуются технические средства, относящиеся уже не к программному обеспечению, а к аппаратуре.

Попробуем определить, что нужно для создания автотренажёра кроме программы для симуляции автогонок. Ведь и в реальной жизни далеко не уедешь, если просто нажимать кнопки.

РУЛЬ С ОБРАТНОЙ СВЯЗЬЮ

В 90-х гг. XX в. в залах игровых автоматов можно было вести машинку, движущуюся по гоночной трассе. В качестве «устройства ввода» использовался руль, похожий на настоящий, но поворачивался он без каких-либо усилий, поэтому особого эффекта от такой «езды» не чувствовалось. Прошло несколько лет. Конструкторы, занимающиеся разработками игровых манипуляторов, придумали замечательное изделие — манипуляторы с обратной связью. Руль, джойстик (рукоятка для управления игрой), GamePad (специальное устройство для видеоигр, которое держат двумя руками; под пальцами одной находятся кнопки «огня», под другой — маленькая рукоятка для управления движением) теперь вели себя в зависимости от условий игры. Например, когда игрок выходил из крутого виража, руль про-

ворачивался с заметным усилием. То есть достигался совершенно потрясающий эффект присутствия.

Внутри устройства с обратной связью находятся довольно сложные компоненты: сервомоторы, виброприводы, электромагниты. Всем этим управляют одна или несколько небольших микросхем-контроллеров, принимающих сигналы от ЭВМ, где работает основная программа. Контроллер включает и выключает компоненты, позволяющие ощутить тяжесть штурвала или испытать, как ведёт себя машина на повороте.

Кроме руля в настоящей машине есть ещё масса других устройств, которые можно симулировать для создания тренажёра, например педали газа и тормоза, рычаг переключения передач.

В принципе такого оборудования уже вполне достаточно (пожалуй, ещё бы монитор побольше), чтобы почувствовать себя настоящим пилотом гоночного болида. Многоканальный звук и эффекты визга тормозов и рёва двигателя довершат картину.

Подобный тренажёр вполне можно собрать и дома, а в профессиональных тренажёрах обычно используется несколько мониторов для наблюдения за окружающим пространством впереди, по бокам и сзади. Помимо этого, кресло «пилота» точно повторяет оригинальное, со всеми неудобствами и нюансами.

Возможно, для проведения гонок «Формула-1» в будущем не понадобятся дорогостоящие трассы. И это не фантастика. Действительно, существуют соревнования по гребле в зале, когда спортсмены гребут, сидя на одном месте, перегоняя вёслами воду из бассейна в бассейн, а зрители и судьи наблюдают за результатами «гонок» на экранах мониторов. Поверьте, от этого соревнования не стали менее зрелищными или напряжёнными.

По сообщениям прессы, у 13-летнего мальчика врачи обнаружили признаки профессионального заболевания рабочих, имеющих дело с отбойными молотками. Это объяснялось тем, что подросток много времени (по шесть-семь часов в день) уделял видеоиграм, в которых использовался джойстик с обратной связью.

ИСКУССТВЕННЫЕ РЕАЛЬНОСТИ

«...Только вчера, только вчера я открыл эту сумасшедшую планету, находящуюся в миллионе световых лет от дома. Я выбивался из последних сил, ругал себя, что пошёл обследовать планету, не взяв с собой робота. По инструкции, действующей на всём звёздном флоте Земли, я не имел права даже шагу ступить без него на новой планете... Стояла страшная жара, задыхаясь, я бежал по пыльной, выжженной жарким солнцем пустыне. Липкий пот норовил попасть в глаза, я смахнул прилипшие ко лбу спутавшиеся волосы. Колени и руки разбиты в кровь. Оглушительный рёв, от которого кровь стыла в жилах и страх парализовывал тело, снова прервал мои мысли. Только нечеловеческим усилием воли мне удалось заставить ноги двигаться. До моего корабля оставалась пара сотен метров, там спасительное

укрытие и оружие, способное сразить любое чудовище. Но по тому, как всё сильнее вибрировала земля, я всё отчётливее понимал: шансов у меня нет... Я побежал.

Меня пронзила дикая, всепоглощающая боль, спину жгло огнём и сводило судорогой, ноги подкосились, и, теряя сознание, я понял, что на этот раз чудовищу опять удалось меня победить...

...Это был не сон. Когда боль стала отступать, продолжая тяжело дышать, я увидел выросшую перед глазами надпись: "Игра окончена". Тяжело вздохнув, сняв шлем, я с сожалением начал стягивать комбинезон. Тело продолжало гудеть и болеть, дыхание всё ещё было неровным, когда, открыв дверь, я покинул кабину Виртуальной реальности...»

Термин «виртуальность» использовался в разных науках и ранее, в частности в качестве обозначения мнимости некоторых объектов в физике; он также применяется для трёхмерных макромоделей, реализованных с помощью компьютера.

Английское словосочетание virtual reality переводится как «фактическая действительность». Но в русском восприятии термин «виртуальная реальность» соотносится с реальностью, искусственно созданной и очень похожей на настоящую (вопреки значению слова «виртуальный» — «возможный», которое пришло в русский язык из латыни).

Согласно сложившемуся стереотипу, виртуальная реальность — что-то связанное с компьютерными играми и фантастическими фильмами. Отчасти это так, однако учёные посчитали, что виртуальной реальности найдётся место и в настоящей жизни с учётом её способности обрабатывать очень большие объёмы данных и делать «видимыми» результаты научных исследований.

Смоделированный объект можно помещать в различные среды, не только имитировать и прослеживать его перемещения в вымышленном пространстве, но и демонстрировать его функционирование. Компьютерное моделирование позволяет создать и усовершенствовать сложное изделие, оценить и опробовать его не на реальном предприятии, а в виртуальной среде. Это особенно актуально для дорогостоящих, сложных, уникальных комплексов.

В настоящее время всё шире применяются технологии виртуального моделирования, т. е. процесса создания виртуальной модели объекта, предназначенного для последующего производства, его оценки на этапе виртуального прототипа (свойств безопасности, функциональности и т. д.), оптимизации процессов его изготовления. Лишь после получения удовлетворительных результатов принимается решение об изготовлении физического объекта. Это необходимо для того, чтобы на научные и инженерные разработки затрачивалось меньше времени и средств. Тогда путь от лаборатории до производственного цеха сократится, а все аспекты внешнего вида будут определяться на компьютерных, а не на готовых моделях.

Существует два основных типа виртуальной реальности — сценическая и экранная (*англ.* scenic и screen). Сценический тип — человек находится как бы внутри виртуального мира, является одним из его объектов. Экранный тип — человек отделён от виртуальной среды, наблюдает её извне через неподвижное поле зрения (как, например, сидящий за дисплеем компьютера).

Экранные системы обеспечивают частичное погружение в виртуальную

среду, т. е. часть среды моделируется на компьютере, а часть имитируется физически — в виде различных пультов, рабочих мест, кабин.

Сценическая виртуальная реальность позволяет достичь практически полного погружения в виртуальную среду. Общение с окружающим миром у человека происходит с помощью пяти основных чувств: зрения, слуха, осязания, обоняния, вкуса. При подмене источников информации для некоторых или всех органов чувств возникает иллюзия присутствия в другом мире. Для формирования такого типа реальности необходимы особые технические средства: системы отображения различных видов, шлемы и очки (обеспечивают неограниченное и подвижное поле зрения), разнообразные датчики (контролируют положение тела в пространстве и тактильную связь с несуществующими объектами). Существенными элементами являются системы воспроизведения пространственного звука. Так, на поверхности виртуального космического аппарата человек будет слышать работу машин, звуки собственного дыхания, живую речь коллег и т. д.

Сложилось мнение, что шлем плюс перчатки — это необходимые и достаточные атрибуты виртуальной реальности, с помощью которых можно «войти» в искусственные миры, созданные с помощью компьютера. На самом деле моделировать виртуальные миры чрезвычайно трудно. Ещё сложнее осуществить «подмену» информации для наших органов чувств. Например, зрение. Мы видим мир с помощью сетчатки глаз. Учёные подсчитали, что разрешение сетчатки приблизительно 8000 × 8000 точек. Чтобы получить такую картинку на компьютере при глубине цвета 24 бит, нужно использовать как минимум 384 Мбайт видеопамяти, а пропускную способность шины иметь 11,5 Гбайт/с для 30 кадров в секунду. И это только для изображения, а ещё есть слух, осязание, «обман» которых требует не меньших ресурсов. Кроме перечисленного существует огромное количество проблем в создании виртуального зрения:

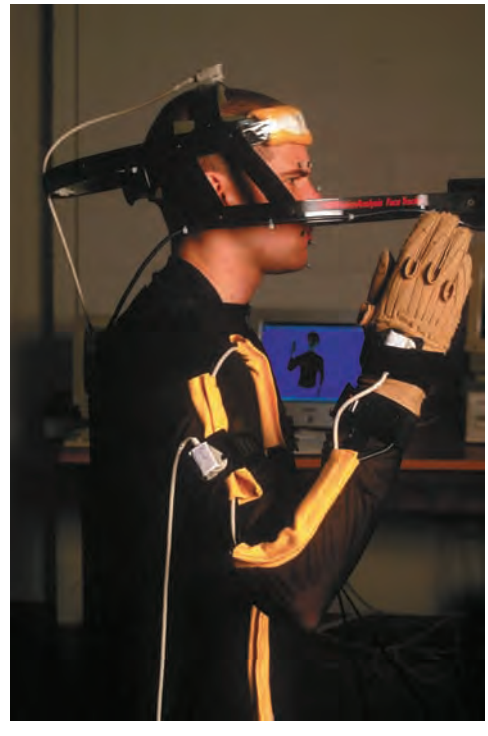
• Картинку нужно подавать прямо на сетчатку, а для этого требуется отслеживать перемещение не только головы, но и зрачков (возможно, даже фокусировку хрусталика). Такой механизм позволит создавать настоящее объёмное изображение, а не простое стереоскопическое.

• Изображение, которое надо построить, полусферическое, причём с переменным разрешением.

• С помощью 24-битного цвета нельзя смоделировать тусклый свет звёзд и ослепительное солнце.

• Для хранения и обработки 3D-моделей нужны процессор с высокой вычислительной мощностью и большая оперативная память.

Простой способ передачи картинки виртуального мира — поочерёдно формировать изображение на дисплее для правого и левого глаза. Для этого необходимо в тот момент, когда на экране появляется нужная картинка, синхронизировать изображение с устройством, которое поочерёдно заслоняет то один, то другой глаз. Чтобы глаза не уставали, частота должна быть 85—100 Гц, для стереоизображения в два раза больше (150—300 Гц).



Другой способ более дорогой — непосредственное построение изображений для каждого глаза на цветных матрицах или отдельном маленьком мониторе, в шлемах или очках виртуальной реальности. В настоящее время довольно успешно развивается имитация осязания. Специальная мышь позволяет почувствовать выпуклость кнопки, проведя по ней курсором, ощутить сопротивление, с которым изменяется размер окна, почувствовать рельеф картинки в графическом редакторе. Джойстики, поддерживающие технологию Force Feedback (силовая обратная связь), заставят, например, при игре в гоночный симулятор «трястись» в соответствии с ухабами на дороге и испытывать трудности при повороте на большой скорости. А играя в «стрелялку», получать отдачу при собственных выстрелах или при попадании пуль в вас. Если зрение, звук, осязание уже неплохо имитируют, хотя и здесь ещё далеко до совершенства, то с другими чувствами намного сложнее. Пока нет ни «запаховых», ни «вкусовых» карточек. Идеальной виртуальной реальности достигнут тогда, когда будет возможно «подключиться» непосредственно к зрительным, обонятельным, осязательным, слуховым окончаниям, которые находятся в спинном и головном мозге.

Тем не менее уже сейчас широко используются достижения в области виртуальной реальности: разработка проектов новых сооружений, медицинское исследование человеческого организма, моделирование интерьера квартир и другие области, где необходимо визуальное представление связанных в систему трёхмерных объектов. Не за горами и такое: хирург, стоя посреди кабинета, делает сложнейшую операцию, но не на человеке, а на его виртуальной модели. Врач отлично чувствует тело и ткани. А всю работу выполняет робот, управляемый компьютером с помощью устройства с обратной связью, которым в свою очередь управляет врач. Естественно, что для медицины подобные устройства должны обеспечивать очень точные и реалистичные ощущения. Хирург может увеличивать изображение любой части тела, рас-

сматривать его в нужном ракурсе и работать, не боясь, что случайно дрогнет рука со скальпелем. При участии микророботов он способен провести операцию прямо внутри человека, например на сердце. Кроме того, такая операция может проводиться на значительном удалении врача и пациента друг от друга, скажем через Интернет: врач находится в России, а пациент — где-нибудь в экспедиции на Северном полюсе. Но для такой деятельности, конечно, нужна очень надёжная связь. Естественно, что достижения в виртуальной реальности используются не только в медицине, но и в местах, где на человека действуют вредные излучения, агрессивная среда, космос и т. д. Эти технологии уже успешно применяются в тренажёрах-имитаторах (автомобилей, самолётов, кораблей, космических аппаратов) и, конечно же, в индустрии развлечений.

Виртуальная реальность — искусственно созданный мир со своими законами и объектами. Вымышленные миры представляют собой плоды человеческого воображения и возникают необязательно посредством компьютера. Пример тому — книги и кинофильмы. Виртуальная реальность давно описана многими писателями-фантастами и учёными. Некоторые из их героев — творцы, создающие виртуальный мир, другие с помощью виртуальной реальности приобретают паранормальные способности, третьи живут в виртуальном пространстве.

В фильме «Газонокосильщик» главный герой, отстающий от сверстников в развитии, с помощью виртуальной реальности приобретает интеллект гения и вместе с тем жестокость маньяка. Фильм «Матрица» рассказывает о том, как компьютеры моделируют мир виртуальной реальности, так что зритель начинает путаться, где мир реальный, а где виртуальный. В фильме «Странный Тома» главный герой просто живёт в виртуальном мире. В силу специфики своей болезни он совсем не выходит из дому, и всё общение с внешним миром происходит только через Интернет. Не является ли это моделью мира будущего?

«... Я с сожалением оглянулся на стоящие в ряд кабины, в одну из них заходил игрок... Я подошёл к автомату, стоящему неподалёку, проверить свою кредитную карту. Автомат проглотил мою кредитку, помигал глазами, погудел, вывел на дисплее остаток на моём счёте и выплюнул

карточку обратно. Я задумался, прикидывая в уме, на сколько мне хватит денег и когда я смогу вновь пополнить счёт... Играть хотелось до дрожи в коленях... Закончив свои подсчёты и придя к малоутешительным выводам, я ещё раз посмотрел на манящие к себе кабины. Глубоко вздохнув, махнув рукой на свои невесёлые мысли, я решительно шагнул в сторону кабины Виртуальной реальности... на моём лице сияла блаженная улыбка...»

ОТ ИГР — К ЗНАНИЯМ

Компьютерные игры, увлекая всё большее число людей, превращаются в своеобразную манию. Играют не только дети, но, что самое печальное, играют и взрослые. И играют, как правило, в рабочее время. С появлением компьютерных игр компьютерные учёные мужи стали тратить на них массу времени. Случайно запертый после работы в одном НИИ научный сотрудник, отец двоих детей, провёл выходные, непрерывно играя, при этом ночами не смыкал глаз. Очевидцы рассказывают, что, когда в понедельник открыли дверь, их коллега выглядел уставшим и был покрыт трёхдневной щетиной, но глаза его светились от счастья.

Запретить игры, конечно, нельзя. Сейчас персональные компьютеры распространены повсеместно, и, наверное, нет ребёнка, который не хотел бы играть в компьютерные игры.

На одном из компьютерных вечеров, проводившихся журналом «Микропроцессорные средства и системы» в Москве в начале 90-х гг. XX в., прозвучал лозунг: «Учись, играя!». Учёные пришли к выводу, что тренажёры, компьютерные учебники, практикумы для школьников по различным дисциплинам, если их делать с элементами игры, существенно улучшат восприятие и запоминание материала учащимися.

Но компьютеры тех лет были ещё недостаточно оснащены, чтобы воплотить подобные замыслы в жизнь. Для этого требовались значительные усовершенствования. Должно было произойти, по крайней мере, два революционных изменения в компьютерах.

Первый революционный шаг —появление *мультимедиа* (Multimedia): к нам пришли, звук, музыка, качество изображения на мониторе стало не хуже телевизионного.

Второй — создание *компьютерных* CD (Compact Disk), компактных, недорогих носителей с большим объёмом информации. Один такой диск вмещает порядка 300 тыс. страниц текста (а на десяти CD умещается результат работы всех математиков мира за год!), около 80 мин цифрового звука, тысячи высококачественных снимков, фильмы телевизионного качества и т. д.

Обычные *информационные* CD, по сути не явились новацией, ведь на школьных уроках и прежде использовалось телевидение, а учебники содержали цветные фотографии.

Зато *гипертекстовые энциклопедии* сродни играм-*квестам,* увлекательным приключениям. Они не только содержат подробный справочник

по изучаемому материалу и информацию в виде красочных фотографий и фрагментов фильмов, но и дают возможность объёмного восприятия изображения с помощью специальных 3D-очков. Благодаря этому ученик вовлекается в игру, во время которой гуляет по зоопарку или музею, спасается от разъярённого тираннозавра, путешествует во времени.

Контрольные вопросы предлагаются также в игровой форме: это может быть поиск конкретных предметов или определение животных по голосам.

Тем не менее энциклопедия остаётся энциклопедией: аналог библиотечного зала с каталогом позволяет непосредственно искать по названию предмета, например всю информацию о буйволах. Каталог фотографий, статей и фрагментов фильмов также удобен в процессе поиска интересующего материала. Если потребуется найти какой-либо дополнительный материал для изучения темы, школьник ещё не раз обратится именно к энциклопедии.

Для самых маленьких создаются живые книги, которые переносят детей в увлекательную игровую среду. Занимательно путешествие по загадочной стране, где ребёнок с помощью мышки всё потрогает на экране. При этом предметы оживают, кролик разговаривает, телевизор легко включить или переключить программу с помощью пульта, по телефону можно позвонить, доска для рисования предлагает порисовать вволю или раскрасить животных, часы скажут, который час... В ненавязчивой форме

происходит процесс обучения. Определи по часам время, узнай, какой музыкальный инструмент как звучит, поговори вежливо с кроликом и т. д. Существуют живые книги и по классическим детским литературным произведениям, например «Винни Пух и все-все-все». И конечно, в рассказ включены игры и с самым забавным медвежонком, и с Кроликом, и со всеми друзьями Пуха.

Компьютер предоставил возможность осуществить мощный прорыв в образовании. Уникальная подача материала и обилие самостоятельных игровых занятий для детей позволят ещё глубже и многограннее понять предмет. Важно только, чтобы в игре Геракл не участвовал во Второй мировой войне, как это порой бывает в телесериалах.

ИСТОРИЯ ДОКОМПЬЮТЕРНОЙ ЭПОХИ

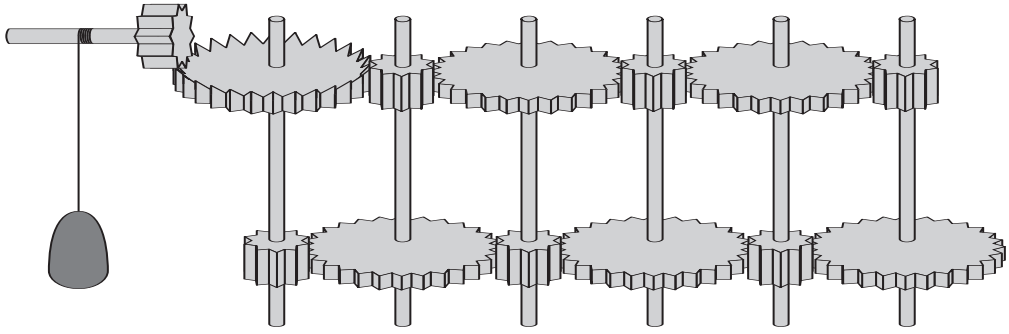
НА ПУТИ К СЧЁТНОЙ МАШИНЕ

Леонардо да Винчи.

На простой вопрос, какой компьютер был первым, можно получить достаточно странный ответ: «IBM PC XT», однако история компьютеров значительно богаче, и первый компьютер, созданный человеческим гением, конечно же, не IBM PC XT, это всего лишь первый персональный компьютер, выпущенный фирмой IBM.

Первоначально компьютер (от *англ.* computer — «вычислитель», «расчётчик») предназначался для выполнения сложных однообразных вычислений. История современной вычислительной техники насчитывает чуть более полувека, но первым механическим компьютером считают абак (который иногда ассоциируется с обычными счётами); упоминание о нём встречается ещё до новой эры. Этот «компьютер» имел память — костяшки на счётах, а для сложения и вычитания «использовался» человек, он производил вычисления, двигая костяшки. Поразрядное сложение, как сложение в столбик, на счётах выполнять удобнее, чем в уме, поэтому они не утратили своей вычислительной «мощи» и по сей день.

В XX столетии были найдены записки великого учёного, блестящего изобретателя и живописца Леонардо да Винчи (1452—1519), где говорится о механической счётной машине. Принцип её работы напоминал устройство одометра (прибор в автомобиле для подсчёта пройденного пути), в котором если покрутить колёсико километров в сторону увеличения, то после 9 будет 0, зато колёсико десятков повернётся на 1, т. е. счётчик покажет 10. Столь «сложный» прибор имеет много общего с первыми арифмометрами, где число кодировалось положением (поворотом) диска с десятью зубцами.



Механизм калькулятора Леонардо да Винчи.

С помощью таких связанных дисков можно построить суммирующее устройство.

В 1642—1643 гг. выдающийся французский философ и математик Блез Паскаль (1623—1662) изобрёл и сконструировал первое механическое счётное устройство, позволяющее складывать в десятичной системе счисления.

Механический сумматор осуществлял сложение чисел на специальных дисках-колёсиках. В машине Паскаля десятичные цифры пятизначного числа задавались поворотами дисков, на которых были нанесены цифровые деления. Результат читался в окошечках. Диски имели один удлинённый зуб, чтобы при сложении можно было учесть перенос единицы в следующий десятичный разряд. Диски были механически связаны через промежуточные шестерни: диски сотен, десятков и единиц вращались в одну сторону — в сторону увеличения. Отклонение специального рычага на определённый угол через храповой механизм позволяло «вводить» однозначное число и суммировать его с числом, хранящимся в калькуляторе. Такой привод был у каждого диска, что позволяло суммировать многозначные числа. В первом калькуляторе Паскаля было пять цифр, затем он увеличил их количество до восьми.

Если счёты — запоминающие устройства, лишь хранящие информацию, тогда как все действия производит человек, то суммирующая машина Паскаля — механический вычислитель, который автоматически выполняет переносы единиц в следующий десятичный разряд.

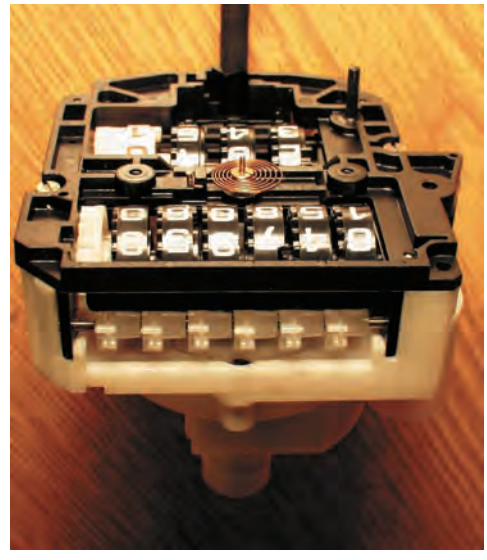
Однако механический сумматор мог только складывать, для вычитаний использовалась техника «вычитания через дополнение». Например, чтобы из 4125 вычесть 737 на пятизначной машине Паскаля, надо к 4125 прибавить дополнение 737 до 100 000, что, как ни странно, легко подсчитать (все цифры, кроме самого младшего разряда, получаются дополнением до 9). Оно равно 99 263:

$$4125 - 737 = 3388,$$

$$4125 + 99\,263 = 103\,388.$$

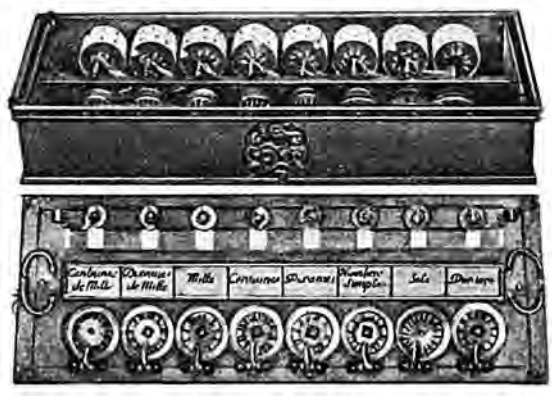
А так как шестого разряда в пятиразрядной машине нет, то результат не 103 388, а 3388.

Ещё одна машина была сконструирована немецким философом, математиком и физиком Готфридом Вильгельмом Лейбницем (1646—1716).



Одометр автомобиля «фольксваген».

Счётная машина Паскаля.

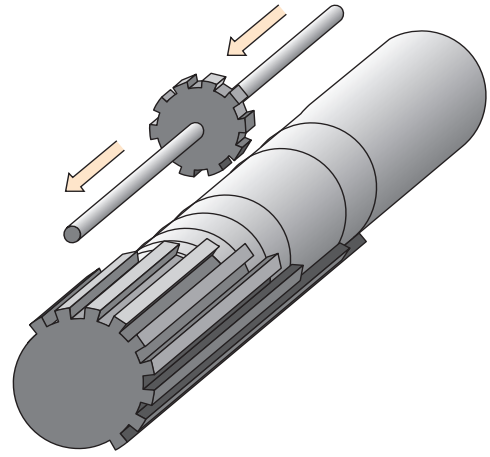


Счётная машина Лейбница.

В 1673 г. Лейбниц изобрёл «компьютер», который не только складывал, но и умножал. Машина Лейбница выполняла сложение практически тем же способом, что и суммирующая машина Паскаля, но в её конструкцию были включены движущаяся часть (подвижная каретка) и ручка, с помощью которой крутились специальное колесо или (в более поздних вариантах) барабаны, расположенные внутри аппарата.

В машине каждый разряд имел собственный механизмам, связанный с механизмами соседних разрядов. Лейбниц использовал *шаговые барабаны* (ступенчатые валики)— цилиндры с девятью зубцами разной длины (длина зубца увеличивалась по возрастающей). Когда барабан поворачивался, связанное с ним передаточное колесо с десятью зубцами поворачивалось от 0 до 9 в зависимости от его позиции по отношению к барабану (колесо могло перемещаться по оси вдоль шагового барабана). Так Лейбниц использовал операцию «сдвига» для поразрядного умножения чисел. Данный метод лёг в основу всех механических калькуляторов последующих веков.

Конечно, по сегодняшним меркам эти машины нельзя назвать настоящими компьютерами (и даже калькуляторами). История механических компьютеров — это история цифрового колеса и привода.



▶ Шаговый барабан счётной машины Лейбница.

АБАК И СЧЁТЫ

Абак — первое счётное приспособление, которое стал применять человек. Идея его устройства заключается в наличии специального *вычислительного поля,* где по определённым правилам перемещают *счётные элементы,* сгруппированные по разрядам. Именно эта идея объединяет столь разные на первый взгляд приборы, как греческий и римский абак, китайские и русские счёты, а также счёт на линиях.

Скорее всего, сначала роль абака (полагают, что корень этого греческого слова означает «пыль») выполняла покрытая пылью или песком доска, на которой можно было чертить линии и перекладывать камешки. Абак был известен ещё в Древнем Египте и, возможно, в Вавилонии, а финикийские купцы завезли его в Грецию.

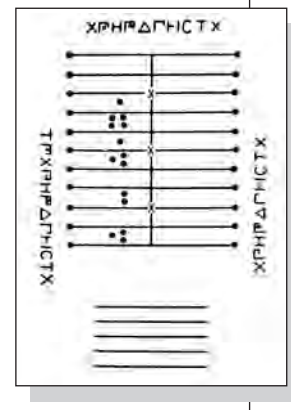
Древнегреческий историк Геродот (между 490 и 480 — около 425 до н. э.) свидетельствует, что «эллины пишут свои буквы и считают слева направо, а египтяне — справа налево». Это первое письменное упоми-

нание об абаке (несомненно, речь здесь идёт именно о нём, поскольку других способов счёта тогда просто не существовало). А вот самое раннее изображение абака можно увидеть на вазе, изготовленной греческими мастерами в III в. до н. э. На ней рядом с персидским царём Дарием I находится казначей, занятый подсчётами на абаке. О том, что изначально данное устройство служило именно для выполнения денежных расчётов, свидетельствует и единственный сохранившийся до наших дней древнегреческий абак — так называемая саламинская плита, найденная при раскопках в 1846 г. Это действительно большая (размером 105×75 см) мраморная плита, на которой прорезаны параллельные линии, образующие несколько колонок. В левой колонке подсчитывали крупные денежные единицы — драхмы и таланты (так что вычисления на греческом абаке велись над именованными числами). Здесь использовалась двоично-пятеричная система счисления: значениями разрядов были 1, 5, 10, 50, 100, 500, 1000, 5000 драхм, 1 и 5 талантов. В правой части считали «мелочь» — халки и оболы, причём значение каждого следующего разряда удваивалось: 1, 2, 4 и 8 халков (что равно 1 оболу). «Стоимость» разрядов обозначена высеченными рядом с ними греческими буквами. Не случайно Полибий (II в. до н. э.), автор «Истории», уподобляет придворных камешкам на счётной доске, цена которых — один халк или целый талант — полностью в руках счётчика. На такой доске легко складывать и вычитать, добавляя или убирая камешки и перенося их из разряда в разряд.

Хотя в Древнем Риме абак и называли calculi — «камешки» (отсюда произошел латинский глагол calculare — «вычислять», а от него — русское слово «калькулятор»), камешки уже не использовали. Абак изменился, превратившись в настоящий счётный прибор. Изготовляли его римляне из бронзы, цветного стекла или слоновой кости в виде доски с двумя рядами прорезей, по которым передвигались косточки (в нижнем ряду по четыре, а в верхнем — по одной). Косточка верхнего ряда «стоила» в пять раз больше, чем расположенные под ней (например, число 80 представлялось одной косточкой верхнего и тремя косточками нижнего ряда колонки десятков: 50 + 3·10). Несколько рядов прорезей служили для представления дробей, кратных 1/12, 1/24, 1/48 и 1/72. В работе римский абак был значительно удобнее греческого.

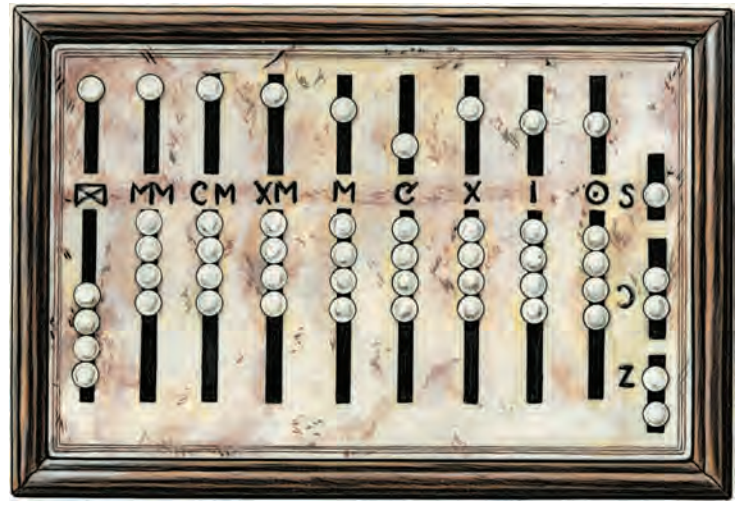
Только в X в., после нескольких столетий упадка науки и культуры, наступивших вслед за падением Римской империи, абак снова распространяется в Европе. Его возрождение связано с именем одного из самых ярких и образованных людей раннего Средневековья Герберта из Орийака (около 940—1003), ставшего в 999 г. Папой Римским Сильвестром II. Вспомним, как на первых страницах романа М. А. Булгакова «Мастер и Маргарита», в сцене на Патриарших прудах, Воланд говорит: «Тут в государственной библиотеке обнаружены подлинные рукописи чернокнижника Герберта Аврилакского, десятого века. Так вот требуется, чтобы я их разобрал. Я единственный в мире специалист». Вполне возможно, что он имел в виду именно рукописи математического содержания, ведь самое известное сочинение Герберта называется «О правилах абака».

Абак Герберта содержал 27 колонок, объединённых по три, и три дополнительные колонки для



Греческий абак.

Римский абак.

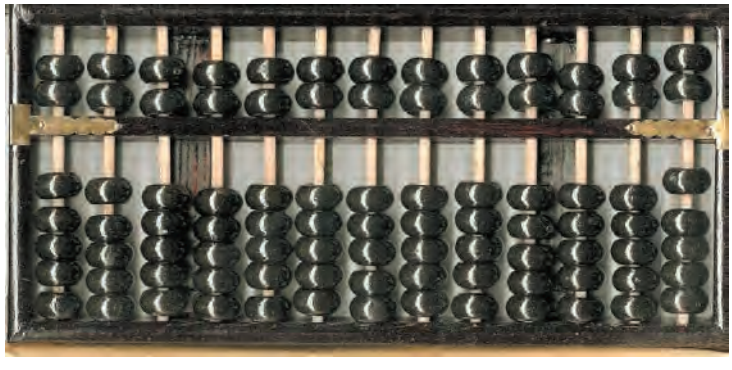


Абак Герберта.

представления дробей. Вместо камешков использовались специальные жетоны с нанесёнными на них цифрами от 1 до 9 (первоначально это ещё не были арабские цифры). Герберт не только усовершенствовал прибор, но и создал целую систему правил, с помощью которых производились весьма сложные вычисления.

Абаком пользовались все: и купцы, и менялы, и ремесленники. К абацистам (в честь Герберта из Орийака их ещё называли гербекистами) принадлежали многие видные учёные. Даже спустя шесть столетий изобретение Герберта оставалось важнейшим инструментом для практических вычислений, не случайно Блез Паскаль писал, что использование его арифметической машины позволит отказаться от утомительных расчётов «с помощью пера и жетонов». А вот создатель другой вычислительной машины — Лейбниц предпочитал жетоны вычислениям на бумаге. В Западной Европе учителей арифметики даже в XVII в. продолжали именовать магистрами абака, как, например, известного математика Никколо Тарталью (около 1499—1557).

Старинные китайские счёты.



Правда, в XII—XIII вв. абак принял форму так называемого счёта на линиях. В нём использовались специальные разлинованные таблицы и жетоны, которые можно было помещать как на линиях, так и между ними. До конца XVIII в. счёт на линиях сохранял свои позиции в некоторых европейских странах и лишь затем окончательно уступил место вычислениям на бумаге.

С IV в. до н. э. абак известен в Китае — в то время использовались счётные палочки, которые выкладывали на специальной доске. Постепенно их сменили разноцветные фишки, доска приобрела иную форму, а в X в. появились китайские счёты — суан-пан. В них место доски заняла рама с нанизанными на прутья косточками (по семь на каждом). Она разделена на две неравные части: в одной — по пять косточек (единицы), а во второй — по две (пятёрки). Из Китая суан-пан в XVI в. пришёл в Японию — здесь он получил название «соробан». Рама соробана тоже состоит из двух частей, но они содержат по одной и по четыре косточки.

Почти одновременно счёты появились в Японии и в России, поэтому бытует мнение, что и к нам их завезли из Китая. Однако это не так. Русские счёты очень сильно отличались от китайских, и сейчас большинство исследователей полагают, что в России они были изобретены самостоятельно. Тем более что возникло это изобретение не на пустом месте.

В России с древних времён был распространён «счёт костьми», близкий европейскому счёту на линиях. Вместо жетонов обычно применялись плодовые косточки (посетивший Россию в 1634—1636 гг. немецкий учёный и путешественник Адам Олеарий отмечал, что писцы для этой цели имеют при себе мешочки с косточками сливы). В XVI в. возник так назы-

ваемый дощаной счёт, первый вариант русских счётов (спустя 100 лет появилось и само слово «счёты»). Их устройство было достаточно сложным: целых четыре счётных поля для рядов по десять косточек. К середине XVII в. относятся хранящиеся в Историческом музее в Москве самые старые из дошедших до нас счётов. Они представляют собой две соединённые рамы, точнее, ящика, каждый из которых разделён на три счётных поля. В ящиках имеется по одному полю, содержащему полные ряды по десять

КТО БЫЛ ПЕРВЫМ?

Более трёх столетий приоритет в изобретении счётной машины безоговорочно признавался за Паскалем. Историкам, конечно, было известно письмо от 20 сентября 1623 г., адресованное знаменитому математику и астроному Иоганну Кеплеру, в котором профессор Тюбингенского университета (в настоящее время эта земля принадлежит Германии) Вильгельм Шиккард сообщал, что сконструировал машину для автоматического производства вычислений. Однако других тому подтверждений не имелось, и долгие годы к этим сведениям относились с недоверием.

В своё время рукописи Кеплера были куплены императрицей Екатериной II и сохранись в архиве Российской академии наук. Уже после Второй мировой войны, в 1957 г., изучая фотокопии этих документов, директор Кеплеровского научного центра, расположенного в Штутгарте, Франц Гаммер обнаружил набросок чертежа некоего механизма, похожего на счётное устройство. Кропотливый и целенаправленный поиск позволил исследователю найти в письме Шиккарда Кеплеру, написанном 25 февраля 1624 г., подробное описание внешнего вида созданной им вычислительной машины.

Из описания можно сделать вывод, что шестиразрядный десятичный вычислитель состоял из десятизубцовых и однозубцовых колёс и был рассчитан на выполнение сложения, вычитания, а также табличного умножения и деления. Шиккарду удалось решить проблему переноса между разрядами: машина накапливала и переносила влево десятки или сотни. При этом механизм передачи десятков был реверсивным, так что вычитание выполнялось вращением установочных колёс в обратном направлении.

В том же письме Шиккард извещал своего корреспондента, что вместе с механиком Вильгельмом Пфистером построил два экземпляра машины (он называет её часами для счёта), которые, к сожалению, сгорели во время большого пожара. Профессор писал о том, как тяжело он переживает эту потерю, и жаловался на отсутствие времени для постройки новой машины (кстати, одно из сгоревших устройств предназначалось в подарок Кеплеру). Но, судя по всему, к работе над вычислителем Шиккард так и не вернулся, хотя он умер в 1635 г. (во время эпидемии холеры).

Научному сообществу той поры не пришлось увидеть счётную машину Шиккарда в действии, но сомнения в её существовании, скорее всего, следует отклонить. Во-первых, трудно понять, зачем бы Шиккард стал вводить Кеплера в заблуждение (тогда требования научной этики соблюдались весьма строго!), а во-вторых, в его пользу убедительно говорят найденные позднее эскизы других чертежей вычислителя и письменные инструкции для механика. Более того, в 60-х гг. XX столетия учёным Тюбингенского университета по этим чертежам и описанию удалось построить действующую модель машины своего великого земляка.

После этого казалось, что историки смело могут увенчать вновь обретённого основоположника лаврами. Однако в 1967 г. были обнаружены неизвестные записные книжки (или дневники) Леонардо да Винчи, где имелся эскизный набросок 13-разрядного суммирующего устройства, построенного на основе десятизубцовых колёс. Таким образом, великий Леонардо обдумывал идею создания вычислительной машины как минимум за 100 лет до Шиккарда! Точно датировать запись не удалось, хотя известно, что дневник был начат ещё до открытия Америки в 1492 г.

История раннего периода развития вычислительной техники до сих пор не написана до конца. Никто не знает, какие ещё интереснейшие открытия сделают учёные — историки и архивисты.

косточек, а также по два поля с рядами, содержащими от одной до четырёх косточек. Уже к началу XVIII в. русские счёты приобрели современный вид. Дальнейшие усовершенствования: изменение размеров, изгиба проволоки, формы рамы — проводились для удобства использования. Но были и попытки сделать счёты более мощным инструментом для вычислений (например, соединить их с разными числовыми таблицами) — последняя из них относится к 1921 г.

Почти 300 лет счёты имели в России широчайшее распространение. Смертельный удар нанесло им только появление дешёвых карманных электронных калькуляторов.

БЛЕЗ ПАСКАЛЬ

В математике издавна известна «улитка Паскаля» — одна из замечательных кривых четвёртого порядка. В этом названии увековечено имя Этьена Паскаля (1588—1651), в семье которого 19 июня 1623 г. родился сын, названный Блезом. Этьен Паскаль, принадлежавший к судейскому сословию, человек богатый и образованный, был неплохим математиком и даже переписывался со знаменитым французским учёным Пьером Ферма). После смерти жены он посвятил жизнь воспитанию детей. Блез с самого раннего возраста проявлял признаки несомненной гениальности. Он обладал феноменальной памятью, в четыре года умел читать и писать, а в десять лет, пытаясь понять, почему при ударе звенит фаянсовая тарелка, написал первую научную работу «Трактат о звуке». Через два года мальчик самостоятельно (отец считал, что сыну пока рано заниматься математикой, и прятал от него учебники) доказал теорему о сумме углов треугольника, а спустя ещё год его как равного приняли в кружок крупнейших парижских математиков.

Блез Паскаль.

В 1640 г. Этьен Паскаль получил назначение «интендантом полиции, юстиции и финансов» в Руан. Эта должность подразумевала и контроль за сбором налогов по всей провинции. Считается, что мысль об арифметической машине возникла у юного Блеза из-за желания помочь отцу в сложных расчётах, которые тот производил на бумаге и на счётной доске «с помощью пера и жетонов». Вполне возможно, что первый толчок к размышлениям дала ему мысль великого философа Рене Декарта о том, что некоторые умственные процессы по сути своей не отличаются от механических. Позднее и сам Паскаль писал: «Действия арифметической машины больше похожи на действия мыслящего существа, нежели животного, но у машины нет собственной воли, а у животных есть». Судя по всему, он первым поставил вопрос о соотношении искусственного и человеческого разума!

Так или иначе, к концу того же года у Блеза сформировалась главная идея конструкции будущей машины — автоматический перенос разряда. «...Каждое колесо... некоторого разряда, совершая движение на десять арифметических цифр, заставляет двигаться следующее только на одну цифру» — именно так будет звучать формула изо-

бретения в полученной позднее, 22 мая 1649 г., королевской привилегии (прообраз современного патента) на арифметическую машину. Она утверждала приоритет Блеза Паскаля в изобретении и закрепляла за ним право производить и продавать машины.

Первая модель была готова уже через несколько месяцев, но... не работала. Паскаль приступил к разработке второй модели, над которой трудился с перерывами до 1642 г. Именно этот год считается датой изобретения, хотя первый экземпляр машины закончили лишь спустя ещё три года.

Позднее Блез Паскаль писал, что он «не экономил ни времени, ни труда, ни средств». Изобретатель руководил мастерами, часто сам вытачивал детали на токарном станке, подбирал наиболее подходящие материалы для машины: эбеновое дерево, медь, слоновую кость. Паскаль пробовал разные конструкции, его машины различались не только размерами, но даже разрядностью.

Автор рассчитывал на коммерческий успех своего детища. Он развернул настоящую рекламную кампанию: опубликовал объявление, приглашая всех желающих воспользоваться арифметической машиной (фактически квартира его друга, математика Роберваля, где в установленные часы был открыт приём посетителей, стала первым в истории вычислительным центром). Интересно, что в этом объявлении Паскаль особенно подчёркивал прочность изделия. Он подверг машину суровому испытанию, провезя её в карете более 1100 км (идея тестирования компьютеров появилась очень давно!).

Паскаль показывал свою машину в салонах самых знатных людей королевства, а для всеобщего обозрения выставил её в Люксембургском дворце. Один экземпляр в 1652 г. он даже послал в подарок шведской королеве Христине. Но настоящего производства наладить так и не удалось. Всего с 1645 по 1653 г. было изготовлено около 50 арифметических машин (из них восемь сохранились до наших дней). Несколько из них Паскаль сумел продать, в частности одну — принцу Конде за 100 ливров. Однако создателю она обошлась вчетверо дороже (а себестоимость других экземпляров доходила до 600 ливров). Да и покупали их не для работы, а скорее как интересную игрушку. Паскаль некоторое время продолжал совершенствовать машину, пытался также создать устройство для извлечения квадратного корня, но после 1653 г. больше к этому не возвращался.

Восьмиразрядная машина Паскаля.

Возможно, он счёл поставленную перед собой задачу решённой, а возможно, просто не видел для неё дальнейших перспектив. Общество ещё не было готово к использованию его изобретения. Кроме того, 15 ноября 1654 г. произошло событие, резко изменившее дальнейшую жизнь учёного. В этот день Паскаль чудом избежал гибели: его карета едва не свалилась с моста в Сену. Пережив сильнейший душевный кризис, Паскаль увидел в случившемся перст Божий, напоминание о своём долге перед Творцом. С 1655 г. он отказался от светской жизни и вёл полумонашеское существование в обители Пор-Руаяль. Иногда он ещё возвращается к занятиям математикой, но в основном его дни проходят в написании философских и богословских сочинений и в оказании помощи бедным. Блез Паскаль скончался 19 августа 1662 г. в Париже. Ему было всего 39 лет.

В историю человечества Блез Паскаль вошёл как один из величайших гениев. Он стоял у истоков математического анализа и теории вероятностей. Выдающийся физик, создатель классической гидростатики, Паскаль также был прекрасным механиком и изобретателем, стремившимся воплотить свои открытия

в жизнь. Ему принадлежит идея гидравлического пресса. Он же первым в истории предложил организовать движение городского транспорта: большие кареты (омнибусы) должны были перевозить по установленным маршрутам пассажиров, причём весьма дёшево. Всем известная обычная тачка тоже была придумана им! А его роль как создателя первой арифметической машины отмечал «отец кибернетики» Норберт Винер.

Язык прозы Паскаля-писателя признан образцовым. («Каждая строка, вышедшая из-под его пера, — это драгоценный камень», — сказал один из его соотечественников.) «Мысли», главное сочинение Паскаля, восхищали своей глубиной Стендаля и Льва Толстого, Бодлера и Тургенева...

Через века прошли названия: «теорема Паскаля» — в проективной геометрии, «треугольник Паскаля» — вматематике, «закон Паскаля» — в гидростатике. Уже в наши дни имя учёного получили единица давления и один из самых популярных языков программирования.

ГОТФРИД ВИЛЬГЕЛЬМ ЛЕЙБНИЦ

О Лейбниц, о мудрец, создатель вещих книг!
Ты выше мира был, как древние пророки.
Твой век, дивясь тебе, пророчеств не постиг
И с лестью смешивал безумные упрёки.

В. Брюсов

Готфрид Вильгельм Лейбниц родился 1 июля 1646 г. в городе Лейпциге. Его отец был профессором этики, а дед — профессором права Лейпцигского университета. Мальчику не исполнилось и семи, когда он потерял отца. Готфрид с детства много занимался: в восемь лет он самостоятельно изучил греческий язык и латынь, а в пятнадцать — окончил гимназию.

Готфрид Вильгельм Лейбниц.

В 1661 г. Лейбниц стал студентом факультета права Лейпцигского университета. Кроме юриспруденции он изучал философию и математику в университетах Лейпцига, Йены и Альтдорфа. В 1666 г. 20-летний студент защитил сразу две диссертации: «Об искусстве комбинаторики» (Лейпцигский университет) и «О запутанных делах» (Альтдорфский университет).

В 1668 г. Лейбниц поступил на службу к курфюрсту Майнца. С 1672 по 1676 г. он находился в Париже с дипломатической миссией, что не помешало ему продолжать заниматься математикой и естествознанием. Здесь Лейбниц познакомился со многими учёными. Одним из них был известный голландский изобретатель и физик Христиан Гюйгенс (1629—1695). Видя, сколько вычис-

лений приходится тому делать и как много времени уходит на это, молодой человек решил облегчить его труд с помощью механического устройства для расчётов.

Сначала он хотел лишь улучшить машину великого Паскаля. Первое описание подобного «арифметического инструмента» Лейбниц сделал ещё в 1670 г. Через два года он предложил новое описание, по которому был изготовлен экземпляр, продемонстрированный в феврале 1673 г. на заседании Лондонского королевского общества. По словам самого учёного, он придумал арифмометр, чтобы надёжно и быстро механически выполнять все арифметические действия, особенно умножение. Однако признавал, что инструмент несовершенен, и обещал его улучшить. В 1676 г. Лейбниц привёз в Лондон новый вариант счётной машины, но и эта модель оказалась малопригодной для практических вычислений. Лейбниц неоднократно возвращался к своему арифмометру. Усовершенствованная машина появилась лишь в 1694 г. в Ганновере.

Лейбниц гордился своим изобретением. Вот что он писал: «Мне посчастливилось построить такую арифметическую машину, которая бесконечно отличается от машины Паскаля, так как моя машина даёт возможность совершать и умножение, и деление над огромными числами мгновенно, притом не прибегая к последовательному сложению и вычитанию». Работа над машиной обошлась ему в 24 тыс. талеров. Для сравнения можно сказать, что годовая зарплата министра равнялась 2 тыс. талеров. Несмотря на прогрессивность изобретения, счётная машина не получила широкого распространения, потому что в конце XVII — начале XVIII в. отсутствовал спрос на такую сложную и дорогую технику.

Один из экземпляров конструкции 1694 г. Лейбниц собирался подарить Петру I, но машина оказалась неисправной, а механик учёного не смог её починить в короткий срок. Существует и другая легенда: арифмометр всё-таки попал в руки Петра I, и тот подарил его китайскому императору, желая поразить техническими достижениями Европы.

В 1676 г. Лейбниц вернулся в Германию и поступил на службу к ганноверским герцогам. Выполняя их разнообразные поручения, он работал сначала придворным библиотекарем, а потом герцогским историографом и тайным советником юстиции.

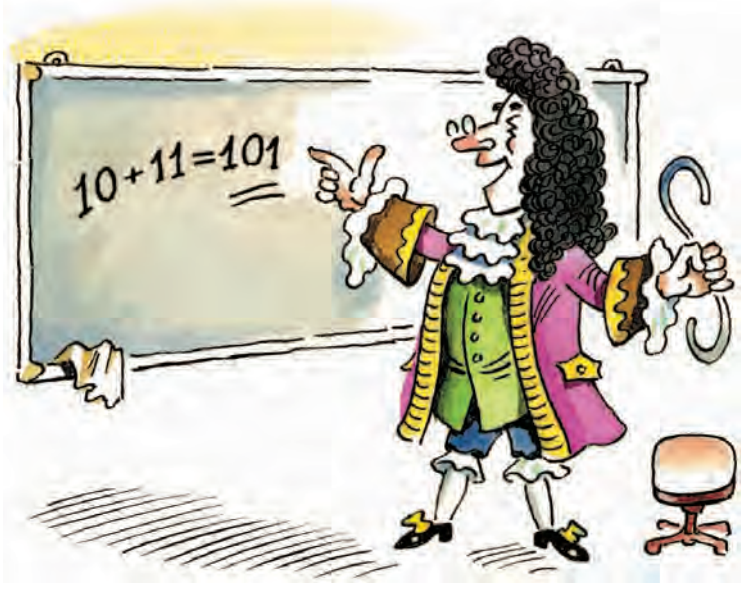
Однако деятельность Лейбница выходила далеко за пределы официальных обязанностей. Он находил время для проведения химических опытов, для медицинских исследований, изобретал различные устройства (в частности, сконструировал ветряной двигатель для насосов, выкачивающих воду из шахт), внёс существенный вклад в геологию, психологию, лингвистику и особенно в философию, физику, механику и математику.

Примечательным фактом является и то, что по просьбе Петра I учёный разработал проекты развития образования и государственного управления в России. «Я не принадлежу к числу тех, — писал Лейбниц русскому царю, — которые питают страсть к своему отечеству или к какой-либо другой нации, мои помыслы направлены на благо всего человеческого рода... и мне приятнее сделать много добра у русских, чем мало у немцев...»

Петр I.

Исаак Ньютон.

На протяжении всей жизни Лейбниц вёл активную переписку. После себя он оставил более 15 тыс. писем к тысяче адресатов на французском, немецком и латинском языках. Лейбниц вступал в переписку и с Ньютоном, но она продолжалась только год и не привела к сотрудничеству великих учёных (см. статью «Создание дифференциального и интегрального исчисления. Ньютон и Лейбниц» в томе «Математика» «Энциклопедии для детей»).



Современников Лейбница поражали его фантастическая эрудиция, почти сверхъестественная память и удивительная работоспособность, его широкий кругозор и диапазон деятельности, умение в любой проблеме увидеть и выделить сущность, основу. Лейбниц, как никто другой из современников, обладал предельной лаконичностью и точностью стиля, творческой энергией и способностью к обобщению.

Потребность обобщения заставляла его искать универсальный метод научного познания. По мнению Лейбница, мир создан разумом Творца как «наилучший из всех возможных миров» и живёт по законам, которые не в силах преступить даже их Создатель. Лейбниц считал, что мир может быть познан разумом человека, и тогда воцарится всеобщая «предустановленная гармония», для чего обязателен и единый метод познания мира.

Лейбниц указал путь для перевода логики из словесного царства, полного неопределённостей, в царство математики, где отношения между объектами или высказываниями определяются совершенно точно. Он предложил употреблять в логике математическую символику и впервые высказал мысль о возможности применения в логике двоичной системы счисления, что позднее стало использоваться в автоматических вычислительных машинах.

Заслуга Готфрида Вильгельма Лейбница ещё и в том, что он ввёл много математических терминов, прочно утвердившихся в научной практике: «функция», «дифференциал», «дифференциальное исчисление», «дифференциальное уравнение», «алгоритм», «абсцисса», «ордината», «координата», а также знаки дифференциала, интеграла, логическую символику.

Умер Готфрид Вильгельм Лейбниц 14 ноября 1716 г. в нищете, как часто случается в мире науки, — заслуги великого учёного не принесли ему богатства.

ЧАРЛЗ БЭББИДЖ

«Я отдаю себе отчёт, что мои утверждения могут рассматриваться как нечто сверхутопическое и что они вызовут в памяти философов Лапуты...» — с грустью признавался выдающийся английский учёный Чарлз Бэббидж.

(Лапута — летающий остров, придуманный Джонатаном Свифтом. Там жили мудрецы, примечательные своей оторванностью от реальной жизни и пространными псевдонаучными рассуждениями.)

Чарлз Бэббидж родился 26 декабря 1791 г. на юго-западе Англии, в маленьком городе Тотнес графства Девоншир, в семье совладельца банкирской фирмы «Прэд, Макворт и Бэббидж» Бенджамина Бэббиджа, оставившего после смерти сыну большое состояние. С детства Бэббидж увлекался всевозможными механизмами. Получая новую игрушку, он обязательно спрашивал: «А что там внутри?». Если ответ не удовлетворял его, игрушка разбиралась на части. Из-за слабого здоровья Чарлз до 11 лет учился дома, с ним занималась мать, урождённая Елизавета Тип. Потом мальчика отдали в одну из лучших частных школ Англии, где Чарлза сразу покорила богатая библиотека. Среди прочих там были прекрасные книги по математике...

Чарлз Бэббидж проявил серьёзные математические способности в период учёбы в Тринити-колледже Кембриджского университета, куда он поступил в 1810 г. Здесь обнаружилось, что Бэббидж знает математику лучше сверстников. Очень быстро Чарлз перегнал по знаниям и своих преподавателей, придя к неутешительному выводу: Британия заметно отстала от континентальных стран по уровню математической подготовки. Для преодоления разрыва в 1812 г. Чарлз и его ближайшие друзья Джон Гершель и Джордж Пикок вместе с другими молодыми математиками основали Аналитическое общество, которое своей активностью фактически инициировало реформу математического образования в Кембридже, а затем и в других университетах. Начинание оказалось плодотворным, поэтому долгие годы после смерти Бэббиджа его имя связывалось именно с формированием новой математической школы в Англии.

В 1813 г. в «Записках Аналитического общества» Бэббидж опубликовал первую научную статью «О бесконечных произведениях». В том же году Чарлз перешёл из Тринити-колледжа в колледж Святого Петра. И в 1814 г., окончив университет, получил степень бакалавра. Вторая его работа «Очерк функционального исчисления» посвящалась изучению функциональных уравнений общего вида. Она была зачитана на двух заседаниях Лондонского королевского общества (15 июня 1815 г. и 14 марта 1816 г.), и в 1816 г. Бэббиджа избрали его членом.

После окончания университета в 1814 г. Чарлз Бэббидж ведёт жизнь свободного джентльмена-философа: он вхож в высшее общество Британской империи. В 1815 г. Бэббидж женился на Джорджии Витмур из рода Шропширов. За годы супружеской жизни у них родилось восемь детей, к сожалению, пять из них умерли в детстве.

Чарлз Бэббидж.

Среди близких друзей Ч. Бэббиджа был известный писатель Чарлз Диккенс (1812—1870).

В 1817—1820 гг. он напечатал ещё несколько математических работ, преимущественно в области функционального анализа. А в 60-х гг. XX в. в Британском музее было найдено неопубликованное сочинение Бэббиджа «Философия анализа», написанное в 1821 г.

Впоследствии Бэббидж меньше занимался математикой, и, как правило, его работы были связаны с прикладными задачами.

В 1818 г. в Плимуте Бэббидж участвовал в погружении под воду в «воздушном колоколе». Позже, в 1826 г., в своей статье «Подводный колокол» он привёл чертёж небольшого подводного судна.

Однажды, слушая оперу «Дон Жуан», он заскучал и через несколько минут ушёл из зала — посмотреть, как устроен механизм управления сценой...

Интересен и другой случай. Фрагмент из стихотворной поэмы Альфреда Теннисона «Видение греха»:

Каждую минуту умирает человек,
Но каждую минуту человек
рождается —

заставил Бэббиджа написать поэту письмо: «...Я должен заметить Вам, что в этом расчёте принимается во внимание суммарное население мира в состоянии постоянного равновесия. В то же время хорошо известен факт, что вышеупомянутое количество постоянно увеличивается. Поэтому я вынужден посоветовать, чтобы в следующем издании Вашей прекрасной поэмы ошибочный расчёт, о котором я говорю, был уточнён следующим образом:

Каждое мгновение умирает
человек,
А один и шестнадцать сотых
рождается.

Я могу добавить, что точной цифрой будет 1,167, но для законов рифмы возможны приближения».

Бэббидж неоднократно посещал Францию, где познакомился с известным астрономом и физиком Пьером Лапласом, физиком и математиком Жаном Батистом Фурье. Но наибольшее влияние на Бэббиджа оказал менее именитый французский учёный-математик Г. Прони. Его печальная история началась в конце XIX в., когда в стране было задумано проведение радикальных общественных реформ. Прони было поручено (он руководил в это время Бюро переписи) создание новых логарифмических и тригонометрических таблиц для перехода на метрическую систему мер.

Прони перенёс идею разделения труда на вычислительный процесс, распределив исполнителей по трём квалификационным уровням. Заслуга Прони в том, что он нашёл алгоритмический подход для сведения сложных вычислений к простым, рутинным операциям, не требующим от исполнителей творческого подхода. (Но к тому времени, когда работы завершились, во Франции не нашлось денег, чтобы напечатать грандиозные таблицы, занимающие 17 огромных рукописных томов; колоссальный труд так и остался невостребованным.) Бэббидж застал дела Прони в плачевном состоянии, но это не помешало ему уловить главное: возможность упрощения процедуры сложных вычислений путём механического выполнения однообразных рутинных действий. Идеи учёного навели его на мысль о замене малообразованных исполнителей механическим устройством и вдохновили на создание первой дифференциальной машины.

В 1822 г. Бэббидж закончил описание *разностной машины*, которая смогла бы производить вычисления с точностью до 18-го знака. Чертёж лёг на стол премьер-министра. В 1823 г. была выплачена первая субсидия на её постройку. Строительство продолжалось десять лет, конструкция машины всё более усложнялась, творческие успехи перемежались ожесточёнными схватками между Бэббиджем и Джозефом Клементом (талантливый инженер и государственный чиновник, назначенный руководить работами). Сейчас трудно указать причину, почему первая версия разностной машины не была доведена до конца. Заслуженные английские учёные выступили против Бэббиджа, опровергая саму возможность построения машины. В 1833 г. финансирование было прекращено, а недостроенная машина перешла в государственную собственность и после была заброшена. Возможно, причиной неудач послужила излишняя разносторонность и разбросанность Бэббиджа: он то увлекался железными дорогами (сделал несколько важных изобретений, в том числе спидометр для паровоза), то занимался технологией массового производства компонентов для своей машины, требующей сотен одинаковых элементов.

Так или иначе, с первой разностной машиной было покончено, но самому автору она послужила основой для новых открытий.

Позже Бэббидж предпринял новую попытку построить разностную машину-2, но организационные проблемы опять помешали сбыться мечте.

В 1834 г. у него возникла мысль создать универсальную вычислительную машину, которую он назвал *аналитической*. В этом проекте Бэббидж впервые пришёл к идее программного управления вычислительным процессом. Он задумал сделать механическое устройство, способное не просто считать, но и управлять ходом собственной работы в зависимости от заложенной программы и результатов промежуточных вычислений. Это изобретение опередило свою эпоху на 100 лет! Возможности такой машины потрясли самого автора. В мае 1835 г. учёный писал: «Шесть месяцев я составлял проект машины более совершенной, чем первая. Я сам поражён той вычислительной мощностью, которой она будет обладать, ещё год назад я не смог бы в это поверить».

В 1840 г. Бэббидж едет с визитом в Турин (Италия), куда был приглашён с лекциями о своей машине. Лекции имели огромный успех, и один из слушателей, молодой инженер Луиджи Менабреа, издал во Франции их конспект.

Несмотря на то что Бэббидж написал более 80 научных трудов, подробное изложение принципов работы разностной и аналитической машин принадлежат не его перу — он был слишком занят созданием машин, чтобы ещё заниматься и их описанием. Разностная машина описана Ларднером, а аналитическая — Луиджи Менабреа.

Однако аналитическая машина так и не была построена. В 1851 г. Бэббидж скажет: «Все разработки, связанные с аналитической машиной, выполнены за мой счёт. Я провёл целый ряд экспериментов и дошёл до черты, за которой моих возможностей не хватает. В связи с этим я вынужден отказаться от дальнейшей работы».

Он готов был пойти на любые авантюры, чтобы добыть средства на постройку аналитической машины. Сначала был первый проект вместе с леди Адой Лавлейс, которую Бэббидж заразил идеей создания вычислительной машины. Бэббидж рассчитывал заработать деньги, написав роман в трёх томах, но вскоре забросил эту затею. Потом загорелся новым проектом — автоматом для игры в крестики-нолики. Однако и этой идее не суждено было сбыться.

Бэббидж работал над своей аналитической машиной до последних дней жизни. Учёный скончался 18 октября 1871 г., не дожив всего двух месяцев до 80-летия.

Чарлз Бэббидж был не только математиком, но и философом, экономистом и даже политэкономом. Ещё в 1832 г. Бэббидж написал книгу

Барон де Прони.

Ада Августа Лавлейс.

Генри Провост Бэббидж.

«Экономика машин и производств». (Он объездил всю Европу в поисках подходящих для своей цели научных и технических решений.)

В 1872 г. специальный комитет Британской ассоциации содействия развитию науки отметил: «Мы полагаем, что существование подобных устройств помимо экономии труда при выполнении обычных (арифметических) операций сделает осуществимым то многое, что, будучи практически осуществимым, находится слишком близко к пределам человеческих возможностей».

Сын изобретателя генерал Генри Провост Бэббидж (1824—1918), выйдя в 1874 г. в отставку, посвятил несколько лет изучению имеющегося наследия и в 1880 г. начал строить по чертежам отца центральный узел аналитической машины. В 1888 г. он вычислил произведения числа π на числа натурального ряда от 1 до 32 с точностью до 29 знаков!

Генри Бэббидж с переменным успехом продолжал работу над машиной отца до 1896 г. После десятилетнего перерыва она была возобновлена, и появился действующий образец аналитической машины, способный печатать результаты вычислений.

Машина Бэббиджа оказалась работоспособной, но изобретатель этого уже не увидел.

МАШИНЫ БЭББИДЖА

РАЗНОСТНАЯ МАШИНА

В начале XIX в. французские учёные предложили любопытный метод вычислений. Задача разбивалась на малые части, состоящие лишь из простых операций, которые поручались большому числу людей, знающих в математике только арифметические действия и выполняющих их почти механически. Чарлз Бэббидж, декан кафедры математики Кембриджского университета (её когда-то возглавлял Исаак Ньютон), заметил, что эти монотонные математические расчёты, особенно проводимые для построения различных таблиц, состоят из набора повторяющихся действий. И решил, что во избежание вычислительных ошибок и опечаток в таблицах для таких операций можно приспособить машины. Он начал проектировать автоматическую механическую вычислительную машину. В основу её работы учёный положил свойство многочленов степени $n - 1$: конечные разности n-го порядка равны нулю. Данное свойство конечных разностей можно проиллюстрировать следующим примером. Пусть надо вычислить многочлен второй степени:

$$y = x^2 + 3x - 4.$$

Для этого строится таблица из пяти колонок. В первой колонке находится значение аргумента, во второй — значение многочлена, в остальных — значения конечных разностей:

$\Delta^1_x = Y_x - Y_{x-1}$ (первые разности),

$\Delta^2_x = \Delta^1_x - \Delta^1_{x-1}$ (вторые разности),

$\Delta^3_x = \Delta^2_x - \Delta^2_{x-1}$ (третьи разности).

X	Y	Δ^1	Δ^2	Δ^3
0	–4	4	2	0
1	0	6	2	0
2	6	8	2	0
3	14	10	2	0
4	24	12	2	
5	34	14		
6	50			
7	**66**			

Первые разности получают, вычитая из значения функции её предыдущее значение. Так, в первой строке

$$\Delta^1_1 = Y_1 - Y_0 = 0 - (-4) = 4.$$

Для второй строки

$$\Delta^1_2 = Y_2 - Y_1 = 6 - 0 = 6.$$

Вторые разности получают, вычитая из первой разности её предыдущее значение. Тогда для первой строки

$$\Delta_1^2 = \Delta_1^1 - \Delta_0^1 = 6 - 4 = 2.$$

По тому же принципу строят и третьи разности. Как видно из таблицы, третьи разности равны нулю, а вторые — постоянны. Если суммировать значения разностей и предыдущее значение функции, то можно получить следующее значение многочлена. При $x = 7$

$$Y_7 = Y_6 + \Delta_5^1 + \Delta_4^2 + \Delta_3^3.$$

Или, так как все $\Delta^3 = 0$,

$$Y_7 = Y_6 + \Delta_5^1 + \Delta_4^2 = 50 + 14 + 2 = 66.$$

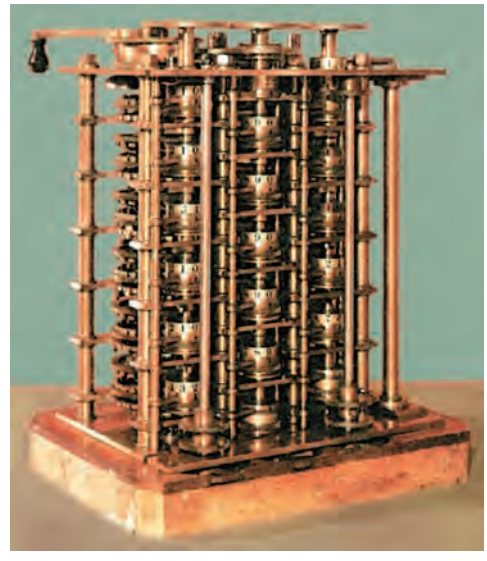
То есть можно получить новое значение функции путём суммирования ранее вычисленных значений.

Машину для автоматизации процесса составления таблиц Бэббидж назвал разностной. Он с юмором замечал, что на вопрос о принципах работы задуманной машины лучше ответить шестью знаками: $\Delta^n U_x = 0$, но такой ответ мало кому был понятен. (Это выражение означает, что для многочлена $n-1$ степени U_x n-е разности равны нулю.) Основное назначение разностной машины Бэббидж видел в составлении таблиц, однако она могла и проверять уже существующие таблицы. Для этого операции должны были проводиться в обратном порядке — от значения полинома (многочлена) к конечным разностям. Если при вычислении значений многочлена допускалась ошибка, то вместо постоянных разностей старшего порядка получались разные числа. Однако подчас проще пересчитать заново всю страницу и сверить результаты, чем выполнить обратные вычисления, на что указывал Бэббидж в своих комментариях.

Наиболее важным элементом машины Бэббиджа являлся *регистр* — устройство для хранения десятичных чисел. Поскольку предполагалось обрабатывать 18-разрядные числа, то регистр одного числа состоял из 18 зубчатых колёс. Для вычисления полинома n-й степени необходимо было иметь $n + 1$ регистр (1 — для хранения предыдущего значения функции и n — для хранения разностей). Машина предназначалась для вычисления многочленов шестой степени, поэтому в ней предусматривалось наличие семи 18-разрядных регистров.

Поскольку основная операция в машине — операция суммирования, то, естественно, имелось и суммирующее устройство, точнее, устройства, совмещённые с каждым из регистров. Конструктивно машина состояла из трёх рядов зубчатых колёс: первый ряд — зубчатые колёса регистров, второй ряд — зубчатые счётные колёса на осях для выполнения операции суммирования и третий ряд — оси с так называемыми *установочными пальцами* для подготовки очередной операции суммирования. Бэббидж разделил выполнение операции переноса десятков на две части. Впоследствии это использовалось практически во всех механических калькуляторах. Размеры механического блока машины составляли: длина — 3 м, ширина — 1,5 м; диаметр зубчатого счётного колеса около 13 см.

В машине предполагалось использование устройства, выводящего результаты на печать параллельно с проведением вычислений.



Разностная машина Чарлза Бэббиджа.

Разностная машина Бэббиджа, в отличие от всех предыдущих разработок, могла наряду с одним заранее заданным действием (вычисление полиномов по методу конечных разностей) вычислять значения некоторых функций с помощью специально подобранных формул.

Постройка разностной машины представляла большую проблему: требовалось разрабатывать не только новые конструкции и узлы, но и нередко инструменты для их изготовления; точность обработки металла была очень приближённой, да и рабочих нужной квалификации не хватало.

В 1822 г. Бэббидж опубликовал статью «Заметки о применении машин в вычислениях астрономических и математических таблиц» с описанием машины для вычисления и печати таблиц математических функций и в том же году продемонстрировал работающую модель. Она содержала три пятиразрядных регистра, т. е. предназначалась для вычисления полиномов второй степени. На этой машине Бэббидж рассчитал таблицу квадратов. Автор был полон энтузиазма: Лондонское королевское общество высоко оценило его изобретение.

В 1823 г. Бэббидж, получив финансовую поддержку британского правительства, начал постройку настоящей разностной машины. Он считал, что на её создание уйдёт три года. Но машину не удалось построить и через десять лет. Бюджет был превышен в пять раз.

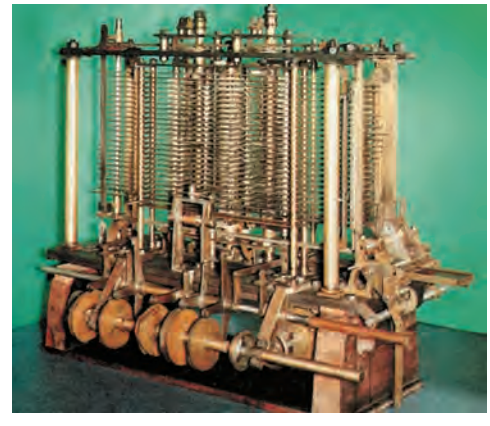
В 1833 г. испытания построенной части машины убедили автора в правильности расчётов: машина выполняла действия с запланированной точностью и скоростью. Но государство отказало учёному-пионеру в поддержке. В 1843 г. незавершённую разностную машину со всеми чертежами поместили на хранение в музей Королевского колледжа в Лондоне. Именно из частей этой машины была построена действующая модель, находящаяся сейчас в Кембридже.

АНАЛИТИЧЕСКАЯ МАШИНА

После испытаний 1833 г. Бэббидж задумал создать принципиально новую машину, способную выполнять различные действия в соответствии с предварительно составленным планом работ — программой. Первые чертежи появились в 1834 г., машину назвали *аналитической*.

Бэббидж решил слегка изменить конструкцию разностной машины: расположить оси по окружности так, чтобы колонка результата располагалась рядом с колонкой последней разности и легко связывалась с ней. Это позволило бы управлять процессом вычислений, воздействуя на порядок расчёта. Вскоре учёный сформулировал идею независимого устройства управления вычислениями, с помощью которого можно было производить все арифметические действия.

Аналитическая машина задумывалась как чисто механическое устройство, однако учёный хотел выполнять расчёты не вручную, а с применением внешнего источника энергии, в частности парового двигателя.



▶ Аналитическая машина Чарлза Бэббиджа.

Период с 1842 по 1848 г. был полностью посвящён созданию новой машины. Автор продолжал работу без какой-либо финансовой поддержки от правительства. В 1849 г. Бэббидж представил схему аналитической машины. Машина состояла из трёх основных блоков.

• Склад (Store) — память для хранения чисел на регистрах, состоящих из механических колёс. Предполагалось, что память должна хранить тысячу 50-разрядных десятичных чисел. Если в дальнейших вычислениях содержимое регистра не требовалось, число можно было напечатать и использовать регистр для других вычислений. Если памяти не хватало, чтобы вместить все числа, необходимые для вычисления, то они записывались на перфорированные карты (разновидность жаккардовых карт), при этом число таких карт (см. статью «Автоматизация и перфокарты») ничем не ограничивалось.

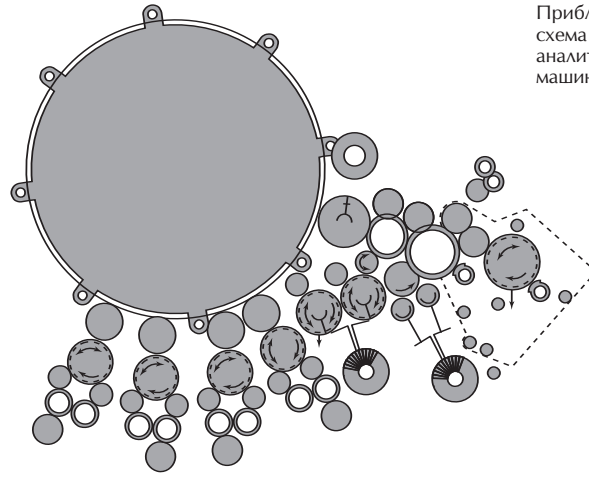
• Фабрика (Mill) — блок для выполнения арифметических операций. Бэббидж так оценивал производительность своей машины: умножение двух 50-разрядных чисел и деление 100-разрядного числа на 50-разрядное вычисляются со скоростью одна операция в минуту, а сложение (и вычитание) 50-разрядных чисел — одна операция в секунду.

• Устройство, оставленное автором без названия, для управления процессом вычисления, осуществления выборки чисел из памяти, выполнения вычислений и вывода результатов.

Перфокарты, используемые в машине, разделялись на две группы: *операционные* — для выполнения арифметических операций и *управляющие* — для осуществления загрузки чисел из регистров в арифметическое устройство и выгрузки обратно в память, а также ввода-вывода.

Соратница Бэббиджа Ада Августа Лавлейс предложила назвать перфокарты, управляющие передачей чисел в машине, *переменными:*

• «нулевая карта» загружает числа из регистра в арифметическое устройство, при этом содержимое регистра очищается, т. е. устанавливается в нулевую позицию;

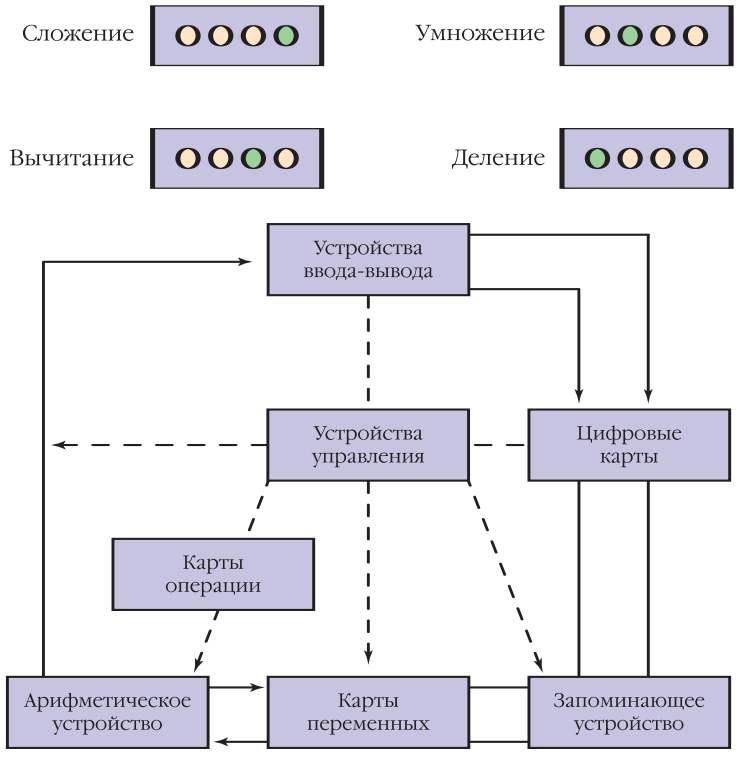


Приблизительная схема механизма аналитической машины.

• «удерживающая карта» производит то же действие, что и «нулевая карта», только содержимое регистра сохраняется;

• «доставляющая карта» передаёт результат из арифметического устройства обратно в память.

Карты, применяемые для ввода чисел в память, назывались *цифровыми*. Вот как Бэббидж описывал механизм чтения с перфокарт: «Два больших ящика, один из которых пустой, а другой — наполненный перфорированными картами, располагаются спереди и сзади многогранной призмы. Эта призма прерывисто вращается и каждый раз продвигается вперёд на короткое расстояние, после чего немедленно возвращается. Карта проходит над призмой только перед каждым ходом механизма, такого же, как и челнок (в станке Жаккарда. — *Прим. ред.*).



Кодирование арифметических операций на перфокартах.

Карты, которые прошли, падают вниз, пока не достигнут пустого ящика для сбора карт, в котором они располагаются одна над другой...».

Сын изобретателя Генри Бэббидж приводил такой пример вычислений на аналитической машине: «Если взять формулу $(ab + c)d$ в качестве иллюстрации, то потребуются карты всех видов. Будет использован следующий порядок их загрузки: четыре карты исходных данных — a, b, c и d — собраны вместе и уложены на ролик (или призму). Эти числа будут помещены на колонки (зубчатого механизма. — *Прим. ред.)*, предназначенные для них в части машины, называемой «Склад». Каждое полученное число запоминается и готово к использованию... Применяются три операционные карты и четырнадцать управляющих карт, каждый набор карт, собранный вместе, укладывается на свой ролик (или призму)...».

Приведённая таблица показывает последовательность операций в аналитической машине для вычисления

$$y = (ab + c)d.$$

Число занимает один регистр памяти (колонка из десятичных цифровых колёс). По окружности колёс нанесены цифры от 0 до 9. Каждое колесо вращается независимо от других. Управляющие карты поднимают колёса выбранных колонок так, чтобы ввести их в зацепление с зубчатыми рейками. При этом каждое колесо передвигается на число зубьев, соответствующее разряду числа. В результате число, нанесённое на перфокарту, записывается в память. Для хранения знака используется самое верхнее колесо в колонке. За один оборот главного вала можно сохранить число в колонке или передать из памяти к Складу.

При сложении чисел в арифметическом устройстве число уменьшается до нуля в одной колонке, заставляя поворачиваться на эту величину другую колонку. При прибавлении 4 к 8 результатом будет 2, при повороте колеса последовательно «промелькнут» цифры 9, 0 и 1. Когда 9 перейдёт в 0, будет сдвинут рычаг переноса, который осуществится позднее.

Несомненная заслуга Бэббиджа — введение программного управления ходом вычислений. Если необходимо выбрать из двух чисел большее (операция сравнения), то надо вычесть из одного числа другое. В результате, когда уменьшаемое меньше вычитаемого, должен возникнуть перенос. При этом рычаг перемещается в самое верхнее положение, что означает отрицательный результат. Такое положение рычага позволяет ввести в работу предварительно подготовленный набор перфокарт.

Леди Лавлейс предложила способ возврата одной или нескольких «отработанных» перфокарт из ящика-приёмника обратно в источник для последующего считывания и выполнения действий. Таким образом, стало возможно многократно повторять целые участки программ, т. е. организовывать программные циклы.

К сожалению, как разностная, так и аналитическая машины при жизни автора не были построены. Одной из причин считают то, что мыслитель Бэббидж намного опередил своё время: при существующей технике и про-

Управляющие карты	Операционные карты	Операция
1	—	Установить *a* в колонке 1 Склада
2	—	Установить *b* в колонке 2 Склада
3	—	Установить *c* в колонке 3 Склада
4	—	Установить *d* в колонке 4 Склада
5	—	Прислать *a* из Склада на Фабрику (арифметическое устройство)
6	—	Прислать *b* из Склада на Фабрику
—	1	Умножить *a* на *b*, $ab=p$
7	—	Положить *p* в колонку 5 Склада и сохранить для последующего использования
8	—	Прислать *p* на Фабрику
9	—	Прислать *c* на Фабрику
—	2	Сложить *p* и *c*, $p+c=q$
10	—	Взять *q* в колонке 6 Склада
11	—	Прислать *d* на Фабрику
12	—	Прислать *q* на Фабрику
—	3	Умножить *d* на *q*, $dq=p_2$
13	—	Взять p_2 в колонке 7 Склада
14	—	Напечатать p_2

изводстве нельзя было детально и точно воспроизвести чертежи.

В аналитической машине предусматривались все основные элементы, присущие современным компьютерам: Склад — *память,* Фабрика — *арифметическое устройство процессора,* устройство для управления — *управляющее устройство процессора*. Архитектура машины практически соответствует архитектуре современных ЭВМ, а команды, которые выполняла аналитическая машина, в основном включают все команды современных процессоров, в том числе и изменение порядка выполнения программы, условный переход, цикл. Поэтому машину Бэббиджа хочется назвать первым настоящим компьютером.

ГЕРМАН ХОЛЛЕРИТ

29 февраля 1860 г. в американском городе Буффало в семье немецких иммигрантов Холлеритов родился сын Герман. Когда мальчик пошёл в школу, выяснилось, что он страдает весьма неприятным заболеванием — дисграфией и с трудом пишет (в то же время рисовал он весьма неплохо). И хотя математика и естественные науки давались Герману гораздо легче, мучения на уроках грамматики в конце концов вынудили его в возрасте 14 лет оставить школу. Усиленные домашние занятия позволили ему год спустя поступить в Горную школу при Колумбийском университете. На способного юношу обратил внимание один из преподавателей, профессор У. Траубридж, и пригласил после окончания школы в 1879 г. на работу в возглавляемое им Национальное бюро по переписи населения. Это стало решающим моментом в судьбе Германа Холлерита.

Бюро занималось сбором и статистической обработкой информации при проведении переписей населения. Переписи в США имели давнюю традицию. Первая из них состоялась

Герман Холлерит.

ПЕРВАЯ ПЕРЕПИСЬ НАСЕЛЕНИЯ В РОССИИ

5 июня 1895 г. император Николай II утвердил Положение о первой всеобщей переписи населения, в котором говорилось, что перепись «есть дело для России совершенно новое, никогда в ней до сих пор небывалое. Перепись эта будет состоять в том, что будут пересчитаны поодиночке все жители государства для того, чтобы Правительство могло знать верные числа и счёт населения как во всём государстве, так и в каждом уезде, в каждой волости, селении и усадьбе».

Перепись была назначена на 9 февраля 1897 г. Планировалось собрать сведения о каждом жителе страны по следующим признакам: имя, семейное положение, отношение к главе хозяйства, пол, возраст, сословие, вероисповедание, место рождения, место жительства, родной язык, грамотность, занятие, физические изъяны.

В августе 1895 г. на сессии Международного статистического института в Берне (Швейцария) директор Центрального статистического комитета (ЦСК) Н. А. Тройницкий познакомился с американским инженером Германом Холлеритом, чей доклад был посвящён использованию счётных машин при проведении переписи. Заинтересованность в сотрудничестве оказалась обоюдной. Холлерит надеялся на расширение рынка сбыта своих машин, а Тройницкий прекрасно понимал, что вручную обработать данные переписи невозможно.


Безплатно.

№ Листа

ПЕРВАЯ ВСЕОБЩАЯ ПЕРЕПИСЬ

населенія Россійской Имперіи,

ПЕРЕПИСНОЙ ЛИСТЪ

ФОРМА А.

В начале ноября 1896 г. Холлерит получил от российского правительства приглашение участвовать в переписи и уже 15 декабря прибыл в Санкт-Петербург. Договориться об условиях контракта удалось очень быстро. Согласно ему, фирма Холлерита предоставляла России в аренду 35 старых машин, уже использовавшихся в других странах. Машины необходимо было вернуть к апрелю 1900 г., до начала очередной переписи в США. Кроме того, Россия закупала 70 новых табуляторов в комплекте с сортировочными машинами, а также 500 перфораторов.

Сборку техники под руководством самого Холлерита решили производить в Петербурге. Здесь же в целях экономии по его чертежам изготовляли отдельные узлы и детали. Гонорар Германа Холлерита составил 67 571 доллар, из них 59 500 долларов причиталось за новые машины.

Перепись состоялась в назначенный срок. Её проводили сотни тысяч счётчиков: переписные листы заполняли местные чиновники, служащие, интеллигенция. Например, участком по проведению переписи в Мелихове заведовал А. П. Чехов. Он в шутку называл себя «ротным командиром» и «боцманом» местных счётчиков.

«Выдали счётчикам отвратительные чернильницы, отвратительные аляповатые знаки, похожие на ярлыки пивного завода, и портфели, в которые не лезут переписные листы, — и впечатление такое, будто сабля не лезет в ножны. Срам. С утра хожу по избам, с непривычки стукаюсь головой о притолоки, и, как нарочно, голова трещит адски...» — писал Антон Павлович 11 января. «Замучила перепись», — жалуется он 30 января. А 8 февраля с облегчением подводит итог: «Перепись кончилась. Это дело изрядно надоело мне, так как приходилось и считать, и писать до боли в пальцах, и читать лекции 15 счётчикам. Счётчики работали превосходно, педантично до смешного. Зато земские начальники, которым вверена была перепись в уездах, вели себя отвратительно. Они ничего не делали, мало понимали и в самые тяжёлые минуты сказывались больными».

Но для ЦСК тяжёлые времена были ещё впереди. Заполненные переписные листы свозили для обработки в столицу. Работа была поставлена из рук вон плохо, многие сотрудники просто не знали, что и как они должны делать... Работа двигалась настолько медленно, что летом 1897 г. Холлерита вновь вызвали в Россию. И хотя его консультации несколько поправили положение, в январе 1902 г. план обработки всё же пришлось значительно сократить. Подсчёты продолжались почти восемь лет — общий свод результатов переписи по империи был опубликован лишь в конце 1905 г.

В истории первая всероссийская перепись населения осталась примером того, как не надо организовывать дело. А последний свидетель тех событий, единственный сохранившийся табулятор Германа Холлерита, с 1952 г. находится в собрании Политехнического музея в Москве.

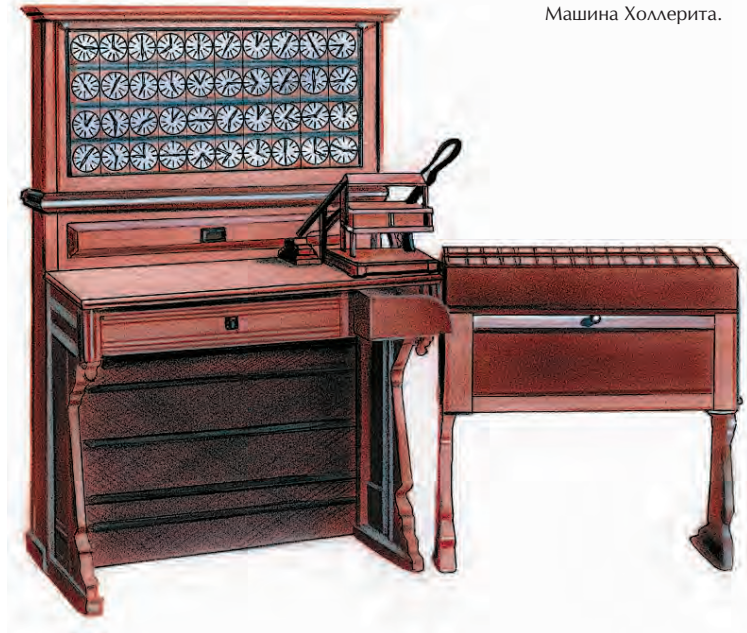
ещё при Джордже Вашингтоне, в 1790 г. В XIX в. переписи проводились каждые десять лет. Население постоянно росло, а бурно развивавшаяся американская экономика остро нуждалась в большом количестве разнообразных данных о его составе. Например, в 1880 г. численность населения составила 50 млн человек, и на каждого требовалось завести и заполнить карточку, содержавшую 210 рубрик! Подсчёты и обработка результатов переписи затянулись на семь с половиной лет, почти до следующей переписи. Это был тупик. Срочно требовались новые методы организации работы.

На мысль механизировать труд счётчиков Холлерита навёл доктор Джон Биллингс, возглавлявший в бюро департамент, в котором составляли сводные данные. Он же предложил использовать для записи информации перфокарты. Правда, сначала Холлерит хотел использовать бумажную ленту, намотанную на барабан и скользившую по металлическому столу. Сверху её прижимала металлическая полоса с рядом слабо закреплённых тупых гвоздей. При попадании гвоздя в отверстие на ленте электрический контакт замыкался, приводя в движение счётный механизм. Скоро выяснилось, что бумага часто рвётся, информация не успевает считываться, да и перематывать ленту в поисках нужных данных слишком долго. Так что от лент пришлось отказаться в пользу перфокарт. Позднее Холлерит писал, что окончательное решение он принял, наблюдая за работой кондуктора в поезде. Тот компостером пробивал на билете дырки в условных местах, отмечая таким образом пол, цвет волос и глаз пассажира. В результате получалось что-то вроде перфокарты, содержащей описание внешности владельца билета.

Первый *табулятор* был опробован в 1887 г. в статистическом бюро Балтимора. Результаты оказались весьма обнадёживающими, и через два года состоялось ещё одно испытание — частичная перепись населения в Сент-Луисе. Выигрыш во времени по сравнению с ручным подсчётом был двукратным, а на некоторых участках работы — десятикратным! Правительство заключило с Холлеритом контракт на поставку оборудования, и уже в июне 1890 г. началась первая в истории перепись населения с применением машин. Обработка её результатов, занесённых на 62 млн карточек, заняла менее двух лет, а экономия составила 5 млн долларов — огромнаю сумма, ведь государственный бюджет США не превышал тогда 100 млн долларов. Система Холлерита не только обеспечивала высокую скорость, но и позволяла сравнивать статистические данные по самым разным параметрам. Холлерит совершенствовал свою машину. Он разработал удобный клавишный перфоратор, автоматизировал процедуры подачи и сортировки перфокарт. В общей сложности изобретатель получил более 30 патентов.

Джон Биллингс.

Заслуги Холлерита признаны во всём мире. Институт Франклина в Филадельфии наградил его медалью, ему вручили золотую медаль на Парижской выставке 1893 г. Альма-матер изобретателя Горная школа присудила Холлериту, не имевшему высшего образования, учёную степень доктора философии, а научные общества Европы и Америки избрали своим почётным членом. Табулятор Холлерита использовался на переписях



Машина Холлерита.

Офис IBM. 1948 г.

Томас Уотсон.

в Австро-Венгрии, Канаде, Норвегии, Англии и других странах.

В 1896 г. Холлерит основал компанию Tabulating Machine Company (TMC). Его машины применялись повсюду: на железных дорогах и промышленных предприятиях, в крупных торговых фирмах и страховых компаниях. С их помощью начисляли заработную плату и вели складской учёт, решали множество других задач. В 1900 г. Госдепартамент США вновь выбрал систему Холлерита для переписи. Однако вскоре государственные чиновники решили отказаться от системы Холлерита. Новые машины создал инженер Джеймс Пауэрс, сотрудник Национального бюро по переписи населения. Они были чисто механическими и имели весьма сложную конструкцию. Холлерит, потеряв государственные заказы, в том же году отошёл от предпринимательской деятельности. Он продал фирму, оставшись консультантом в поглотившей её компании Computer Tabulating Recording Company (CTRC). Когда в 1921 г. консультант CTRC Герман Холлерит уходил в отставку, фирма уже уверенно смотрела в будущее. А ещё три года спустя директор Томас Уотсон переименовал фирму в IBM (International Business Machines).

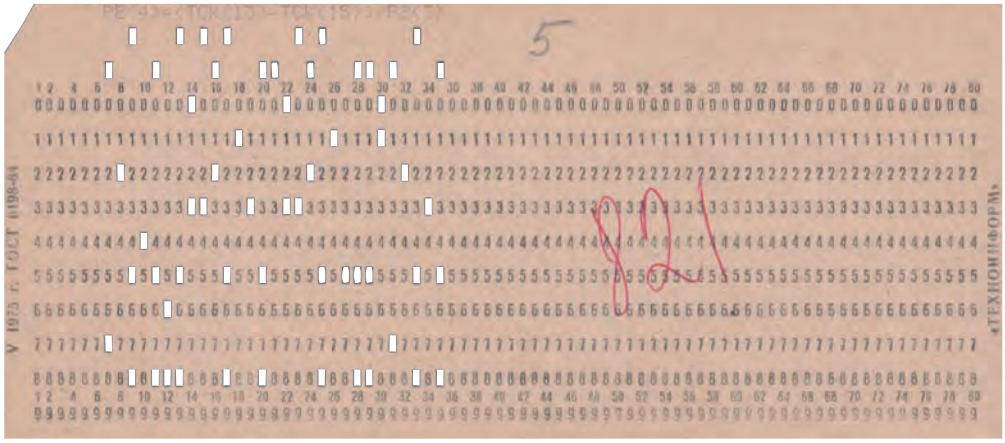
Герман Холлерит скончался 17 ноября 1929 г. в Вашингтоне. И хотя официально основателем могущественной империи IBM считается Томас Уотсон, многочисленные награды и патенты Холлерита занимают в музее компании одно из самых почётных мест.

АВТОМАТИЗАЦИЯ И ПЕРФОКАРТЫ

Одна из простейших «машин», которой можно программно управлять, — обычная электрическая лампочка. С помощью выключателя можно зажечь или погасить её. Если разобрать выключатель, то обнаружатся контакты, замыкание которых приводит к включению, а размыкание — к выключению лампочки.

То же действие легко продемонстрировать, вставляя между замкнутыми контактами кусок картона. Он разомкнёт цепь, и лампочка погаснет. Если в картоне проделать отверстие (например, дыроколом), то лампочка будет гореть. Можно составить программу для лампочки и записать её на куске картона: есть дырка — горит лампочка, нет — не горит. Подобная идея использована в *перфокар-*

▶ Подготовка перфокарт для станка Жаккарда.



Перфокарта 70-х гг. XX в. (написанный от руки номер помогал не перепутать карту в колоде).

max — картонных карточках с отверстиями.

В XVIII в. во Франции конструкторы ткацких станков пытались при использовании перфокарт сделать так, чтобы челнок (главная деталь станка) работал автоматически, по программе. Преуспел в этом сын лионского ткача Жозеф Мари Жаккар. В 1801 г. он создал автоматический ткацкий станок, управляемый перфокартами.

Наличие или отсутствие отверстий в карте заставляло нить подниматься и опускаться при каждом ходе челнока. Таким образом, поперечная нить могла обходить каждую продольную с той или иной стороны в зависимости от программы на перфокарте. Сотни таких нитей использовались для создания затейливых узоров. Имя Жозефа Мари Жаккара получил не только станок, но и плетение «жаккард», которое по праву относят к наиболее сложным и запутанным плетениям, исполняемым на ткацком станке.

Этот станок был первым массовым промышленным устройством, автоматически работающим по заданной программе. Его считают одной из самых совершенных машин, когда-либо созданных человеком. Модель станка для крупноузорчатых тканей отмечена медалью Парижской выставки 1801 г., а сам Жаккар был приглашён на работу в парижский музей (Консерватория искусств и ремёсел), и вскоре в одной лишь Франции работало более 10 тыс. таких станков. В 1808 г. он создаёт более совершенный станок для узорчатого ткачества, известный по сей день. Но с этим изобретением связана и печальная история. Восстание лионских ткачей в 1803 г. было спровоцировано установкой станка в мастерской Жаккара в Лионе. Рост производства тканей, а следовательно, их дешевизна и доступность вызвали социальный взрыв.

Жозеф Мари Жаккар.

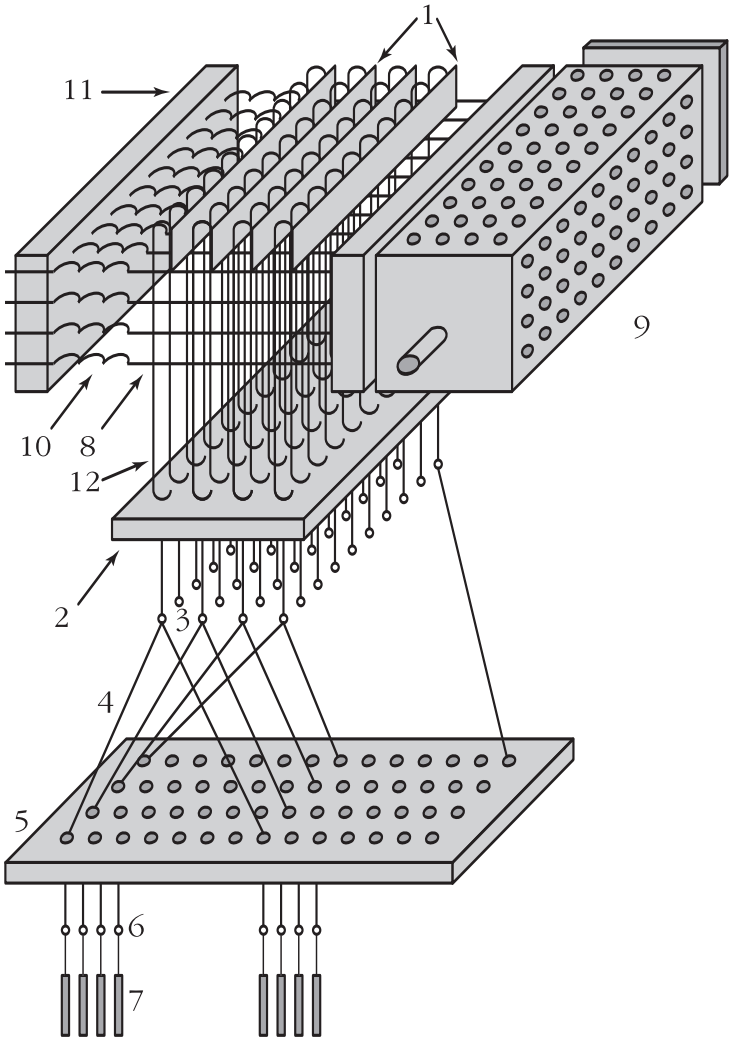
Перфокарты произвели переворот не только в ткацком деле, но и в статистике. Статистика постоянно сталкивается с проблемами обработки огромного количества информации. Герман Холлерит, служащий американского Национального бюро по переписи населения, впервые применил «компьютер» для решения практических задач.

Машина Холлерита включала:

• клавишный перфоратор, позволяющий пробивать (перфорировать)

Машина Холлерита.

около ста отверстий в минуту одновременно на нескольких картах;



• машину для сортировки, которая представляла собой набор ящиков с крышками (карты продвигались по своеобразному конвейеру; с одной стороны карты находились считывающие штыри на пружинах, с другой — резервуар со ртутью; когда штырь попадал в отверстие на перфокарте, то благодаря пружине слегка касался ртути, находившейся на другой стороне, и замыкал электрическую цепь, открывая крышку соответствующего ящика; туда и попадала перфокарта);

• табулятор, который работал аналогичным образом, только замыкание электрической цепи приводило к увеличению показаний соответствующего счётчика на единицу.

Перфокарты были размером с долларовую купюру, они имели 240 позиций для перфорации (12 рядов по 20 позиций). Сейчас бы сказали, что в машине Холлерита использовалось кодирование информации с помощью перфокарт.

Схема станка Жаккарда

1 — ножи;
2 — рамная доска;
3 — рамные шнуры;
4 — аркатные шнуры;
5 — делительная доска;
6 — лицы;
7 — грузики;
8 — иглы;
9 — перфорированная призма;
10 — пружина;
11 — доска;
12 — крючки

«НАУЧНЫЕ» КАЛЬКУЛЯТОРЫ

Томас де Колмер.

В 1820 г. во Франции Карл Ксавье Томас де Колмер (1785—1870), используя идеи машины Лейбница, построил первый механический компьютер-калькулятор — *арифмометр*. Машина могла не только складывать и вычитать, но и производила деление и умножение в ручном режиме. Томас де Колмер 18 ноября 1820 г. получил во Франции патент №1420 на своё детище.

Арифмометр имел настоящий коммерческий успех. Более полутора тысяч машин было изготовлено в течение последующих 90 лет. В Германии в 1878 г. Артур Бургхард наладил производство подобных машин.

В России над созданием арифмометра с 1874 г. работал швед Вильгод Однер, а в 1890 г. механические настольные счётные машины стали производить на заводе. (Их модифика-

С. С.! Вильгельм/Вильгодт????

ция «Феликс» выпускалась в России вплоть до 80-х гг. XX столетия. Основная деталь — зубчатое колесо носит имя Однера.) В начале XX в. арифмометры Однера выпускались во всём мире, только в России к 1914 г. их работало более 20 тыс.

В Швеции с 1913 г. производится MADAS (от *англ.* multiplication, automatic division, addition and Subtraction — «умножение», «деление», «сложение» и «вычитание»). Полностью автоматическое умножение и деление арифмометры стали производить в 1927 г.

Термин «арифмометр» подразумевает механическое устройство, которое способно выполнять четыре арифметических действия. С усовершенствованием механических калькуляторов был добавлен ряд функций: запоминание промежуточных результатов, последующие операции над ними, печать результатов и т. п. Это явилось следствием расширяющегося коммерческого спроса на настольные счётные машины, а не результатом научных исследований. Научный результат состоит в том, что и в современных компьютерах есть арифметическое устройство (или устройства), которое умеет производить простые (в некоторых машинах сложные) арифметические действия. Максимальное количество разрядов числа, с которым может оперировать арифметическое устройство, называется *размером машинного слова.*

В конце XIX в. прогресс требовал развития математической физики. В проектировании железных дорог и строительстве пароходов, текстильных фабрик и мостов нужны были машины, способные эффективно производить многократно повторяющиеся вычисления. Но до таких машин было ещё далеко.

В 1880 г. американский изобретатель Томас Алва Эдисон (1847—1931) начал выпуск безопасных электрических лампочек. В 1883 г. он ввёл в вакуумный баллон платиновый электрод, подал напряжение и обнаружил, что между электродом и угольной нитью протекает ток. Явление, известное как термоэлектронная эмиссия, получило название «эффект Эдисона».

Арифмометр Колмера.

В 1896 г. компания General Electric разработала новый стандарт в электричестве: использование переменного тока для питания электрических приборов. Калькуляторы оснащаются электрическими моторами для проведения вычислений и печати на бумаге. В компьютерах всё большее применение находят компоненты, работающие на электрическом токе.

В 1904 г. английский физик Джон Амброз Флеминг (1849—1945) на основе открытия Эдисона создал диод — электронное устройство, пропускающее ток в одну сторону и не пропускающее в другую. Через два года американский инженер Ли де Форест (1873—1961) изобрёл триоды.

XIX век и вся предыдущая история показали необходимость создания автоматического вычислительного устройства, работающего автономно, без участия человека, по заранее подготовленной программе. Однако первые компьютеры можно считать механическими, так как в качестве запоминающего устройства (памяти) использовались механические приспособления.

Английский математик Джордж Буль (1815—1864) ещё в 1848 г. описал постулаты (правила) логики, оперирующей алгебраическими элементами, которые могут принимать только два возможных состояния («да» или «нет», 0 или 1), названной впоследствии булевой алгеброй. Фактически

Вильгельм Однер.

Арифмометр «Феликс».

благодаря его алгебре стало возможно конструирование логической схемы компьютера.

Таким образом, в XX в. в компьютерах стало применяться двоичное кодирование информации, где минимальный объём памяти — 1 бит — представлял собой переключатель с двумя состояниями — 0 и 1, который легко «реализовался» инженерами. Однако встал вопрос об электронной «версии» переключателя.

Русский учёный Михаил Александрович Бонч-Бруевич в 1918 г. и английские учёные В. Икклз и Ф. Джордан чуть позже, в 1919 г., независимо друг от друга создали электронное реле, которое могло находиться в одном из двух состояний (0 или 1). Реле, сейчас известное как *триггер,* стало основой памяти ЭВМ.

К XX столетию всё было подготовлено для рождения ЭВМ.

КОНРАД ЦУЗЕ

Лампа Эдисона.

Конрад Цузе родился 22 июня 1910 г. в Берлине. В 1935 г. получил диплом инженера-строителя, специалиста по прочности в Техническом университете в Берлине. Работу над созданием вычислительных устройств Цузе начал ещё будучи студентом, в 1934 г. Так что его без преувеличения можно назвать пионером современной вычислительной техники на европейском континенте.

Цузе отличался аналитическим, научным складом мышления. Для проведения сложных инженерных расчётов он хотел создать более мощный, чем существующие, вычислитель. Обдумывая проблему вычислений, Цузе вскоре пришёл к нескольким выводам. Во-первых, вычислитель должен управляться с помощью программ, причём предпочтительно использовать двоичную систему счисления и арифметику с плавающей запятой. Во-вторых, необходимо иметь полностью автоматическое арифметическое устройство, память большого объёма и элементы с двумя устойчивыми состояниями. Таким образом, Цузе фактически воспроизвёл основные идеи аналитической машины Чарлза Бэббиджа (о которой он узнает лишь спустя несколько лет). Только в одном отношении идеи Цузе уступали идеям его гениального предшественника: он не предусмотрел команды условного перехода.

Цузе понял также, что вычисление — универсальное понятие, что это просто преобразование данных, которые можно представить в виде комбинации двоичных разрядов (сейчас мы называем их битами, а сам Цузе использовал выражение «состояние да/нет»). Теории нуждались в проверке практикой, и Цузе начал с создания устройства памяти. Со времён Паскаля числа в вычислителях представлялись набором дисков с нанесёнными на обод десятью цифрами. Выбор двоичной системы позволял использовать другие элементы, например реле, принимающие только два положения: «открыто» и «закрыто». Однако сначала Цузе всё-таки остановился на механическом решении. Сконструированная им память состояла из набора металлических пластин, способных перемещаться в определённом направлении. Значения обрабатываемых величин и выполняемая операция задавались смещением пла-

стин. Оно вызывало смещение других пластин, соответствующее вырабатываемому результату. Этот результат сохранялся в наборе битов памяти и, в свою очередь, мог быть использован в дальнейших вычислениях. В 1936 г. устройство памяти было запатентовано, причём в заявке подчёркивалось, что хранить в ней можно наряду с числами произвольную информацию, в том числе и сами программы, записанные в двоичном виде.

Память оказалась достаточно компактной и могла быть расширена до 1 тыс. слов (релейная память такого объёма потребовала бы 40 тыс. реле, заняв целую комнату).

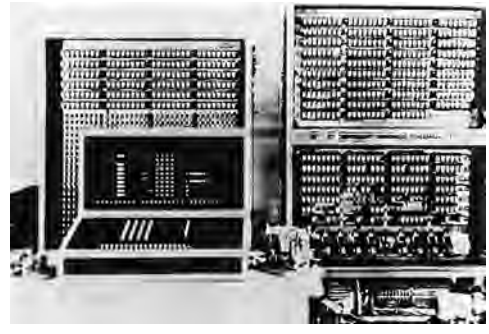
Первая полностью механическая машина Z1 была построена в 1936—1938 гг. Управление ею осуществлялось от перфоленты, на которую записывались трёхадресные команды программы. Память имела объём 16 чисел по 24 бит. Надёжностью машина не отличалась, поэтому Цузе решил перейти на новую элементную базу, заменив все механические элементы на электромеханические реле. Но сначала он построил небольшую машину Z2, оперировавшую 16-разрядными двоичными числами с фиксированной точкой. Работа над ней завершилась в 1939 г. Арифметическое устройство, выполненное из 200 старых телефонных реле, соединялось в машине с механической памятью.

Этот опыт придал Цузе уверенность в своих силах. Он понял, что готов создать машину без использования механических элементов. Хотя в 1939 г. работа над Z3 прервалась (Цузе призвали на военную службу), упорно трудясь по вечерам и в выходные дни, он сумел в 1941 г. завершить разработку. Это была электромеханическая программно управляемая универсальная машина и фактически первое в истории реальное воплощение идеи аналитической машины Бэббиджа!

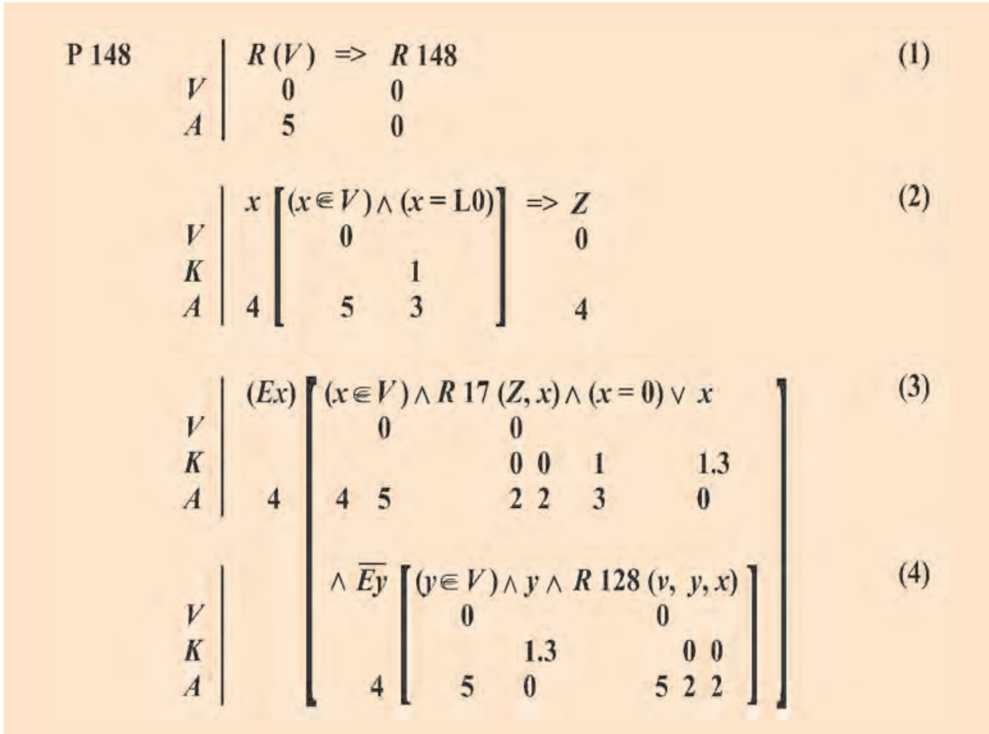
Как и в предыдущих моделях, здесь использовалась двоичная система счисления. Команды были одноадресными (это во многих случаях увеличивало скорость вычислений благодаря отсутствию лишних записей

Конрад Цузе.

в память), они вводились с перфоленты (её роль выполняла киноплёнка). Исходные данные задавались с клавиатуры, а результаты вычислений высвечивались на специальном табло. Операции над числами с плавающей точкой реализовывались аппаратно (всего имелось девять арифметических команд: сложение, вычитание, деление, извлечение квадратного корня, а также умножение на $^1/_2$, 2, $^1/_{10}$, 10 и на -1). Сложение выполнялось за 0,3 с, а умножение занимало от 4 до 5 с. Арифметическое устройство было построено из 600 реле. Ещё 1800 реле потребовалось для памяти объёмом 64 числа по 22 бит (один знаковый разряд, 14 бит на мантиссу и 7 бит на порядок).



Машина Z3.



Одна из подпрограмм игры в шахматы, написанная К. Цузе на языке Plankalkul.

Z3 использовалась для весьма трудоёмких расчётов, связанных с определением прочности конструкций самолётов. Машина погибла во время бомбёжки Берлина, но её копия, сделанная спустя 20 лет, экспонируется сейчас в одном из музеев Мюнхена.

В 1942 г. Цузе приступил к постройке более мощной машины Z4. Правда, память на 1024 слова была в ней механической, но длина чисел увеличивалась до 32 бит. Планировалось также улучшить систему ввода-вывода и предоставить программисту больше возможностей по написанию гибких программ. К сожалению, эти идеи ему так и не удалось реализовать.

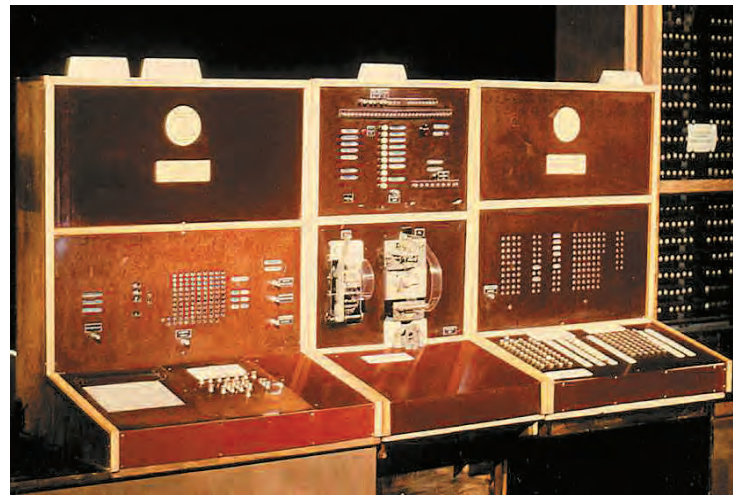
Судьба Z4 сложилась удачнее, чем у её предшественниц. В конце войны, спасая машину от бомбёжек, Цузе покинул Берлин. Переезжая с места на место, он ухитрялся при этом не прекращать работу. За несколько недель до капитуляции Германии Конрад Цузе нашёл убежище в старинном университетском городе Гёттингене. Там 28 апреля 1945 г. работающая машина впервые была продемонстрирована местным учёным. Несколькими днями позже, отказавшись отправить машину на подземные заводы в горах Гарца, где фашисты лихорадочно дорабатывали «оружие возмездия», с помощью сотрудников Вернера фон Брауна (изобретателя реактивных снарядов Фау-1 и Фау-2) Цузе снова увозит её в никуда. Странствия по горящей, разрушенной стране закончились далеко на юге, в маленькой альпийской деревушке Хинтерштейн. Именно здесь, в помещении бывшей конюшни, постройка Z4 была завершена. Здесь же Цузе создал первый в истории язык программирования Plankalkul.

До 1950 г. Z4 оставалась практически единственным работающим компьютером в Европе. С 1950 по 1954 г. она успешно эксплуатировалась в Техническом университете в Цюрихе (Швейцария). Ещё позднее в знак признания исторического значения Z4 поместили в музей в Мюнхене.

Цузе был пионером во многих направлениях развития и применения компьютеров. Вероятно, он пер-

вым в мире построил вычислительное устройство, предназначенное для управления и контроля технологических процессов. Во время войны Цузе (как инженер, специалист по прочности конструкций) был привлечён к работам по производству управляемых реактивных снарядов. Недостаточная точность изготовления ухудшала аэродинамические характеристики снарядов, поэтому в конце производственной линии каждое изделие подвергалось тщательному обследованию с помощью 80 датчиков. Затем производились достаточно сложные вычисления, определявшие степень отклонения от проектных норм и необходимую корректировку изделий. Сначала Цузе построил специализированный вычислитель из 500 реле, выполнявший фиксированную последовательность операций. Машина заменяла 12 человек и безотказно работала в течение двух лет, с 1942 по 1944 г., по две смены ежедневно. Однако при этом требовалось участие техника, который считывал, записывал и передавал показания датчиков оператору, вводившему их в вычислитель. Цузе удалось полностью устранить вмешательство человека в этот процесс. Данные с датчиков считывались автоматически и передавались в вычислитель через изобретённое им устройство, которое сейчас мы бы назвали аналого-цифровым преобразователем. Таким образом, Цузе удалось встроить вычислитель в замкнутый технологический процесс. Так что и полностью автоматизированные огромные современные производства, и миниатюрные встроенные микропроцессоры в бытовых приборах — все они ведут свою родословную от скромного релейного устройства Конрада Цузе.

Ещё до войны вместе другом Гельмутом Шрейером Цузе выполнил проект полностью электронной машины, содержавшей 2 тыс. электронных ламп. Их идея казалась в то время нереализуемой и не нашла поддержки в официальных инстанциях, но к 1945 г. им всё же удалось собрать макет арифметического устройства на ста лампах, работавшего с десяти-



Машина Z4.

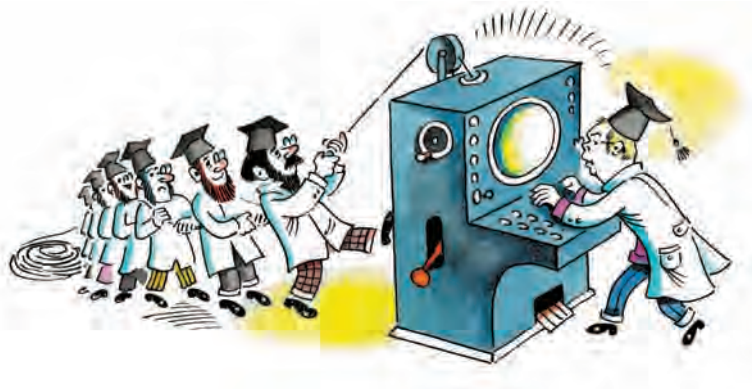
разрядными числами. (Когда после войны стало известно о появлении компьютера ENIAC, содержавшего 18 тыс. ламп, Цузе был потрясён — настолько невероятным представлялось создание такого сложного электронного устройства.) Эта работа Цузе не получила развития, поскольку в побеждённой Германии исследования в области электроники запрещались.

В истории вычислительной техники имя Конрада Цузе стоит особняком. Один человек, работавший с несколькими помощниками, не только повторил путь, проделанный большими коллективами учёных в других странах, но часто опережал их. Начав с механических и релейных машин, он пришёл к пониманию роли электронных устройств. Цузе стал первопроходцем в теории и практике программирования, в области систем искусственного интеллекта, в создании управляющих вычислительных машин... В истории науки часто случалось, что опередившие время идеи оставались невостребованными современниками. Идеи Цузе были актуальны и могли быть востребованы. Но этому помешали обстоятельства. И хотя его не обошло прижизненное признание (удостоенный многих престижных международных наград, Цузе скончался в 1995 г.), вычислительная техника развивалась своим путём. Кто знает, каким оказался бы этот путь, стань работы Цузе известны своевременно.

АНАЛОГОВЫЕ МАШИНЫ

В отличие от машин Бэббиджа в аналоговых машинах числа заданы какими-либо физическими величинами. Говорят, что аналоговая машина обрабатывает информацию, которая представлена в непрерывной (аналоговой) форме.

Развитие аналоговых машин, как и положено, шло от простых решающих устройств к сложным. Ещё около IV тысячелетия до н. э. в местах, где впоследствии возникло государство Вавилония, при проведении землемерных работ и составлении карт применялись принципы аналоговых вычислений. Около 80 г. до н. э. греки, используя геоцентрическую модель Солнечной системы, построили планетарий. И, таким образом, стало возможным определять положение Солнца и планет. Существует легенда о механическом устройстве для решения логических задач, автором которого якобы являлся философ Платон.



Тем не менее первым аналоговым вычислительным устройством принято считать логарифмическую линейку. Логарифмические шкалы изобрёл английский математик Эдмунд Гюнтер, применявший их для умножения и деления чисел путём сложения и вычитания отрезков. В 1654 г. Роберт Биссакер изготовил первую логарифмическую линейку.

На линейке отрезки представляют собой логарифмы чисел. Поэтому при умножении длины отрезков складывались ($\ln ab = \ln a + \ln b$), при делении — вычитались ($\ln {a}/{b} = \ln a - \ln b$), а при извлечении квадратного корня — делились пополам ($\ln \sqrt{a} = {1}/{2} \ln a$). Если бы шкалы линейки были равномерными, то на ней производили бы только сложение и вычитание.

Графики и *номограммы* можно считать ещё одной разновидностью аналоговых вычислительных устройств. В 1791 г. графики впервые стали использоваться в руководствах по навигации для нахождения значения функции нескольких переменных.

С той поры изобретено множество аналоговых устройств для решения самых разных задач. В XIX столетии применялись наполненные водой сообщающиеся сосуды для решения обычных уравнений. В 1814 г. английский инженер Дж. Герман разработал аналоговый прибор — *планиметр* для

вычисления площадей плоских фигур, ограниченных замкнутой кривой. Планиметр усовершенствовал в 1854 г. швейцарский учёный Якоб Амслер, предложивший интегрирующий прибор с катящимся колесом. А английский физик Джеймс Томсон, изобрёл так называемый *фрикционный интегратор*.

В 1876 г. его родной брат, тоже физик, Уильям Томсон барон Кельвин придумал прибор, определявший время и высоту прилива в различных портах, применив фрикционный интегратор. Вычисления основывались на положении Солнца и Луны, а механизмы, отображающие их, приводились в движение электромотором. Кельвин является и автором машины для решения системы линейных уравнений. Кроме того, он показал возможность решения дифференциальных уравнений путём соединения нескольких интеграторов, однако это не было реализовано из-за низкого уровня техники.

Первую механическую вычислительную машину для решения дифференциальных уравнений построил в 1904 г. русский механик и математик А. Н. Крылов. Он использовал идею польского математика Бруно Абданк-Абакановича, придумавшего в 1879 г. механический аналоговый интегрирующий прибор — интеграф для интегрирования произвольных функций, заданных графически.

В 1909 г. Артур Райт сконструировал устройство для сложения и вычитания, работающее на основе закона последовательного и параллельного соединения электрических сопротивлений в цепи ($R = R_1 + R_2$ — для последовательного, $1/R = 1/R_1 + 1/R_2$ — для параллельного).

Особенно интенсивно велась разработка аналоговых машин во время Второй мировой войны. В те годы были созданы устройства для управления артиллерийским огнём и наведения бомб. Аналоговые устройства хорошо отвечали требованиям армии: выдавать решение сразу же после ввода данных.

Предельная точность аналоговых машин достигала 0,1 %, что, с одной стороны, позволяло успешно решать большинство поставленных задач, а с другой — являлось непреодолимым барьером, так как требовало точности изготовления узлов самой машины.

Установка новой задачи на коммутационной панели машины фирмы «Элиот».

В аналоговых машинах часто числа кодируются величиной напряжения. С помощью обычного потенциометра (переменного сопротивления) производят операции сложения и вычитания, поворачивая ручку прибора на соответствующую величину по или против часовой стрелки. Величина напряжения на выходе потенциометра пропорциональна величине напряжения на входе и углу поворота ручки ($U_{вых} = U_{вх} f(\varphi)$, где φ — угол поворота ручки, а f — функция изменения отношения сопротивления потенциометра к его полному сопротивлению), т. е. потенциометр можно применять как множительное (или делительное) устройство: $z = x f(y)$. Специализированные синусно-косинусные потенциометры позволяют реализовать сразу две функции $z = x \cos y$ и $z = x \sin y$.

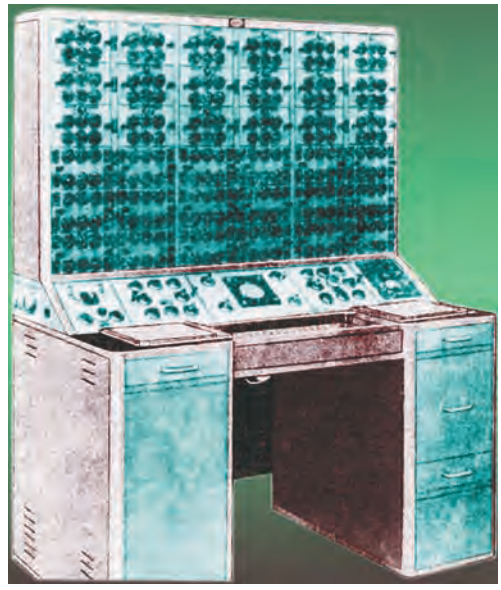
Другой метод перемножения использует пару катушек, при этом сомножителям соответствуют токи в катушках, а произведению — ток между катушками. Прибор для измерения мощности — ваттметр требует перемножения не двух, а трёх величин: $p = UI \cos \varphi$, где φ — фазовый угол между U и I. По одной катушке ваттметра пропускают ток I, по другой — ток, пропорциональный напряжению U.

Номограммы — это специальные чертежи, предназначенные для решения определённого типа задач, например приведённых квадратных уравнений.

Домашний электросчётчик делает ещё более сложную работу, он интегрирует *UI* cos φ по времени, т. е. в каждый момент времени умножает *UI* cos φ на бесконечно малую продолжительность этого момента и суммирует все произведения. Скорость диска соответствует потребляемой мощности, а число оборотов — интегралу по времени. Домашний счётчик — электромеханический прибор. В отличие от него в электронных аналоговых машинах интегрирование обычно производится путём накопления заряда на конденсаторе. В *дифференциальных анализаторах*, предназначенных для решения дифференциальных уравнений, основным процессом является интегрирование по времени.

Аналоговая машина не в состоянии решить задачу в общем виде, из которого потом можно было бы получать частные решения, подставляя новые исходные данные. По принципу действия аналоговые машины могут оперировать только конкретными числами и давать только частные решения.

Усилители постоянного тока являются важным элементом электронных аналоговых машин. На них основаны схемы операций суммирования, вычитания, интегрирования, дифференцирования, изменения шкалы

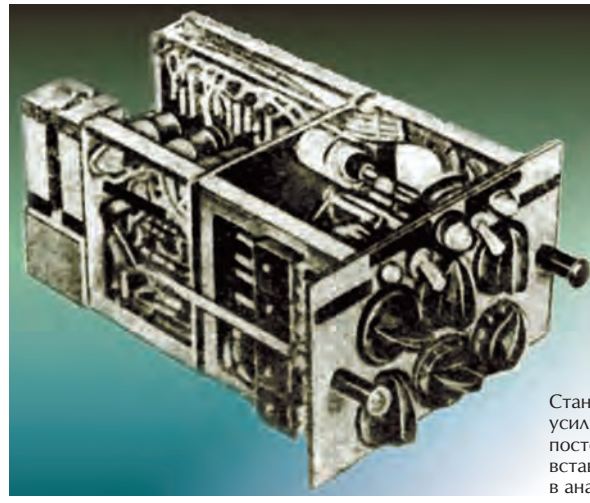


Универсальная аналоговая машина фирмы «Шот Бразерс».

времени и пр. В больших машинах иногда использовалось более 500 усилителей. Подобные приборы смонтированы в специальных ячейках, легко устанавливающихся в машине. Это позволяет собирать любую схему, применяя базовые ячейки усилителей, конденсаторы и сопротивления. Соединения происходят на коммутационной панели (она похожа на панель телефонной станции), на которую выводятся провода от всех участвующих в схеме элементов. Соединения между гнёздами производят проводами с быстросъёмными штекерами. Решение задачи можно получить, включив питание схемы, на точном стрелочном приборе, регистрирующем самописце или электронно-лучевой трубке.

К середине XX в. кроме дифференциальных анализаторов существовал ещё большой класс аналоговых машин, используемых в моделировании. Отдельные элементы машины намеренно соединяются в схемы так, чтобы она представляла модель изучаемой системы. Каждый вычислительный элемент машины решает математическое уравнение, описывающее физическое поведение аналога в системе. Это свойство аналоговых машин позволяет учёным увидеть за математическими формулами действительную физическую систему. Аналоговые машины активно использовались в авиации для моделирования аэродинамических систем и систем управления самолётом.

Аналоговые машины применяют и сегодня. Во многих автомобилях имеется прибор, объединяющий в себе два типа устройств — цифровое и ана-

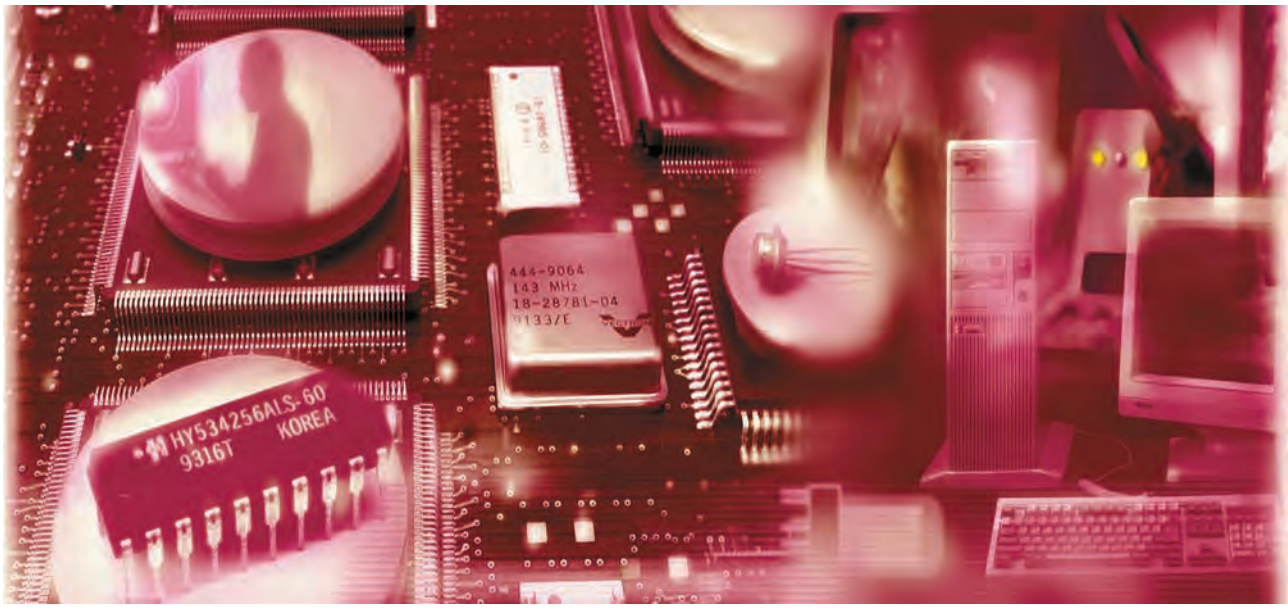


Стандартный блок усилителя постоянного тока, вставляемый в аналоговую машину.

логовое. Это всем известный спидометр, аналоговый прибор, где стрелка указывает скорость движения автомобиля. На табло спидометра чуть ниже центра находятся окошки одометра — цифрового счётчика пройденного расстояния.

В конце XX в. некоторые производители автомобилей предприняли попытку заменить панель приборов целиком на цифровую. Однако человеку естественнее видеть аналоговые шкалы, ведь и цифровые часы не вытеснили часы с привычным стрелочным циферблатом и, вероятно, не вытеснят никогда.

Возможно, и аналоговые машины ждёт второе рождение, ведь квантовые компьютеры тоже относятся к этому классу машин.



ЭВОЛЮЦИЯ КОМПЬЮТЕРА В XX ВЕКЕ

ПЕРВОЕ ПОКОЛЕНИЕ КОМПЬЮТЕРОВ

MARK I

В конце 30-х гг. XX в. многочисленные устройства, использующие перфокарты в роли носителей закодированной информации, стали достаточно надёжны и по меркам того времени обладали хорошей скоростью считывания (порядка нескольких перфокарт в минуту). Их начали применять для хранения информации и в качестве устройства ввода-вывода, когда учёные приступили к созданию цифровых машин-компьютеров. Одна из таких машин была сделана группой учёных из лаборатории фирмы IBM под руководством Говарда Айкена и получила название Mark I. Построили её из стандартных электромеханических частей, применяемых в калькуляторах. Mark I мог выполнять не только четыре основных арифметических действия с 23 разрядными десятичными числами, но и специальные встроенные алгоритмы для вычисления тригонометрических функций и логарифмов. Машина «программировалась» с перфоленты (информация на перфоленте кодировалась пробиванием отверстий в определённых местах, как и на перфокартах), которая двигалась только вперёд. Поэтому выполнение циклов или передача управления назад были невозможны.

Результат выдавался на перфоратор или обычную электрическую пишущую машинку. Mark I использовал для выполнения арифметических операций вращающиеся диски-счётчики (как в арифмометре) и для некоторых функций— *электромагнитные реле,* так что машину уже можно было классифицировать как «релейный» компьютер. Его габариты ужасают! Mark I размещался вдоль 20-метровой стены и весил около 5 т. Машина работа-

Говард Айкен и Грейс Хоппер.

Электромеханическая вычислительная машина Mark I.

ла по современным меркам медленно (ей требовалось от 3 до 5 с для умножения), но полностью автоматически и могла использоваться для длинных вычислений, где отсутствовали циклы (их приходилось вручную разбивать на последовательность операций). Так, вычисление n! = n(n2)...1 состояло из n–1-строки умножения в отличие от нескольких строк вычисления n! в цикле:

```
m:=1
факториал: = 1
нц пока m < n + 1
| факториал: = факториал * m
| m:=m + 1
кц
```

ENIAC

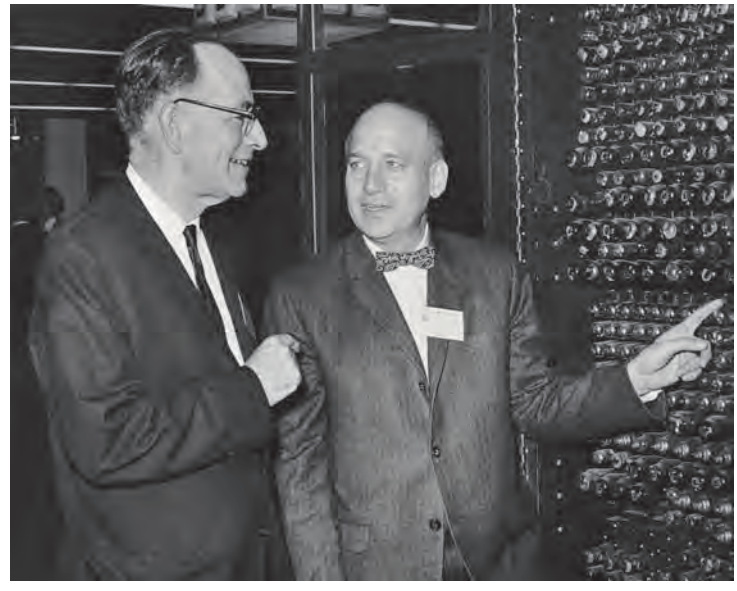
Вторая мировая война вынуждала вести постоянные разработки автоматических вычислительных машин. Военная техника требовала быстрых математических расчётов, например для систем наведения при управлении зенитным огнём. Механические калькуляторы не могли обеспечить нужной скорости вычислений, поэтому военные настаивали на проведении скорейших разработок и немедленной постройке электронных вычислительных машин. В 1942 г. Преспер Экерт и Джон Моучли со своими сотрудниками-единомышленниками в Школе электрических разработок Университета штата Пенсильвания (США) задумали создать быстродействующую ЭВМ, получившую название ENIAC (аббревиатура определения, переводится с английского как «электронный числовой интегратор и вычислитель»).

Несмотря на то что размер машинного слова был всего 10 десятичных цифр (у Mark I— 23), ENIAC производил 300 операций умножения за одну секунду и около 5000 сложений и вычитаний! Такой производительности удалось достичь при помощи хранения в памяти машины готовых результатов таблиц умножения. Вместо «медленных» реле ENIAC использовал 18 тыс. электронных вакуумных ламп, а для ввода-вывода закодированной информации— хорошо знакомые перфокарты.

Как и многих современных вычислительных машин, у ENIAC вычислитель состоял из нескольких блоков-устройств: один блок складывал и вычитал, другой умножал, третий делил и даже мог извлечь квадратный корень и т. д. Кроме того, имелось ещё 20 десятичных *регистров*-счётчиков, которые использовались для сложения и временного хранения результатов. Чтение чисел из регистров и запись в них— так называемое время выборки из регистра— происходили за 0,2 миллисекунды. Благодаря тому что время доступа к числам, содержащимся в памяти, на несколько порядков превышало время доступа к регистрам, можно было хранить промежуточные результаты в регистрах,

ENIAC изумил бы сегодняшних школьников не столько своими возможностями, сколько размерами. Он размещался на площади более 150 м^2 и потреблял свыше 180 кВт электроэнергии!

Джон Моучли и Преспер Экерт.



Компьютер ENIAC.

что уменьшало время счёта. Например, если надо сложить 123 + 102 + 45, то 123 и 102 выбираются из памяти, результат суммы первых двух слагаемых хранится в регистре. При сложении 225 + 45 лишь 45 надо «достать» из памяти.

Программа вычислений на ENIAC задавалась вручную с помощью механических переключателей и гибких кабелей со штекерами, которые вставлялись в нужные разъёмы (кабельные соединения), что сильно напоминало телефонные станции начала XX в. Фактически программы на этой машине не записывались, а «навтыкивались».

Электронные лампы.

Изменение программы вычислений требовало немалых (в том числе и физических) усилий. Ещё до окончания постройки ENIAC машиной заинтересовался видный американский математик Джон фон Нейман, принявший участие в работе команды Моучли — Экерта. Он произвёл существенное усовершенствование машины, предложив создать блок со стандартным набором кабельных соединений, куда входили все команды вычислений и все функции. Управлять процессом вычислений стала программа, хранящаяся в выделенной области памяти. Она представляла собой набор двоичных чисел, а поскольку была малопонятна неспециалисту, то получила название *машинной программы.* Каждая из её команд соответствовала определённой функции, т. е. определённому кабельному соединению из блока соединений. Теперь для загрузки программы не требовалось производить новые кабельные соединения или убирать старые. Оставалось лишь поместить новую программу в память. Таким образом, фон Нейман сформулировал принцип, лежащий в основе функционирования современных вычислительных машин: в памяти ЭВМ содержатся не только обрабатываемые числа, но и сама программа; и то и другое хранится в виде многозначных двоичных чисел. Внушительный масштаб решённых проблем и широкие возможности применения ENIAC положили начало первому поколению компьютеров.

Позднее, в 1971 г., было оспорено первенство ENIAC как первой цифровой вычислительной машины. Более простой вычислитель, построенный под руководством Джона Атанасова в 30-х гг., использовал те же принципы конструирования электронных переключателей на вакуумных лампах. В 1973 г. состоялось судебное слушание по этому вопросу, и суд принял решение, что патентные права на основные идеи цифровых электронных машин принадлежат Джону Атанасову.

Тем не менее ENIAC можно назвать первым удачным быстродействующим электронным цифровым компьютером, который успешно работал с 1946 по 1955 г.

IAS

После Второй мировой войны Джон фон Нейман приступил к разработке собственного компьютера, который получил название IAS (от *англ.* Institute for Advanced Studies — Институт передовых исследований). Он впервые был представлен в 1952 г. в Принстоне (США).

Машинные слова стали измеряться в битах, т. е. двоичных разрядах, в отличие от десятичных разрядов в ENIAC. В этом «виноваты» инженеры, которым удобнее было проектировать память вычислительных машин из элементов, имеющих два состояния, как в электрической (не путать с электронной) лампочке: «горит» или «не горит». Машинное слово в IAS составляло 40 бит. Именно такое количество информации могло передаваться между памятью и процессором за одну передачу.

Память состояла из 4096 таких слов. В соответствии с идеями фон Неймана слово, записанное в память, может представлять собой либо команду процессору *(инструкцию),* либо данные (например, для вычислений). Инструкция, как правило, состояла из числового кода команды (8 бит) и адреса ячейки памяти (12 бит). Вся инструкция — в два раза меньше, чем машинное слово IAS (20 бит).

Команда	Адрес
8 бит	12 бит

Таким образом, число команд машины IAS не более $2^8 = 256$, а размер адресуемой памяти (называемый ещё математическим адресным пространством), $212 = 4096$, совпадает с реальным физическим размером памяти, т. е. в команде IAS можно указать одну из ячеек памяти, содержимое которой будет использоваться в операции.

При сложении одно слагаемое помещается в специальном регистре, предназначенном для суммирования, — *аккумуляторе,* а второе — где-то в памяти. Пусть надо сложить число 28, находящееся в аккумуляторе, и число 33, находящееся в 342-й ячейке памяти. Тогда в 20-битной команде будут записаны двоичный код операции «сложить» и адрес второго слагаемого «342», представленный в двоичной записи: 000101010110. Чтобы сложить два числа, помещающиеся в памяти, первое сначала надо «загрузить» в регистр (как бы провести операцию сложения с нулём), а затем уже выполнить команду сложения, т. е. потребуются две команды.

Компьютеры IAS.

Команда	Адрес
Сложить	342

Память	Адрес
	341
33	342
	343
	...

Регистры процессора использовались, чтобы хранить данные и результаты команд. Процессор при выполнении команды в качестве аргумента автоматически получал тот регистр, который требовался. Другими словами, «адрес» регистра определялся самой командой.

Процессор IAS компьютера состоял из *блока обработки данных* (арифметические операции и др.) и *управляющего устройства,* которое и осуществляло выполнение программы. Процессор также содержал быстродействующие регистры со «странными» именами— *AC*, *MQ*, *DR*, *IBR*, *PC*, *IR*, *AR*, предназначенные для временного хранения инструкций, адресов иданных. Для синхронизации работы всех устройств машины использовались электронные часы. Они подавали электрические сигналы через равные промежутки времени, называемые *машинными тактами.* За один такт можно было, например, прочитать из памяти 40-битное слово в 40-битный регистр *DR* (или записать в память), при этом

```
нц
| если очередная инструкция не в буфере инструкций (IBR)
| | то прочитать из памяти с адресом (PC) в (DR)
| | занести инструкцию в буфер инструкций (IR)
| | занести следующую инструкцию в регистр инструкций (IBR)
| | увеличить счётчик команд (PC) на единицу
| всё
| выполнить команду
кц
```

Регистр *DR* ещё носит название *MBR* (memory buffer register — «буферный регистр памяти»), а регистр *AR* называют *MAR* (memory address register — «регистр адреса памяти»).

регистр *AR* содержал 12-битный адрес машинного слова. Регистры *AC* (аккумулятор) и *MQ* (множитель) использовались для временного хранения данных и результатов операции. Эти регистры участвовали в суммировании и операциях умножения и деления.

Так как в одном машинном слове помещались две инструкции, то при выполнении они одновременно загружались из памяти. Первая инструкция выполнялась, помещаясь врегистр инструкций *IR*. Вторая помещалась в «буфер инструкций» — регистр *IBR* и, возможно, выполнялась на следующем шаге. Счётчик команд *PC* (*англ.* program counter — «программный счётчик») хранил адрес следующей инструкции программы впамяти.

Выполнение программы IAS состояло в циклическом исполнении компьютером двух шагов — подготовительного и основного (называемого ещё шагом исполнения). На подготовительном шаге происходила загрузка инструкции, на основном — её исполнение.

Машина IAS увеличивала счётчик команд *PC* всякий раз, когда следующей инструкции не было в буфере инструкций *IBR*. По окончании подготовительного шага следовал основной шаг, и все последующие действия машины зависели от результатов последней инструкции. Могла встретиться инструкция перехода к «другому месту» в программе, выполнение которой состояло в изменении значения счётчика команд.

Инструкции программы хранились в памяти приблизительно в той же последовательности, в которой потом исполнялись. Таким образом, в IAS использовались основные принципы, повлиявшие на все последующие цифровые машины:

- наличие арифметического устройства для выполнения арифметических действий;
- расположение программы и данных в общей памяти;
- стандартный цикл выполнения программы;
- последовательное расположение элементов программы впамяти;
- регистры (маленькая быстрая память).

Машина IAS работала эффективно — выполняла умножение за 100 микросекунд, а доступ к памяти (чтение из памяти и запись) осуществлялся за 50 микросекунд.

Джон фон Нейман в 1954 г. предложил основы алгоритмического языка, который нашёл применение гораздо позднее в популярных языках программирования. Гениальный учёный не прекращал работ по конструированию вычислительных машин, будучи консультантом в IBM. Его компьютер IAS можно назвать основным представителем компьютеров первого поколения.

Джон фон Нейман.

ДЖОН ВИНСЕНТ АТАНАСОВ

В1971г. в Федеральном суде США началось разбирательство по установлению авторства электронной вычислительной машины. Ответчиком была фирма HONEYWELL — обладатель патента на машину ENIAC. Процесс длился два года и в 1973 г. закончился победой Джона Винсента Атанасова.

Джон Винсент Атанасов появился на свет 4 октября 1903 г. на севере США, под Нью-Йорком, в семье болгарских эмигрантов. (Фамилия Атанасов имиграционными властями в 1989 г. была изменена на Атанасофф.)

Любовь к математике Джон унаследовал от матери, Айвы Люцены

Парди, преподававшей этот предмет в школе. Отец научил сына разбираться вэлектрике: девятилетний мальчик легко находил и исправлял повреждённую проводку в доме.

В 1921 г. Атанасов поступил в Университет штата Флорида, а в 1925 г., окончив его, получил степень бакалавра. Его приглашали преподавать во многие учебные заведения, даже в Гарвард, но Атанасов выбрал Университет штата Айова.

В 1929 г. Атанасов получил место в докторантуре в Мэдисоне, в 1930 г. защитил докторскую диссертацию, посвящённую электронной структуре атомов гелия. Работа была связана с большими объёмами вычислений, и это натолкнуло Атанасова на мысль создать электронное устройство для автоматизации вычислений. Вскоре он получил место профессора-ассистента по математике и физике в Университете штата Айова, куда Джон возвратился с намерением построить компьютер. Там он провёл эксперименты с электронными вакуумными лампами и изучил разработанные к тому времени устройства для математических расчётов. Атанасов условно разделил их на аналоговые и «свойственные вычислительным машинам», т. е. цифровые (термин «цифровые» появился позже). В 1936 г. Атанасов вместе с физиком-атомщиком Гленом Мэрфи построили модель небольшого аналогового калькулятора для анализа геометрических поверхностей.

В декабре 1939 г. Атанасов и его ассистент Клиффорд Берри создали первый образец машины, которая предназначалась для решения линейных уравнений. Они назвали её ABC (Atanasoff Berry Computer). Весила машина более 300 кг и состояла из:

• двух барабанов по 30 50-разрядных двоичных слов памяти, к каждому барабану можно было обращаться независимо. Запоминающее устройство использовало конденсаторы с автоматическим восстановлением заряда (сейчас это называется DRAM — «динамическая память»);

• блока управления, собранного на 300 электронных лампах и обеспечивающего поразрядное сложение и вычитание чисел при тактовой частоте 60 Гц (примерно одно сложение за секунду);

• перфокарт, которые использовались в качестве вторичной памяти, они вставлялись вручную (дырки на перфокартах не пробивались, как обычно, а прожигались со скоростью до 1500 бит данных в секунду с помощью свечек, похожих на автомобильные. Эта технология была настолько удачной, что ошибка при вводе случалась только 1 раз на 100 тыс. чисел).

Попытки запатентовать изобретение оказались безрезультатными. После демонстрации ABC Атанасов получил от Университета штата Айова 650 долларов на построение компьютера. В январе 1941 г. в газете «Des Moines Tribune» появилась небольшая заметка, в которой сообщалось, что «Д-р Джон Атанасов, профессор физики Университета Айовы, строит ЭВМ, которая по принципу своей работы ближе к человеческому мозгу, чем любая другая машина». Однако тогда на это обратили внимание только специалисты. (При создании машины ENIAC разработчики использовали идеи Атанасова.) Но из-за Второй мировой войны работы над машиной были прекращены.

Мысли о создании цифровой машины не покидали Атанасова ни на минуту. Как-то после очередных неудач он бросился в автомобиль и, в исступлении проехав 300 км, внезапно остановился у неприметного придорожного кафе. Там, успокоившись, Атанасов нашёл долгожданное решение, окончательный вариант концепции цифровой машины.

Джон Винсент Атанасов.

Клиффорд Берри.

◄ Часть компьютера ABC.

В августе 1996 г. Гарри Слидж и Джон Эриксон, используя идеи Атанасова, построили компьютер ABC-2, который предназначался для решения системы из 5 линейных уравнений с 5 неизвестными. При сборке они использовали около 500 вакуумных ламп, десятки метров электрического кабеля и старый перфоратор фирмы IBM.

Во время войны Атанасов возглавлял управление акустики лаборатории Военно-морских сил США в штате Вашингтон. Одновременно он участвовал в проекте испытаний атомной бомбы в Тихом океане. Когда в 1948 г. Атанасов вернулся в Университет Айовы, то обнаружил, что его компьютер демонтирован и все чертежи и диаграммы машины утеряны. Удалось спасти лишь немногие части машины.

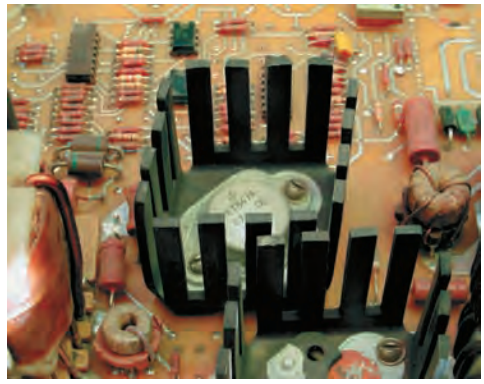
Умер Джон Винсент Атанасов 15 июня 1995 г., так и не получив патента на ABC.

Оригинальные идеи Атанасова предвосхитили инженерные решения, положенные в основу универсальных ЭВМ. Компьютер Атанасова, в отличие от электронномеханических машин 40-х гг. XX столетия, был полностью собран из электронных элементов.

ВТОРОЕ ПОКОЛЕНИЕ КОМПЬЮТЕРОВ

Компьютеры второго поколения производились приблизительно с 1955 по 1964 г. В 1948 г. специалистами американской компании AT&T Bell Laboratories (Bell Labs) был изобретён транзистор, и его авторы — Джон Бардин, Уолтер Браттейн и Уильям Шокли — получили патент на это изобретение. Данное открытие в очередной раз перевернуло мир. Транзистор полностью вытеснил вакуумные лампы из конструкции ЭВМ. Он занимал в десятки раз меньше места, выделял меньше тепла, потреблял меньше электроэнергии, работал более надёжно. Переход с технологии вакуумных ламп на транзисторную считается особенностью машин второго поколения.

Усовершенствовалась и память машин. Электронно-лучевые трубки и ультразвуковые линии задержки со временем были заменены ферритовыми сердечниками и магнитными барабанами. Изменения произошли и в процессоре — сердце вычислительной машины. Увеличился набор команд. Стало возможным читать иписать не только по адресам в памяти, указанным в команде, но и вычислять их, например складывая значение специального, *индексного,* регистра с числом.



Транзистор.

Были разработаны специальные процессоры IO (*англ.* input-output — «ввод-вывод») для управления вводом-выводом и передачей информации. Ими командовал центральный процессор, при этом ничего не передавая.

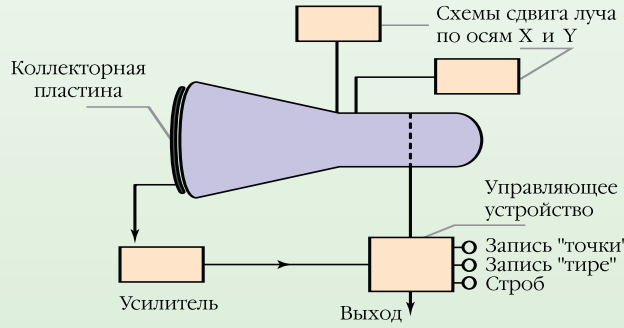
Оказалось, что для математических расчётов не хватает четырёх основных арифметических действий (сложения, вычитания, деления и умножения), которые выполнялись, как правило, лишь с целыми числами. Поэтому дополнительно были разработаны специальные процессоры для эффективного выполнения арифметических действий над действительными числами (числами с *плавающей запятой*).

Программирование на машинном языке процессора очень трудоёмкая задача. И программы, работавшие на одних ЭВМ, было тяжело адаптировать для выполнения на других, так как языки ЭВМ отличались друг от друга. Нужны были новые машинно-независимые языки программирования. Появились языки высокого уровня: FORTRAN, созданный Джоном Бэкусом в 1954—1957 гг. в IBM, и Algol (*англ.* Algorithmic Language), первая версия Algol-58 опиралась на идеи синтаксиса FORTRAN.

Началась разработка программного обеспечения для вычислительных

ЗАПОМИНАЮЩИЕ ЭЛЕКТРОННО-ЛУЧЕВЫЕ ТРУБКИ И УЛЬТРАЗВУКОВЫЕ ЛИНИИ ЗАДЕРЖКИ

В компьютерах первого поколения запоминающей средой часто являлся люминофор экрана электронно-лучевых трубок. Запись и чтение в отдельные биты памяти производились с помощью электронной пушки. Пушка при записи «зажигала» небольшую область экрана в форме тире, устанавливая значение бита в 1, или в форме точки для записи 0. Для считывания хранящейся информации используется вспомогательный электрод, который находится в виде тонкого проводящего слоя на внешней части экрана трубки. Луч электрода направляется в считывающую область (бит), на вспомогательном электроде возникает положительный импульс, если в данном месте было «засвечено» тире, или потенциал электрода не меняется, если была записана точка, т. е. 1 или 0. Обычный объём памяти трубки до смешного мал — 1024 или 2048 бит. Однако в 50—60-х гг. XX в. такая память являлась основной для ЭВМ.



Альтернатива памяти на электронно-лучевых трубках — память на ультразвуковых линиях задержки, где в качестве запоминающей среды были электрические ультразвуковые импульсы, непрерывно «путешествующие» от одного конца трубки с ртутью до другого. Это напоминало волну, бегущую от берега к берегу. Электрический сигнал запускал её. Точно такой же сигнал получали на другом конце трубки, когда волна достигала его. Таким образом, сигнал запоминался на время прохождения по ультразвуковой линии задержки. В одном таком элементе могло храниться 400-значное двоичное число.

машин. В первую очередь это относилось к системам так называемой *пакетной обработки* (когда задачи для машины собирались в пакеты: ввод задач, обработка, вывод), которые являлись зачатками современных операционных систем. Первым компьютером, использующим в качестве элементной базы транзисторы, стала экспериментальная машина TX-O (*англ.* Transistorized Experimental Computer— «транзисторный экспериментальный компьютер»). О производстве TX-O заявили в 1953 г.— машину построили в Массачусетсском технологическом институте.

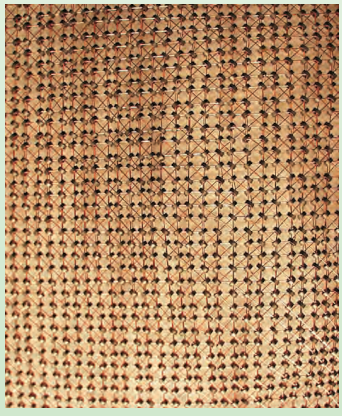
Компьютер IBM 704 на электронных вакуумных лампах имел процессор с индексными регистрами и выполнял действия над числами с плавающей запятой. Более поздняя модель этого ряда— IBM 709 обладала процессорами ввода-вывода, названными впоследствии *каналами,* или *канальными процессорами.* Они осуществляли операции по обмену информацией, а управлял ими центральный процессор, это получило название программирования ввода-вывода.

Первые машины второго поколения фактически являлись гибридными, т. е. создавались с использованием и старой, и новой технологий.

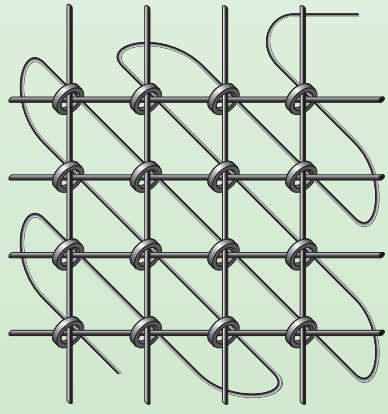
Компьютер TX-O.

ФЕРРИТОВЫЕ СЕРДЕЧНИКИ И МАГНИТНЫЕ БАРАБАНЫ

Магнитный (ферритовый) сердечник представлял собой кольцо небольшого диаметра. Он легко намагничивался и долго сохранял такое состояние. Сердечники нанизывались на провод наподобие бус. Постоянный ток (около 0,5 А) мог изменить намагниченность сердечника, если был пропущен через провод в одном направлении, или оставлял её без изменений, если имел противоположное направление (намагниченность в одном направлении — 0, в другом — 1). Если составить матрицу сердечников, пропустив провода через каждый вертикальный и горизонтальный ряды, и пустить по ним ток силой в половину нормы (0,25 А), то можно изменить намагниченность только одного сердечника (0,25 + 0,25 = 0,5), находящегося на пересечении горизонтальной и вертикальной линий. Остальные не будут реагировать, так как ток меньше 0,5 А.Чтобы определить значение сердечника, его переводят всостояние 0 (намагниченность одного направления). Если он уже в этом состоянии, то ничего не произойдёт. Переход в другое состояние вызовет возникновение тока в дополнительном проводе, проходящем через все сердечники на матрице (провод считывания).



Накопительная матрица на магнитных сердечниках.

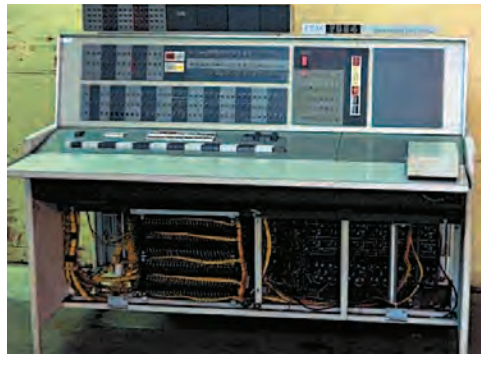


Увеличенный фрагмент матрицы ферритовых сердечников.

Магнитные барабаны имели тонкий слой (около 0,025 мм) магнитного материала по поверхности немагнитного барабана. Двоичные цифры записываются на вращающийся с постоянной угловой скоростью барабан путём намагничивания маленьких участков, когда те проходят мимо головки записи-чтения. Записанное число может быть считано через один оборот барабана. На вращающемся со скоростью 6000 оборотов в минуту магнитном барабане диаметром 15 см и длиной 30 см при плотности записи 60 цифр на 1 см^2 можно записать около 4 кбайт.

IBM 7094

Модель IBM 7094 — научный компьютер второго поколения IBM был построен целиком на транзисторах. В середине 60-х гг. XX в. он являлся одним из самых быстрых компьютеров на рынке: выполнял до 350 тыс. операций в секунду над числами с плавающей запятой. Стоимость машины составляла около 3,5 млн долларов США. Стандартный IBM 7094 имел 32 × 1024 = = 32 768 36-битных слов памяти на ферритовых сердечниках, каждый из сердечников мог запомнить один бит информации.



Компьютер IBM 7090.

В качестве внешней памяти использовался накопитель на жёстких магнитных дисках модели IBM 2302. Жёсткие магнитные пластины диска размером 24 дюйма в диаметре вращались в вакууме. Магнитные головки считывали и записывали информацию на магнитную поверхность дисков (так же как пишут и воспроизводят звук на магнитной кассете). (Одна из последних в этом ряду моделей — IBM 2302 могла хранить более 200 Мбайт.) Процессор IBM 7094 отличался от процессора IAS наличием индексных регистров и арифметического устройства, которое выполняло все операции с

◄ Устройство магнитного диска.

действительными числами (с плавающей и фиксированной запятой). Канальные процессоры осуществляли прямой доступ к памяти. Так как память не могла быть одновременно доступна центральному процессору и каналам, было предусмотрено специальное управляющее устройство, переключающее доступ к памяти между процессорами.

Большинство регистров процессора IBM 7094 подобны процессору IAS, например:

- *IC* (*англ.* instruction counter)— счётчик команд, аналог *PC*;
- *AR* (*англ.* address register)— адресный регистр, содержит адрес памяти, указанный в инструкции;
- *SR* (*англ.* storage register)— буферный регистр, аналог *DR* и *IBR*;
- *RI* (*англ.* register instruction)— регистр инструкций, содержит текущий код операции.

За один такт процессор загружал две команды-инструкции из памяти, вторая инструкция запоминалась в регистре *SR*, который был скрыт от программ и имел служебное назначение.

ИНДЕКСНЫЕ РЕГИСТРЫ

Обычно в команде указывается адрес ячейки памяти (откуда надо взять или куда надо положить данные). Однако если требуется заполнить некоторую таблицу *(массив)* из 10 элементов нулями, выполняют 10 одинаковых операций, где явно указан адрес каждой ячейки таблицы:

Положить 0 в ячейку №101
Положить 0 в ячейку №102
...
Положить 0 в ячейку №110

Если в процессоре отсутствуют индексные регистры, то другого пути нет. Однако индексный регистр позволяет вычислить адрес ячейки, складывая число и значение, которое находится в индексном регистре. Построим цикл:

```
нц для Индексный регистр от 1 до 10
| Положить 0 в ячейку по адресу
| Индексный регистр + 100
кц
```

У IBM 7094 было семь индексных регистров (*XR*), каждый из которых мог хранить 15-битный адрес.

Команда	Признак	Адрес
18 бит	3 бит	15 бит

Для того чтобы процессор знал, какой из индексных регистров используется при вычислении адреса, в команде выделено место — 3 бит (2^3=8), что позволяет закодировать *признак адресации* — число от 0 до 7. Число 1 соответствует первому регистру, 2 — второму, ... , 7 — седьмому, по числу индексных регистров. Если в признаке стоит 0, то адресация обычная, без суммирования с индексным регистром.

Например, в команде используется четвёртый индексный регистр (при этом 3-битный признак равен 100_2, или 4_{10}). Тогда содержимое четвёртого индексного регистра складывается с адресом, который находится в инструкции (15 бит). Это и будет адресом в памяти.

Вот пример сложения числа 28, находящегося в аккумуляторе, и числа 33, находящегося в 342-м слове памяти.

Команда	XR	Адрес
Сложить	4	2

Индексный регистр XR 4

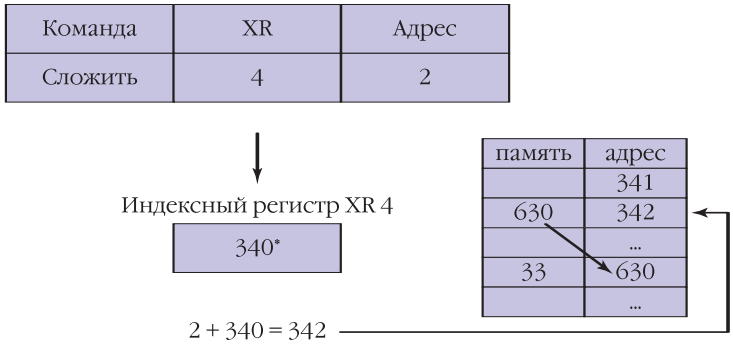
340

память	адрес
	341
33	342
	343
	...

2 + 340 = 342

КОСВЕННАЯ АДРЕСАЦИЯ

Предыдущий пример относился к *прямой адресации,* когда данные для команды расположены по указанному адресу. В нашем примере при *косвенной адресации* (на что указывает знак *) в 342-м слове содержится *адрес аргумента.* Подобно тому как, придя к другу Пете домой, застанешь не его самого, а записку, в которой сказано, что он пошёл к Васе. Если тот находится у Васи, то это косвенная адресация первого уровня. Если же и у Васи записка, что они с Петей у Тани, то это косвенность второго уровня и т. д. Реально в процессорах используется косвенность трёх уровней.



Индексная и косвенная адресации незаменимы при работе с массивами и таблицами.

Всего у машины IBM 7094 насчитывалось более 200 команд, которые можно разделить на несколько классов:

• инструкции пересылки (передачи) данных, служащие для обмена информацией между процессором и памятью или между двумя регистрами процессора;

• арифметические инструкции: сложение, вычитание и т. д. (в том числе и с плавающей запятой);

• логические (не числовые) инструкции типа операции ИЛИ, НЕ и т. д.;

• инструкции для изменения значений индексных регистров, такие, как увеличение на 1, уменьшение на 1 и др.

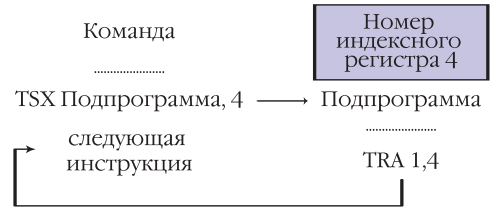
• условный и безусловный переход (передача управления): изменение порядка выполнения программы, циклов;

• операции ввода-вывода: для передачи данных между устройствами ввода-вывода и памятью компьютера.

ПОДПРОГРАММЫ

Важная особенность машин второго поколения — наличие специальных инструкций для перехода к подпрограммам (вспомогательным алгоритмам) и возврата из них. В IBM 7094 для вызова подпрограммы использовалась специальная инструкция — TSX, обратная (возврат из подпрограммы) инструкция — TRA.

При выполнении команды TSX в индексный регистр, применяемый в качестве аргумента, запоминался адрес текущей инструкции (адрес возврата). Затем начиналось выполнение подпрограммы (т. е. происходила передача управления). При завершении в ней должна была находиться инструкция TRA, где в качестве аргумента присутствовал тот же индексный регистр, что и в команде TSX. Происходил возврат в программу.



ВВОД-ВЫВОД

Канальные процессоры (каналы) выполняют специальные программы — программы ввода-вывода. Центральный процессор производит запуск канала и «присматривает» за операциями ввода-вывода (осуществляет их мониторинг). Хотя центральный процессор сам в процессе ввода-вывода не принимает участия, он может выставить запрос на прерывание — остановку текущей работы. Обычно это происходит по окончании операции ввода-вывода, когда получены данные, ожидаемые центральным процессором. Аппарат прерываний, появившийся у некоторых компьютеров второго поколения, сущест-

венно повысил их производительность..

Каждый канальный процессор может одновременно управлять десятью магнитными лентами, печатающим устройством, устройством ввода с перфокарт и даже дисководом с жёстким магнитным диском. Канал, как и центральный процессор, имеет похожий набор скрытых регистров. Естественно, у него отсутствуют обычные регистры, предназначенные для арифметических операций (аккумулятор и др.), и индексные регистры. Канальные процессоры служат для унификации ввода-вывода.

В IBM 7094 данные между канальным процессором и основной памятью передавались словами по 36 бит, а между каналами и устройствами ввода-вывода, такими, например, как магнитные ленты,— уже по 6 бит. Поэтому каналы осуществляли сборку поступающих данных (по 6 бит) в слова по 36 бит. Роль буфера для сборки-разборки выполнял 36-битный *регистр данных.* Число слов, которое должно передаваться в процессе ввода-вывода между памятью и каналом, запоминалось в регистре-счётчике. Состояние ввода-вывода содержалось в *статусном регистре.* Если вдруг возникали ошибки в процессе ввода-вывода, то центральный процессор мог узнать о случившемся по коду, который содержался в этом регистре.

Процесс ввода-вывода происходил следующим способом:

1. Центральный процессор встречает инструкцию ввода-вывода. По ней он определяет, где находится программа ввода-вывода.

АППАРАТ ПРЕРЫВАНИЙ

В классе 20 учеников выполняют контрольную работу (ученики — это каналы, учитель — центральный процессор, контрольные работы — программы каналов или инструкции). Учитель ожидает окончания выполнения контрольных работ и одновременно заполняет журнал. В случае отсутствия аппарата прерываний учитель через равные промежутки времени приостанавливает заполнение журнала и спрашивает у каждого отдельного ученика о готовности его работы. Те, кто ответил на вопрос утвердительно (закончил выполнение своих программ), передают учителю тетрадки (данные ввода-вывода находятся в памяти). Остальные продолжают выполнять свои контрольные работы. Очевидно, что при этом механизме учитель тратит много времени на опрос учеников.

Напротив, при наличии аппарата прерываний учитель даёт ученикам контрольную работу и просит, чтобы они позвали его по окончании выполнения, подняв руку. В этом случае сам он просто заполняет журнал, никого не опрашивая. Когда кто-нибудь из учеников заканчивает работу, то поднимает руку, подавая сигнал (какой-то из процессов ввода-вывода завершён). Учитель приостанавливает заполнение журнала и занимается контрольной ученика.

2. Центральный процессор передаёт в качестве параметра имя устройства ввода-вывода и запускает программу.

3. Канальный процессор выполняет программу и по окончании действий (правильных или с ошибкой) сохраняет код завершения в статусном регистре. Затем посылает прерывание центральному процессору, сообщая о выполнении ввода-вывода.

БОЛЬШИЕ СИСТЕМЫ

На ранних стадиях все программы работали автономно и компьютер надо было останавливать для ручной перенастройки на каждую новую программу. В машинах второго поколения появилась возможность подготовки пакета заданий-программ, предварительно сохранив их, например, на магнитной ленте и затем обработав на компьютере. Результат при этом записывался на другую магнитную ленту. Такой метод получил название пакетного метода обработки. Для подготовки ленты с заданиями стали пользоваться вспомогательным «небольшим» компьютером. Пакетный метод обработки требовал наличия специальной программы-монитора, постоянно находящейся в памяти компьютера, которая управляла потоком заданий. Управление велось в режиме мультипрограммирования: процессор временно приостанавливал выполнение текущей программы, начинал следующую и позже возобновлял прерванную программу. Это повышало эффективность работы компьютера, особенно если текущая задача ожидала завершения ввода-вывода и её решение временно прерывалось, а центральный процессор бездействовал. Мультипрограммирование получило развитие в компьютерах последующих поколений.

Команды ввода-вывода делятся на 3 группы:

• Инструкции, управляющие устройствами ввода-вывода.

К ним относятся, например, команды на перемотку магнитной ленты назад или позиционирование головки дисковода на указанное место перед чтением/записью.

• Инструкции передачи данных.

Они имеют вид «передайте *n* слов между устройством ввода-вывода и основной памятью». Каждая такая инструкция содержит некое количество слов *n*, которое сохранится в счётчике (регистре-счётчике) канала, и адрес основной памяти (куда или откуда поступят данные).

После этого осуществляется передача данных:

```
нц пока счётчик > 0
| передать слово между каналом и
|    памятью (адресный регистр)
| счётчик:= счётчик - 1
| адресный регистр:= адресный
|    регистр + 1
кц
```

• Управляющие инструкции или инструкции условного и безусловного перехода.

Каналы и центральный процессор разделяют между собой основную память компьютера. Управление осуществляется при помощи специального устройства и набора линий связи (проводов), идущих одновременно к процессору и каналам. Набор линий связи называется системной шиной. Операции ввода-вывода проходят медленно по сравнению со скоростью обработки данных центральным процессором. Поэтому память доступна в основном ему, а не каналам.

К 1965 г. около 300 компьютеров IBM 7090 и IBM 7094 были установлены в различных компаниях, несмотря на достаточно солидную стоимость.

Одну из IBM 7094 Вооружённые силы США «сняли с дежурства» в середине 80-х гг., после 30 лет бесперебойной работы.

ТРЕТЬЕ ПОКОЛЕНИЕ КОМПЬЮТЕРОВ

1965 год считается годом появления компьютеров третьего поколения. Различия между вторым и третьим поколениями кажутся не такими большими, как между первым и вторым.

Производство вычислительных машин было хлопотным и дорогим делом. Огромное количество основных элементов ЭВМ— транзисторов надо было собрать воедино; бесчисленное множество проводов опутывало блоки и части компьютера. Сложность ЭВМ росла с каждым днём, это приводило к увеличению числа транзисторов, трудностям проектирования и монтажа.

Революцию в технологии производства ЭВМ третьего поколения вызвало создание интегральных схем. В их основе лежат идеи, впервые высказанные и проверенные в 1958—1961 гг. американскими специалистами.

Джек Килби собрал компоненты электронной схемы (транзисторы, конденсаторы и резисторы) в едином куске полупроводника. Так появилась первая интегральная схема, произведённая американской корпорацией Texas Instruments. Контакты прикреплялись к конструкции схемы воском. Это было не очень удобно инадёжно, особенно в производстве, и технология не получила дальнейшего развития.

Джин Херни и Курт Леховец усовершенствовали операцию изготовления интегральных схем, добившись особых успехов в изоляции элементов друг от друга (изолятором служила тончайшая плёнка оксида кремния).

Роберт Нойс изобрёл способ, позволяющий делать выводы всех элементов интегральной схемы на одной, верхней, поверхности с помощью напыления металла. Это дало возможность обрабатывать сразу целую кремниевую пластину, содержащую десятки и сотни кристаллов интегральных схем, нанося шаг за шагом «рисунок» — слои будущей схемы. Оставалось только «разрезать» её на кристаллы, упаковать в пластик и припаять «ножки» — и интегральная схема готова. Поэтому и технологию назвали *планарной,* что означает «плоская».

В 1954 г. Гордон Тил предложил изготавливать транзисторы из недорогого материала — обычного кремния, что позволило удешевить производство интегральных схем.

Ещё одна ключевая идея состояла в том, что все компоненты интегральной схемы, будь то диоды, резисторы или конденсаторы, проектировались в ней в виде частей самого сложного элемента — транзистора. И когда говорят околичестве элементов на интегральной схеме, имеют в виду транзисторы.

В небольшой объём интегральной схемы удалось поместить сразу сотни и тысячи транзисторов, в результате уменьшились размеры ЭВМ и упростился монтаж. Благодаря этому полупроводниковая память на интегральных схемах вытеснила громоздкую память на ферритовых сердечниках.

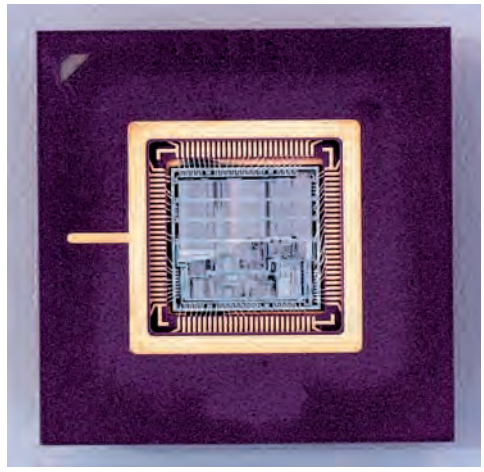
Но не все достижения были связаны только с интегральными схемами. Одновременно возник ещё один класс памяти ЭВМ — ПЗУ (постоянное запоминающее устройство; *англ.* ROM — Read Only Memory). Из ПЗУ допускалось лишь чтение данных. Удалось не только удешевить компьютерную память, но и решить проблему защиты некоторых важных программ ЭВМ, помещая их при производстве в ПЗУ. Обычная память, доступная и для записи, и для чтения, получила название ОЗУ (оперативное запоминающее устройство; *англ.* RAM — Random Access Memory).

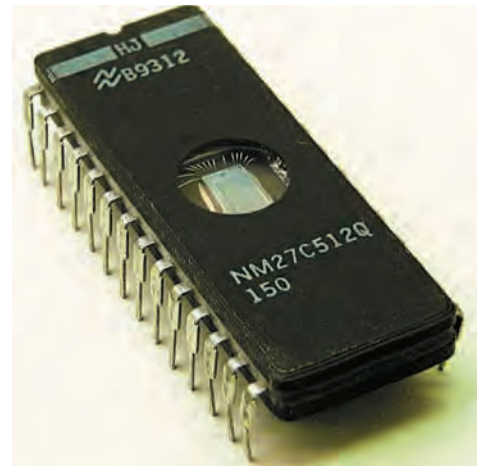
Джек Килби.

Роберт Нойс.

МИКРОПРОГРАММИРОВАНИЕ

Техника *микропрограммирования* (конструирование сложных команд процессора из простых— микрокоманд) стала использоваться при проектировании процессоров. Сами команды процессора реализовывались как микропрограммы (программы, состоящие из микрокоманд).





Интегральные схемы: микропроцессор (слева), ПЗУ (справа).

Программирование, таким образом, «спустилось» внутрь процессора. Теперь не требовалось переписывать целиком всю программу при изменении лишь некоторых свойств алгоритма. Это позволяло обеспечить большую «гибкость» процессора, соответствующую «гибкости» программы. Однако микрокоманды, как правило, оставались недоступны внешним (по отношению к процессору) программам, они были «видны» только в ходе проектирования.

Морис Вилкс, руководитель компьютерной лаборатории Кембриджского университета, ещё в 1951 г. сформулировал основные концепции микропрограммирования:

Морис Вилкс.

• Каждая инструкция процессора (микропрограмма) состоит из последовательности микроинструкций и содержится в специальной памяти, называемой управляющей памятью.

• Микроинструкция осуществляет некоторое *микродействие* (например, передачу данных между регистрами), необходимое для выполнения инструкции процессора.

• Каждое выполнение инструкции процессора, содержащейся в основной памяти, приводит к последовательному выполнению микроинструкций из управляющей памяти.

Микропрограммирование обеспечивает более простой способ проектирования процессора. Набор инструкций «микрозапрограммированного» процессора можно изменить всего лишь заменой *микрокода,* содержащегося в управляющей памяти. Это делает возможным перенастройку такого процессора на систему команд другого и позволяет выполнять программы с «чужой» машины. Подобно тому как сегодня, устанавливая новую версию ОС, расширяют возможности ЭВМ.

Хотя были попытки использовать микропрограммирование при проектировании некоторых машин первого и второго поколений, только в середине 60-х гг. XX в. с появлением моделей IBM System/360 оно получило широкое применение.

IBM SYSTEM/360

Третье поколение компьютеров впечатляет многообразием моделей и количеством фирм-производителей. В числе наиболее известных представителей третьего поколения — ряд компьютеров производства IBM — модели 360 (*англ.* System/360). Это семейство пыталось охватить широкий спектр приложений вычислительной техники. Различные модели были в значительной степени совместимы, и можно уже говорить о мобильности программного обеспечения. То есть программа, написанная для одной модели ряда 360, должна была работать (практически без изменений) на любой другой модели этого ряда. Отличалось, конечно, время выполнения, и могли возникнуть некоторые сложности из-за нехватки места в памяти.

Весь ряд IBM 360 оказал достаточно сильное влияние на стандарты производимых компьютеров.

Простейшие инструкции процессоров IBM 360 состояли из одного полуслова, т. е. из 16 бит:

Команда	Адрес 1	Адрес 2
8 бит	4 бит	4 бит

Здесь адрес 1 и адрес 2 — адреса-имена регистров процессора. Инструкция IBM 360, в отличие от IBM 7094, двухадресная.

Адреса в памяти задавались сложнее. Поскольку можно адресовать любой из 16,7 млн байт, адрес 24-битный (2^{24}=16 777 216), хотя обычно память IBM 360 из-за дороговизны не превышала 32 кбайт. Таким образом, для адресации требовалось только 16 бит (2^{16} = 65 536). В целях экономии памяти при этом использовали так называемый базовый регистр.

Команда	Адрес 1	Индекс	Базовый	Смещение
8 бит	4 бит	4 бит	4 бит	12 бит

В простейшем случае исполнительный адрес, т. е. участвующий в выполнении инструкции, является суммой базового регистра и 12-битного адреса (смещения), указанного в команде. При этом инструкция имеет длину уже 32 бита.

IBM 360.

У IBM 360 было 16 32-битных регистров (общего назначения), используемых в качестве сумматоров индексных и базовых регистров, и 4 регистра с плавающей запятой, каждый размером в двойное слово.

Несмотря на то что IBM 360 относились к компьютерам третьего поколения, в них, как правило, использовалась память на «старых и добрых» ферритовых сердечниках. Сердечники были небольших размеров: внутренний диаметр — 0,48 мм и внешний — 0,76 мм. Компания IBM выпускала матрицы размером 8192 и 284 912 сердечников.

Хотя с 1960 г. IBM поставила пользователям тысячи транзисторных компьютеров, окончательную победу фирма смогла праздновать лишь с появлением ряда 360. Теперь можно было адресовать 32-битное слово памяти, начиная с любого байта.

В новых машинах IBM впервые использовала микропрограммирование при конструировании своих процессоров. Благодаря этому удалось создать шесть моделей ряда 360, каждая из которых могла использовать ту же *периферию* (внешние устройства) и то же программное обеспечение.

К компьютерам ряда 360 осуществлялся удалённый доступ (например, по телефонным линиям). Теперь отпала необходимость ездить в компьютерный зал для запуска вычислительных задач и получения результатов.

Для производства ряда 360 потребовалось вложить 5 млрд долларов — в постройку пяти заводов и привлечение 60 тыс. рабочих. Цель себя оправдала. IBM 360 в 70-х гг. XX в. можно было найти везде: в больницах и университетах, банках и корпорациях, библиотеках и лабораториях.

ОПЕРАЦИОННЫЕ СИСТЕМЫ

Выполнение одновременно нескольких параллельных процессов в компьютере требует наличия достаточно сложной программы, которая управляет ими и распределяет ресурсы компьютера между операциями, устанавливает их очерёдность, взаимодействует с различными программами-процессами и т. п. Специалисты фирмы Burroughs (США) назвали её управляющей мастер-программой (обычно она именуется *операционной системой*). Операционная система большого компьютера — это комплекс программ, хотя некоторые из её функций могут быть выполнены аппаратно. Широкое распространение этих систем — важная характеристика ЭВМ третьего поколения. Развитие операционных систем началось с появления пакетных операционных систем (программы-мониторы) в 50-х гг.

Компьютер Atlas.

Компьютер Atlas (Манчестерский университет, 1961 г.) имел одну из первых таких систем. К проектированию систем разделения времени, когда несколько пользователей одновременно работают на компьютере, практически не замечая друг друга, приступили в начале 60-х гг. *Мультипрограммирование* и *мультиобработка* обычно заключаются в параллельном выполнении программ, делящих между собой основную память компьютера. (Ёмкость основной памяти ограничена по соображениям стоимости.) Её объёма часто не хватает для одновременного хранения всех выполняемых программ с их данными. Поэтому необходимо динамично, «на ходу», выделять память программам. Те программы, для которых она не выделена, должны временно находиться во внешней памяти (например, на диске).

Автоматически (под управлением операционной системы) осуществлялись подгрузка программ из внешней памяти и выгрузка отработавших своё время программ из оперативной памяти на диск. Это явилось одной из главных функций операционной системы ЭВМ третьего поколения.

МИНИ-КОМПЬЮТЕРЫ

Контрастом на фоне больших ЭВМ было начало производства маленьких дешёвых компьютеров, называемых мини-компьютерами. Они характеризовались «коротким» машинным словом длиной от 8 до 32 бит, ограниченными аппаратными и программными средствами, а также небольшими размерами. Невысокая цена обусловила их широкое применение. Например, они использовались как индустриальные контроллеры — машины, управляющие станками и даже целыми цехами относительно простого производственного процесса, где бессмысленно было использо-

◄
Компьютер PDP-1.

вать многоцелевые «тяжёлые» компьютеры. По скорости мини-компьютеры были сопоставимы с большими компьютерами второго поколения и превосходили лучшие показатели машин первого поколения. Одним из представителей мини-компьютеров являлась машина PDP-1 (*англ.* Programmed Data Processor) фирмы DEC. PDP-1 могла управлять многими внешними устройствами. Установка машины не требовала ни специального перепроектирования системы энергоснабжения (в США те же 110 вольт), ни другой системы кондиционирования, ни усиления пола: компьютер занимал всего около 2 м^2 и состоял из 4 блоков весом в несколько десятков килограммов.

Центральный процессор выполнял арифметические операции, осуществлял адресацию, управление и доступ кпамяти. Машинное слово было длиной 18 бит.

Команды сложения и вычитания выполнялись за 10 мкс, команда умножения — примерно за 20 мкс. Скорость выполнения достигала 100 тыс. команд сложения в секунду. Инструкции процессора были простыми — одноадресными.

Основная память машины — на магнитных сердечниках, стандартно 4096 18-битных слов. Память можно было увеличить блоками по 4096 слов, её максимальный размер составлял 65 536 слов. Чтение/запись данных из памяти происходили за 5 мкс.

PDP-1 создали именно для управления устройствами ввода-вывода. Все операции ввода-вывода осуществлялись через высокоскоростные каналы. Мини-компьютер использовал и аппарат прерываний, хорошо зарекомендовавший себя в машинах второго и третьего поколений.

PDP-11

Наибольшую популярность получила серия ЭВМ PDP-11, выпущенная в 70-х гг. Машинное слово этих компьютеров было длиной 16 бит.

Процессор имел 8 одинаковых 16-битных регистров общего назначения. Правда, на два последних регистра возлагали дополнительные функции: регистр R7 являлся счётчиком команд, а R6 — указателем стека. Стек — это место в памяти компьютера, которое используется для временного хранения содержимого регистров и для хранения адреса возврата из подпрограммы (нужно запомнить, куда вернуться). Вызов подпрограммы обычно происходит по инструкции JSR R7 (*англ.* Jump to SubRoutine — «переход к подпрограмме»), а возврат— по RTS R7 (*англ.* ReTurn from Subroutine — «возврат из подпрограммы»). Адреса в программе восьмеричные, по стандарту PDP-11.

Стрелка указывает на текущую исполняемую инструкцию (до выполнения входа в подпрограмму).

Дисковод сменного магнитного диска PDP-11 объёмом 2,5 Мбайт.

```
100000 Программа          ПАМЯТЬ   адрес
. . . . . . . . . . . . . . . . . . . . . .
101000 JSR R7, #110000    000000   776
101004 . . . . . . . . . . . . . . .  000000   1000
. . . . . . . . . . . . . . . . . . . . . .  170000   1002
110000 Подпрограмма       . . .
. . . . . . . . . . . . . . . . . . . . . .
110100 RTS R7

R6= 001002      R7= 101000
```

Следующий шаг — запись в стек адреса возврата (содержится в R7) и запись в R7 адреса подпрограммы.

```
100000 Программа          ПАМЯТЬ   адрес
. . . . . . . . . . . . . . . . . . . . . .
101000 JSR R7, #110000    000000   776
101004 . . . . . . . . . . . . . . .  101004   1000
. . . . . . . . . . . . . . . . . . . . . .  170000   1002
110000 Подпрограмма       . . .
. . . . . . . . . . . . . . . . . . . . . .
110100 RTS R7

R6= 001000      R7= 110000
```

После выполнения подпрограммы по команде RTS происходят «вспоминание» адреса возврата из стека и передача его в R7.

```
100000 Программа          ПАМЯТЬ   адрес
. . . . . . . . . . . . . . . . . . . . . .
101000 JSR R7, #110000    000000   776
101004 . . . . . . . . . . . . . . .  101004   1000
. . . . . . . . . . . . . . . . . . . . . .  170000   1002
110000 Подпрограмма       . . .
. . . . . . . . . . . . . . . . . . . . . .
110100 RTS R7

R6= 001002      R7= 101004
```

Инструкции процессора были двухадресные. В некоторых моделях добавлялось устройство для выполнения арифметических вычислений с плавающей запятой, существенно увеличивающее скорость вычислений.

Основная память машины составляла 32 768 16-битных слов. Её также можно было расширять. Доступ к памяти, превышающей 64 кбайт, осуществлялся при помощи устройства ММ (*англ.* Memory Management — «управление памятью»). ММ «делило» программу на восемь частей, размещая их в различных областях оперативной памяти. Программа этого «деления» не замечала, считая, что выполнялась воедино, в 64 кбайт. Дополнительно код программы можно было отделять от кода данных. Таким образом, максимальный размер программы может составлять 64 кбайт и максимальный размер области данных тоже 64 кбайт. Так как основная память была достаточно медленной (память на магнитных сердечниках), быстродействие обеспечивала полупроводниковая *кэш*-память (ёмкостью 2048 слов).

▶ Компьютеры PDP-1140 и PDP-1170.

ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ

Способность проводить вычисления параллельно — это ещё одна характерная черта компьютеров третьего поколения.

Помимо IBM с её нашумевшим рядом 360 многие фирмы приступили к производству вычислительной техники, добиваясь увеличения произво-

дительности машин. Одним из способов повышения скорости счёта была технология *многомашинных комплексов,* когда для решения общей вычислительной задачи объединяли несколько одинаковых машин. Так, две машины IBM 7094 соединяли друг с другом для ускорения вычислений, однако это требовало разработки специальных методов программирования. Кроме того, дублирование внешних устройств и расположение общих данных задачи в двух разных компьютерах не давали нужной эффективности.

Для решения вычислительных задач также широко применялась новая, *многопроцессорная* техника параллельной обработки данных. В 1963 г. фирма Burroughs (США) начала выпуск высокопроизводительных машин, построенных по этому принципу. Особенностями этих машин были не только необычные концепции разработки компьютеров («компьютер, как единое целое, должен проектировать один человек»), но и применение нескольких однотипных процессоров для повышения производительности системы.

Первой стала машина B 5000, затем появилось ещё несколько моделей, а уже более совершенную B 7700 выпустили в 1973 г. В некоторых экземплярах B 7700 было установлено семь быстродействующих процессоров, осуществляющих между собой обмен данными через память. Оперативная память разделялась на восемь независимых модулей. Модульная конструкция обеспечивала одновременное обращение в память с восьми независимых входов. Как и в IBM, в Burroughs использовалась техника канальных процессоров, отвечающих

Компьютер B 7700.

за операции ввода-вывода. Они могли получать данные от 32 внешних устройств.

При программировании многопроцессорных комплексов применялось специализированное программное обеспечение, при котором вычисления *распараллеливались* (т. е. выполнялись одновременно на разных процессорах). Это можно пояснить на примере нахождения суммы четырёх целых чисел.

Алгоритм **Сумма обычная** вычисляет сумму четырёх чисел. Для получения *sum* требуется три раза произвести сложение. При использовании вспомогательных переменных *s1* и *s2* (в алгоритме **Сумма параллельная**) вычисления можно производить параллельно (что показано в примере расположением *s1* и *s2* на одной строке, через знак :). Для этого потребуется всего два раза осуществить операцию сложения, т. е. второй алгоритм выполняется быстрее первого. Для эффективного применения многопроцессорных комплексов были созданы специальные языки программирования. Вычислительный комплекс как бы сам «распараллеливал» вычисления. Поэтому даже алгоритм **Сумма**

```
алг Сумма обычная (цел a, b, c, d)
дано a, b, c, d
надо
нач цел sum
| sum:= a+b
| sum:= sum+c
| sum:= sum+d
| знач:= sum
кон
```

```
алг Сумма параллельная (цел a, b, c, d)
дано a, b, c, d
надо
нач цел sum, s1, s2
| s1:= a+b : s2:= c+d
| sum:= s1+s2
| знач:= sum
кон
```

```
алг Вычисление обычное (цел a, b, c, d, f)
дано a, b, c, d, f
надо
нач цел sum1, sum2, s
| s:=a+b
| sum1:=s*c
| s:=d+e
| sum2:=s*f
| знач:=sum2
кон
```

```
алг Вычисление конвейерное (цел a, b, c, d, f)
дано a, b, c, d, f
надо
нач цел sum1, sum2, s
| s:=a+b
| sum1:=s*c : s:=d+e
| sum2:=s*f
| знач:=sum2
кон
```

Конвейерные системы были названы по аналогии со сборочным автомобильным конвейером. *Специальные* процессоры, осуществлявшие однотипные операции (процессор для сложения и вычитания, процессор для деления и умножения и т. д.), ассоциировались с рабочими, выполняющими специализированные операции сборки. Данные для вычислений как бы продвигались по конвейеру. Сразу со всем потоком различных данных производились соответствующие операции.

обычная выполнялся на многопроцессорном комплексе быстрее, чем на обычной ЭВМ.

Совершенствование многопроцессорных вычислительных комплексов привело к появлению *конвейерных систем.* В них получила развитие идея параллельности вычислений. За счёт специальных методов программирования удавалось добиться значительного ускорения.

Алгоритм **Вычисление обычное** производит два сложения и два умножения, а второй алгоритм (на конвейерном компьютере) выполняет не четыре действия (2+2=4), а три, так как произведение *sum1* и сумма *s* вычисляются одновременно на различных специализированных процессорах комплекса (что показано в примере расположением *sum1* и *s* на одной строке, через знак :). В реальных вычислениях, осуществляемых на конвейерной системе, оба алгоритма будут работать одинаково, так как (аналогично многопроцессорным вычислительным комплексам) система сама «сообразит», где можно сэкономить на вычислениях.

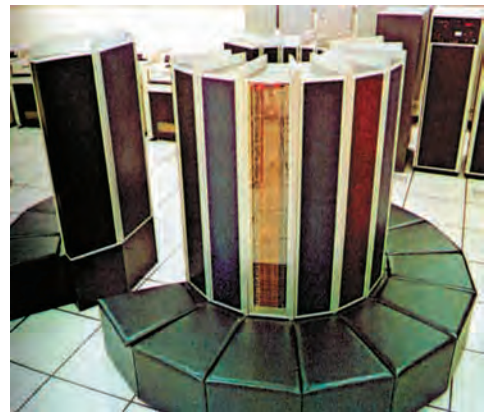
Работа на конвейере компании «Форд мотор». Начало XX в.

Кроме того, параллелизм можно обеспечить на процессорном уровне. Два метода параллельной обработки данных нашли своё место в разработке процессоров:

- Можно предусмотреть наличие нескольких устройств (например, арифметических) для выполнения некоторых действий (например, сложения); *n* независимых сумматоров могут вычислять одновременно *n* сумм. Этот метод позволяет быстро обрабатывать массивы (осуществлять *векторные* операции).
- Сам процессор можно проектировать по принципу конвейера, позволяющего выполнять различные *микродействия* одновременно.

СУПЕРКОМПЬЮТЕРЫ

Корпорация CDC (*англ.* Control Data Corporation), США, производитель конвейерных систем, начала проектирование компьютеров в 1960 г. В 1964 г. она выпустила модель под названием CDC 6600, спустя 5 лет появи-



Компьютер CDC 7600 (слева).
Компьютер CRAY I (справа).

лась более мощная модель CDC 7600. Позднее ряд этих машин продолжили модели CYBER. Для ЭВМ фирмы CDC было характерно наличие нескольких процессоров ввода-вывода (называемых периферийными процессорами), каждый из которых обладал автономией от центрального процессора. Если канал обычно имел память всего в несколько слов и небольшой специфический набор инструкций, то периферийный процессор с обширной памятью и развитой системой команд представлял собой универсальную ЭВМ. Работа канала инициировалась и контролировалась центральным процессором, а на периферийный процессор не мог оказывать влияния. Десять периферийных процессоров CDC 6600 обладали памятью по 4096 12-разрядных слов.

Центральный процессор компьютеров CDC, в свою очередь, состоял из множества элементов — процессоров, которые могли работать параллельно, независимо друг от друга. В процессор CDC 6600 входило десять функциональных устройств: одно для сложения чисел с фиксированной запятой, другое для сложения с плавающей запятой, два устройства умножения, устройство деления, устройство для выполнения логических операций и т. п. Средняя производительность CDC 6600 около 4 млн операций в секунду, а CDC 7600 около 15 млн. (Для суперЭВМ производительность измеряется количеством операций в секунду над числами с плавающей запятой.)

Одним из мировых лидеров в производстве суперкомпьютеров являлась американская фирма Cray Research. Её основал в 1972 г. бывший главный конструктор CDC 6600 и 7600 Сеймур Крей. В 1976 г. фирма установила в Лос-Аламосской научной лаборатории при Калифорнийском университете первый образец своей машины CRAY I. Средняя производительность её составляла 80 млн операций в секунду, при полном задействовании всех вычислительных ресурсов — 138 млн операций с плавающей запятой в секунду, а в короткие промежутки времени — 250 млн.

Такая высокая производительность достигнута благодаря использованию нескольких конвейеров. Система могла выполнять операции с *векторами* (процессор имел векторные регистры), причём функциональные устройства процессора обеспечивали одновременное выполнение векторных и обычных *(скалярных)* операций. Процессор содержал 8 64-битных регистров для выполнения арифметических операций, 64 64-битных регистра для хранения промежуточных результатов и 8 64-элементных векторных регистров (каждый элемент вектора — это 64-разрядное слово).

Роберт Нойс.

Быстродействие CRAY I — заслуга его небольших размеров (высота не более 2 м, диаметр около 2,5 м). Такие габариты позволили связать компоненты суперЭВМ «короткими» линиями (проводами), что сократило задержки при передаче информации между компьютерными блоками. Однако компактность ЭВМ потребовала наличия системы жидкостного охлаждения, похожей на систему охлаждения двигателя автомобиля.

Компьютер ILLIAC IV.

Память насчитывала до 1 млн 64-битных слов. На самом деле слово памяти состояло из 72 бит, где восемь «служебных» битов использовались для обнаружения и исправления сбоев памяти. Другой суперкомпьютер — ILLIAC IV, производящий операции не только с векторами, но даже с матрицами, был разработан в 1967 — 1974 гг. в Иллинойсском университете (США) и изготовлен фирмой Burroughs всего в одном экземпляре. Цель этого проекта — построение системы с производительностью около 1 млрд операций в секунду. Для этого машина должна была содержать 256 процессорных элементов (фактически универсальных процессоров) под общим управлением, сгруппированных в 4 матрицы по 64 процессорных элемента в каждой. Поэтому ILLIAC IV относят к классу *матричных* ЭВМ.

Матричные системы представляют собой двухмерные таблицы-матрицы процессорных модулей. При этом процессорный модуль имел память и мог выполнять множество арифметических операций: сложение и умножение, вычитание и деление, сдвиг, сравнение и др. Если в конвейерных системах все процессоры (функциональные устройства в CDC 6600 и 7600) выполняли различные операции, то матричная система одновременно выполняла одинаковые операции со всей таблицей-матрицей слов. Трудоёмкие операции (например, «умножить все элементы матрицы на какое-то число», «обнулить матрицу», а иногда и более сложные типа «сложить первую и вторую строки матрицы») производились за одно действие, как будто надо умножить число на число.

Из-за своей универсальности матричные системы очень сложны. Поэтому была построена всего одна матрица из 64 процессорных элементов, в результате производительность упала до 200 млн операций в секунду. Тем не менее до своего списания в 80-х гг. ILLIAC IV оставался самым высокопроизводительным компьютером в мире. Стоимость проекта составила почти 100 млн долларов.

АЛЕКСЕЙ АНДРЕЕВИЧ ЛЯПУНОВ

Математик Алексей Андреевич Ляпунов родился в Москве 8 октября 1911 г. По семейным преданиям, род Ляпуновых уходит своими корнями к брату Александра Невского, князю Константину Галицкому. Прадед Алексея Андреевича Михаил Васильевич, ученик Н. И. Лобачевского, был профессором астрономии Казанского университета и директором обсерватории, а позже директором Демидовского лицея (первого высшего учебного заведения Ярославля). Он имел трёх сыновей, которые были широко известны: математик, механик и кораблестроитель Александр Михайлович, композитор Сергей Михайлович, филолог-славист академик Борис Михайлович.

Круг научных интересов самого Алексея Ляпунова в значительной мере определён средой, в которой он рос. Его первым учителем астрономии, физики, математики и минера-

А. А. Ляпунов основал издание серии научных сборников «Проблемы кибернетики», создал и редактировал серию книг «Кибернетика в монографиях», организовал публикацию переводов лучших работ зарубежных авторов в серии «Кибернетический сборник».

логии был отец Андрей Николаевич, который окончил физико-математический факультет Московского университета, а позже учился в Гейдельберге. Он же привил сыну и любовь к математике.

Первые самостоятельные исследования — наблюдения по астрономии, проведённые Алексеем в школьные годы, неоднократно публиковались в «Бюллетене Московского общества любителей астрономии». После окончания школы в 1928 г. Ляпунов поступает в Московский государственный университет на физико-математический факультет, но через полтора года ему пришлось уйти из университета «как лицу дворянского происхождения».

Осенью 1930 г. он начинает работать лаборантом в Государственном геофизическом институте у известного физика, академика П. П. Лазарева (1878—1942). Высшее образование Алексей Ляпунов получает экстерном, сдав экзамены по университетским курсам в МГУ. Поддерживаемый известным математиком, академиком Н. Н. Лузиным (1883—1950), Ляпунов углублённо изучает математику и включается в исследования в области теории множеств, которым посвящены более 60 его научных работ. Алексей Ляпунов публикует работы по теории множеств и незадолго до Великой Отечественной войны защищает кандидатскую диссертацию.

Война прервала на время научную деятельность Ляпунова. В 1942 г. Ляпунов добровольцем уходит на фронт, командиром топографического взвода в артиллерии.

После войны Алексей Андреевич работает преподавателем математики в Артиллерийской академии имени Ф. Э. Дзержинского.

Ляпунов не оставляет науку, защищает докторскую диссертацию в Математическом институте имени В. А. Стеклова, работает в Институте геофизики АН СССР, получает звание профессора. Академик М. В. Келдыш приглашает его в отдел кибернетики Отделения прикладной математики МИАН.

Ляпунов — один из первых отечественных учёных, кто оценил значение *кибернетики,* внёс большой вклад

А. А. Ляпунов.

в её становление и развитие. Работать в области кибернетики значило бороться за её существование.

Профессор кафедры математической логики и вычислительной математики на механико-математическом факультете МГУ, он в 1953 г. организует семинар по программированию, а с 1954 г. — семинар по кибернетике. Семинар стал центром, где объединились учёные и специалисты, а Ляпунов координировал работы в области кибернетики. По его инициативе в 1959 г. при Академии наук СССР создаётся Научный совет по комплексной проблеме «Кибернетика», который возглавляет академик А. И. Берг (1898—1979).

В 1961 г. Алексей Андреевич Ляпунов принимает приглашение создателя Сибирского отделения АН СССР и Академгородка академика М. А. Лаврентьева и переезжает в Новосибирск. Он создаёт отдел кибернетики в Институте математики и кафедру теоретической кибернетики в Новосибирском университете. За годы, прожитые там, Алексей Андреевич сумел осуществить многие из своих замыслов. Он был инициатором создания в 1962 г. первой в стране физико-математической

Медаль «Computer Pioneer». На лицевой стороне – Чарлз Бэббидж, создатель первой (механической) вычислительной машины. На обратной стороне надпись: «Компьютерное общество признало Алексея Андреевича Ляпунова основателем советской кибернетики и программирования».

«За короткий срок отношение к кибернетике прошло следующие фазы: категорическое отрицание — констатация существования — признание полезности, отсутствие задач для математиков — признание некоторой математической проблематики — полное признание математической проблематики кибернетики».

А. А. Ляпунов.
Выступление на IV Всесоюзном математическом съезде. 1966 г.

Ляпунов читает лекцию.

школы-интерната при Новосибирском университете, первым председателем его учёного совета и активным лектором; поддерживал организацию Всесибирских математических олимпиад и летних физматшкол в Академгородке.

В 1964 г. Ляпунов становится членом корреспондентом АН СССР по отделению математики.

Область научных интересов Ляпунова огромна. Его работы посвящены общим вопросам кибернетики, математическим основам программирования и теории алгоритмов, математической лингвистике и машинному переводу, кибернетическим аспектам биологии, а также философским и методологическим проблемам развития научной мысли. Им создан *операторный метод,* по существу первый язык программирования, отличающийся от машинного и разработанный до появления алгоритмических языков типа Algol.

За свою научную, педагогическую и организаторскую деятельность он был награждён правительственными наградами: орденом Ленина, двумя орденами Трудового Красного Знамени и орденом «Знак Почёта». Но больше всего учёный гордился орденом Красной Звезды, который он получил на Великой Отечественной войне за участие в боях по освобождению Крыма.

Алексей Андреевич был не только выдающимся математиком, но и прекрасно знал литературу, интересовался архитектурой и живописью, любил демонстрировать свою минералогическую коллекцию, в совершенстве владел французским языком, глубоко знал историю, искусство и культуру Франции, хотя никогда не выезжал за пределы своей страны.

Алексей Андреевич Ляпунов умер 23 июня 1973 г. и похоронен на Введенском кладбище (аллея, ведущая к его могиле, проходит мимо места, где покоится прах его учителя Н. Н. Лузина).

Первым шагом в международном признании заслуг Алексея Андреевича в области информатики явилось присуждение ему в 1996 г. медали «Computer Pioneer». Это было сделано одной из самых авторитетных профессиональных организаций в сфере высоких технологий — IEEE Computer Society. На лицевой стороне медали изображён Чарлз Бэббидж. Оборотную сторону украшает надпись на английском языке: «Компьютерное общество признало Алексея Андреевича Ляпунова основателем советской кибернетики и программирования».

ТРОИЧНАЯ СИСТЕМА СЧИСЛЕНИЯ И ТРОИЧНАЯ ЭВМ

В 1840 г. французским математиком и изобретателем механических вычислителей Леоном Лаланном была предложена уравновешенная троичная система. Основанием этой системы служит число 3, но вместо цифр 0, 1 и 2 используются 0, 1 и –1 (обычно вместо «минус единицы» используется символ $\bar{1}$). По аналогии с битами двоичной системы счисления

такие обозначения называют *тритами.*

Сразу же можно заметить, что знак числа определяется первым ненулевым символом в его троичной записи. Если это 1, то число положительное, а если $\bar{1}$, то отрицательное. Очень просто, поразрядно, выполняются в уравновешенной троичной системе сложение и вычитание ($1+0=1$, $1+\bar{1}=0$, $1+1=1\bar{1}$, $\bar{1}+\bar{1}=\bar{1}1$). Сложим в уравновешенной троичной системе числа 247 и 122:

$$\begin{array}{r} 100011 \\ + \\ 1\bar{1}\bar{1}1\bar{1}\bar{1} \\ \hline 1\bar{1}\bar{1}\bar{1}\bar{1}00 \end{array}$$

Легко проверить, что $1\bar{1}\bar{1}\bar{1}\bar{1}00_3=369$. Столь же просто производится и вычитание: для этого достаточно изменить знак вычитаемого на противоположный и сложить число с уменьшаемым. Умножение также сводится к простым операциям изменения знака и сложения.

Более 100 лет после Лаланна уравновешенная троичная система не привлекала большого внимания учёных. Но когда на смену ранним механическим и электромеханическим вычислительным устройствам, основанным на десятичной системе счисления, стали приходить электронные компьютеры, она всерьёз рассматривалась как возможная альтернатива двоичной.

Арифметические операции над числами, записанными в этой системе, выполняются очень просто. Обладает она и другими достоинствами: запись числа в ней короче, чем при использовании двоичной системы; нет необходимости вводить специальный знаковый разряд.

Знаменитый американский специалист в области методов вычислений Доналд Кнут писал в 1969 г., что время троичной арифметики в компьютере наступит, если удастся найти замену триггеру (логическая схема, имеющая два состояния и позволяющая хранить один бит информации). Однако задолго до выхода книги Кнута троичная система счисления уже была успешно применена в весьма интересной и оригинальной отечественной разработке. Речь идёт об ЭВМ «Сетунь», созданной в Вычислительном центре МГУ в 1959 г. под руководством академика С. Л. Соболева и Н. П. Брусенцова. Начиная с 1962 г. машина выпускалась серийно, и за три года было изготовлено около 50 экземпляров. Устройства на базе троичной арифметики оказались не только более простыми и быстрыми, но и

ЗАДАЧА БАШЕ О ВЕСАХ

Интересно, что представление чисел в уравновешенной троичной системе неявно присутствует ещё в знаменитой «задаче Баше о весах»— которую, впрочем, Леонардо Пизанский (Фибоначчи) сформулировал за 400 лет до Баше, в XIII в. В задаче требуется найти набор из четырёх гирь, с помощью которого можно взвесить любой груз весом от 1 до 40 кг (или набор из трёх гирь для взвешивания любого груза весом от 1 до 13 кг— задача известна в разных вариантах).

Для решения задачи надо просто найти представление в уравновешенной троичной системе целого числа, равного весу груза. Пусть, например, надо взвесить груз в 33 кг. Представим число 33 в уравновешенной троичной системе: $33=27+9-3+0=11\bar{1}0_3$. Значит, гири 27 кг и 9 кг надо положить на одну чашу весов, а взвешиваемый груз и гирю 3 кг— на другую. Если к этим трём гирям добавить ещё одну весом 1 кг, то мы сможем взвесить любой груз, не превышающий 40 кг.

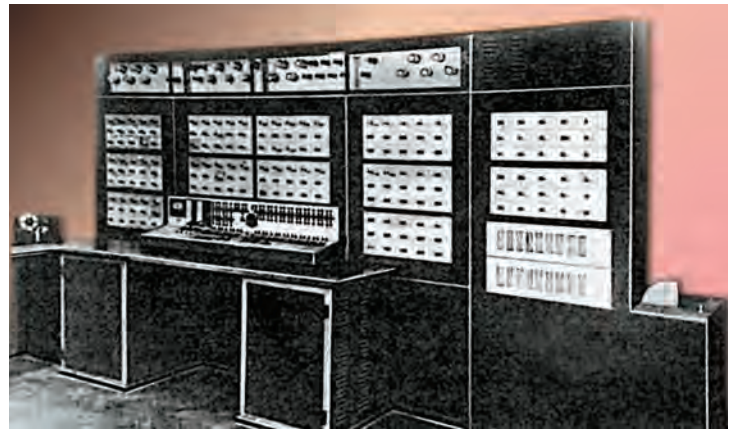
$8_{10}=3^2+0-3^0=10\bar{1}_3$,

$1\bar{1}\bar{1}\bar{1}\bar{1}\bar{1}_3=3^5-3^4-3^3-3^2-3^1-3^0=243-81-27-9-3-1=122_{10}$,

$-17_{10}=3^3+3^2+0+3^0=\bar{1}101_3$,

$\bar{1}000\bar{1}\bar{1}_3=-3^5+0+0+0-3^1-3^0=-243-3-1=-247_{10}$

ЭВМ «Сетунь».



С. Л. Соболев.

весьма надёжными. Они также потребляли меньшую мощность, чем двоичные устройства, реализованные на тех же элементах.

Эта машина выполняла операции с фиксированной запятой, а операции с плавающей запятой реализовывались программно. Система команд насчитывала всего 24 команды (среди них — три команды условного перехода и команда умножения тритов), но, используя их, можно было эффективно программировать самые разные задачи. Память была двухуровневой: оперативная память на ферритовых сердечниках объёмом 162 коротких слова (9 тритов) и основная память на магнитных барабанах объёмом 3888 коротких слов. Время выполнения сложения и вычитания составляло 180 мкс, а умножения — 320 мкс. Конечно, сегодня эти цифры кажутся весьма скромными, но в то время «Сетунь» превосходила другие малые машины по производительности. При этом она была значительно дешевле машин своего класса.

Спустя несколько лет, в 1970 г., был построен экспериментальный образец усовершенствованной машины «Сетунь 70». Эта ЭВМ отличалась рядом интересных архитектурных особенностей, а идеи применения трёхзначной логики получили в ней дальнейшее развитие. И хотя проект «Сетунь» до сих пор остаётся единственным примером использования троичной системы в компьютерах, возможно, время троичных машин ещё придёт.

ОТЕЧЕСТВЕННЫЕ ЭВМ

С. А. Лебедев.

На первый взгляд может показаться, что вычислительная техника разрабатывалась только в США. Но это не так. Действительно, новая научная область требовала больших финансовых вложений, что было не под силу послевоенной Европе, ставшей основным плацдармом Второй мировой войны. Одной из немногих стран, которая, несмотря ни на что, стала участником гонки в компьютеростроении, являлся СССР.

В 1948 г. академик Сергей Алексеевич Лебедев (1902—1974), пионер отечественного производства компьютеров, начал строительство первой советской (и европейской) ЭВМ — малой электронной счётной машины (МЭСМ). Работы по её созданию носили исследовательский, экспериментальный характер. В 1950 г. в Институте электромеханики Академии наук Украины МЭСМ ввели в эксплуатацию. В 1952—1953 гг. она оставалась практически единственной регулярно эксплуатируемой ЭВМ в Европе.

Основные параметры машины: быстродействие — 50 операций в секунду; в памяти можно было хранить 31 16-разрядное число и 63 команды длиной 20 бит; площадь помещения, занимаемого машиной, — 60 м^2; потребляемая мощность — 25 кВт. Только в ОЗУ использовалось 2,5 тыс. триодов и 1,5 тыс. диодов. Для расширения маленькой памяти можно было дополнительно использовать магнитный барабан ёмкостью 5 тыс. слов (по 16 бит). Машина имела сменное так называемое долговременное запоминающее устройство (позже названное ПЗУ) для хранения числовых констант и часто выполняемых команд.

Конечно, машина, по современным меркам работала медленно, но основные принципы её построения (Лебедев предложил их независимо от проводимых в США разработок) использовались при проектировании других ЭВМ. МЭСМ фактически явилась моделью БЭСМ — большой электронной счётной машины. Обе машины (МЭСМ и БЭСМ) были изготовлены в единичном экземпляре.

Практически весь коллектив сотрудников, создавших МЭСМ, стал ядром Вычислительного центра АН УССР, организованного на базе лаборатории С. А. Лебедева.

Работа над БЭСМ в Вычислительном центре закончилась в 1952 г., и через год в АН СССР она уже вошла в строй. БЭСМ по праву признана лучшей европейской ЭВМ 50-х гг. XX в. Машина обрабатывала 39-разрядные слова со средней скоростью 10 тыс. операций в секунду. В качестве внешних запоминающих устройств БЭСМ использовала два магнитных барабана по 5120 слов в каждом. Скорость считывания с барабана составляла 800 слов в секунду. К машине также подключались магнитные ленты общей ёмкостью 120 тыс. слов.

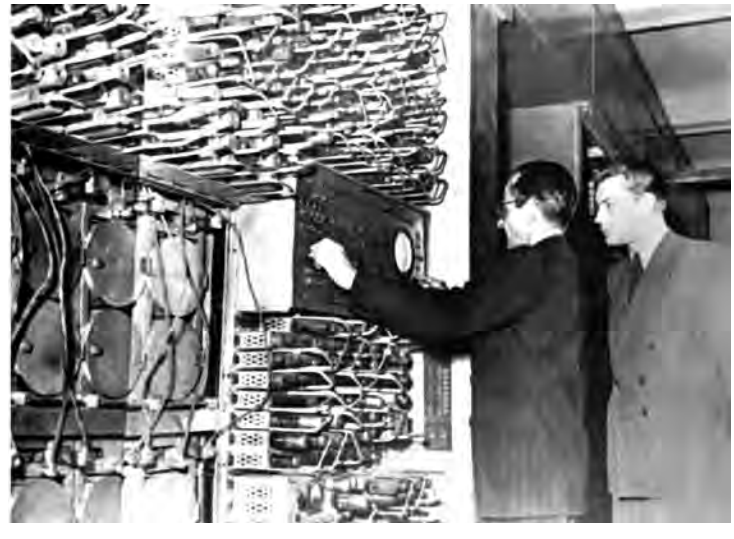
БЭСМ положила начало целой серии цифровых вычислительных машин. Ртутные линии задержки, используемые в качестве элементов оперативной памяти, в 1954 г. были заменены на запоминающие электронно-лучевые трубки. А через два года их сменили ферритовые сердечники объёмом 1024 39-разрядных слова.

БЭСМ-4.

В таком виде машина известна как БЭСМ-1. На ней решались разнообразные задачи, например подсчитывались орбиты движения 700 малых планет Солнечной системы.

Для промышленного изготовления конструкцию машины изменили, и в 1958 г. начался серийный выпуск ламповой машины БЭСМ-2. Её потребляемая мощность составляла 75 кВт.

В 1964 и 1966 гг. появились новые машины этого ряда — БЭСМ-3М и БЭСМ-4. В отличие от своих предшественниц, они собирались из полупроводниковых элементов. Машина БЭСМ-4 имела память 2×4096 45-разрядных слов, четыре магнитных барабана объёмом 16 384 слов и потребляла всего 8 кВт.



С. А. Лебедев и В. А. Мельников у БЭСМ-6.

В 1967 г. для задач, требующих множества сложных вычислений, была создана полупроводниковая машина БЭСМ-6 со средним быстродействием 1 млн операций в секунду. По сравнению с БЭСМ-4 память возросла в 8 раз (разрядов было 48, а не 45), а магнитных барабанов стало 16 по 32 тыс. слов в каждом.

В БЭСМ-6 отразились все передовые тенденции развития вычислительной техники того времени: мультипрограммный режим, аппаратная система прерывания, схема «защиты памяти» и автоматического присвоения адресов (т. е. фактически диспетчер памяти). Любая часть памяти могла использоваться в качестве стека. Центральный процессор использовал одноадресную систему команд и 16 быстродействующих регистров.

Для программирования применялись языки FORTRAN и Algol. Машина оказалась настолько удачна и надёжна, что эксплуатировалась до 90-х гг. Редкий современный компьютер похвастает подобным долголетием!

Под руководством С. А. Лебедева в 1958 г. в Институте точной механики и вычислительной техники АН СССР создали ЭВМ М20. Она стала родоначальницей семейства машин М220 и М222. Среднее быстродействие М20 было 20 тыс. операций в секунду. Память объёмом 4096 45-разрядных слов выполнена на ферритовых сердечниках. Три магнитных барабана запоминали более 12 тыс. слов. Ввод происходил с перфокарт, вывод — на печатающее устройство. Машина была построена по блочному принципу, что упрощало ремонт. В ней использовалось 4,5 тыс. электронных ламп и 3,5 тыс. полупроводниковых диодов.

В 1957 г. в Пензе была создана одноадресная ламповая ЭВМ «Урал-1». Хотя машина отличалась большими размерами, по производительности её отнесли к малым. Можно считать, что с неё началась история малых ЭВМ. При малом быстродействии (100 операций в секунду) машина не нуждалась в быстром запоминающем устройстве, поэтому в качестве основной памяти использовался магнитный барабан объёмом 1024 36-битных слова, который впоследствии заменили на ферритовое запоминающее устройство. В 1964—1971 гг. выпустили ряд программно- и аппаратно-совместимых между собой моделей: «Урал-11», «Урал-14», «Урал-16».

Машины серии «Минск» в 70-х и 80-х гг. XX в. применялись в основном для инженерных и научных расчётов. Одна из них, «Минск-22» (её показатели: 5 тыс. операций в секунду, память — 8 тыс. 37разрядных

МАШИНА ДЛЯ ИНЖЕНЕРНЫХ РАСЧЁТОВ

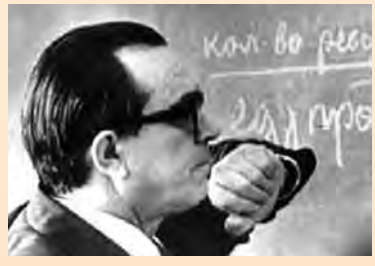
ЭВМ МИР (машина для инженерных расчётов — фактически первый персональный компьютер (мог разместиться в небольшой комнате), задуманный и построенный под руководством академика Виктора Михайловича Глушкова в 1965 г. в Институте кибернетики АН УССР. Первые машины использовали алгоритмический язык МИР, весьма близкий к математическому. В ходе проведения вычислений можно было легко изменять алгоритм, формулы, коэффициенты, подобно тому как это делалось на персональных компьютерах 90-х гг. при использовании языка BASIC. Программист работал за столом с электрической пишущей машинкой, с которой вводилась и выводилась информация.

В машине предусматривалось микропрограммирование (специальные микропрограммные матрицы общей ёмкостью около 700 кбит, описывающие логику команд машины). Основная память располагалась на магнитных сердечниках ёмкостью 4096 байт. Арифметическое устройство было матричным, что позволило сделать произвольную разрядность обрабатываемых чисел, при этом быстродействие составляло около 8 тыс. операций в секунду.

В 1968 г. вышла модификация машины — МИР-1, которая могла вводить и выводить информацию на перфоленту. Экземпляр ЭВМ МИР на выставке в Лондоне приобрела американская фирма IBM.

В 1969 г. была выпущена машина МИР-2, в ней впервые применён дисплей со световым пером, упростивший диалог с ЭВМ и сделавший его более оперативным. В новой машине использовался язык программирования Аналитик, дальнейшее развитие алгоритмического языка МИР. На ней можно было решать задачи линейной алгебры не только в числовом, но и в аналитическом виде. Быстродействие машины возросло до 12 тыс. операций в секунду. Ряд машин МИР имел продолжение.

Машина МИР-1.



В. М. Глушков.

Машина «Урал-1» (слева). «Минск-22» (справа).

слов), долгое время являлась основным компьютером вычислительного центра ГУМа (главного универмага страны). В ней (магнитная лента вмещала 1,6 млн слов) хранилась информация о движении товаров по всем складам магазина и производился расчёт заработной платы. Однако, испытывая некоторое недоверие к вычислительной технике, руководство параллельно держало обширный штат бухгалтеров, проверявших вычисления машины. Ассемблер ЭВМ имел кириллическую мнемонику и назывался ЯСК (язык символического кодирования).

Другой компьютер этого ряда — «Минск-32» обладал быстродействием 25 тыс. операций в секунду и комплектовался памятью до 65 тыс. 37-разрядных слов. Машина была программно-совместима с «Минском-22». Медленные и быстрые каналы позволяли подключать к ней магнитные барабаны, что существенно ускоряло производительность. ЭВМ «Минск-32» уже имела компиляторы с языков программирования высокого уровня — Алгамс (разновидность Algol) и Кобол.

К отечественным суперЭВМ (машины, предназначенные для высокоскоростных вычислений) относят многопроцессорные вычислительные комплексы (МВК) «Эльбрус», разработанные в 70—80-х гг. «Эльбрус-1» достигал производительности 10 млн операций в секунду. В конфигурацию машины входило до десяти центральных процессоров, обращающихся к общей памяти. Обмен с внешними устройствами производили процессоры ввода-вывода, которые фактически представляли собой специализированные ЭВМ. Максимально машина могла управлять четырьмя такими процессорами. Другие спецЭВМ — процессоры передачи данных — обеспечивали связь с пользователями.

В МВК использовано много неординарных решений, например, каждая величина, хранящаяся в памяти, снабжена дополнительным признаком — *тегом* (управляющим разрядом). В нём содержится информация о типе хранимой величины, а также признак защиты от чтения или записи. Архитектура центрального процессора имела много общего с аналогичными комплексами американской фирмы Burroughs.

В конце 70-х гг. в Советском Союзе началось производство универсальных многопроцессорных комплексов четвёртого поколения «Эльбрус-2». Производительность каждого процессора превышала 10 млн операций в секунду. Суммарная производительность могла достигать 100 млн операций в секунду.

Отечественное компьютеростроение испытывало трудности, связанные с необходимостью высококачественного промышленного изготовления электронных компонентов. Вероятно, поэтому был повторён не совсем удачный опыт фирмы IBM System/360 в виде серии ЕС ЭВМ (единой серии). Многие успешные (и не очень) решения копировались с западных аналогов.

Прообразом киевской мини-машины СМ-4 и зеленоградской «Электроники-79» стали машины серии PDP-11 фирмы DEC (США). Однако отечественные образцы уступали по основному критерию потребителя — надёжности. А с появлением персональных компьютеров бороться с всепроникающим IBM PC не смогли ни западноевропейские конкуренты, ни российские разработчики.

В 1961 г. под руководством академика В. М. Глушкова была создана полупроводниковая ЭВМ «Днепр». При этом установлен мировой рекорд разработки: от идеи до производства машины прошло всего три года.

Труды академика В. М. Глушкова по теоретической и прикладной кибернетике, теории цифровых автоматов (монографии «Синтез цифровых автоматов», статья «Абстрактная теория автоматов» в журнале «Успехи математических наук») заложили основы теории ЭВМ. Русский кибернетик Глушков стал известен в мире не меньше, чем американец Н. Винер.

СОВРЕМЕННЫЙ КОМПЬЮТЕР

ЧЕТВЁРТОЕ ПОКОЛЕНИЕ КОМПЬЮТЕРОВ

БОЛЬШИЕ И СВЕРХБОЛЬШИЕ ИНТЕГРАЛЬНЫЕ СХЕМЫ

Ещё в 1982 г. в статье журнала «Scientific American» Ху Мин Туг и Амар Гупта писали: «Если бы за последние 25 лет авиационная промышленность развивалась столь же стремительно, как и вычислительная техника, то "Боинг-767" можно было бы приобрести сегодня за 500 долларов и облететь на нём земной шар за 20 минут, израсходовав при этом 19 литров горючего. По этой аналогии, хотя и не совсем точной, можно судить о темпах снижения стоимости энергопотребления и роста быстродействия вычислительных машин.

Цены на логические элементы компьютеров ежегодно снижаются на 25 %, а для устройств памяти этот показатель ещё выше — 40 %. За 25 лет скорость вычислений возросла в 200 раз; при этом размеры и потребление электроэнергии у современных ЭВМ стали в 10 тыс. раз меньше, чем у машин сравнимой производительности 25-летней давности».

А начиналось всё в 60-х гг. XX столетия, когда доминирующей технологией в производстве компьютерных компонентов стали *чипы* — крошечные прямоугольники полупроводника, изготовленные из кремния и «упакованные» в изолирующий материал. Они получили название «интегральные схемы».

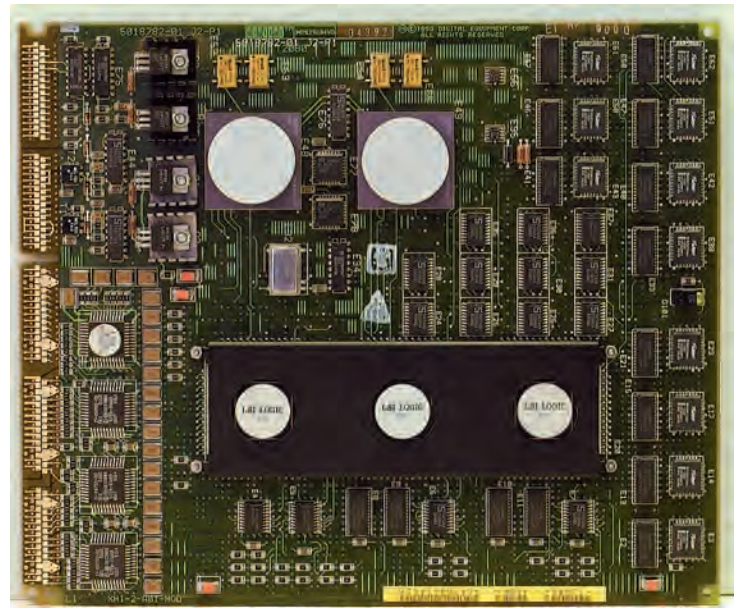
Основным элементом схемы являлся транзистор, поэтому «ёмкость» её измеряли в транзисторах. Технология производства интегральных схем прошла эволюцию от *малых интегральных схем (МИС),* содержащих лишь несколько транзисторов, до *больших интегральных схем (БИС),* объединяющих тысячи, сотни тысяч и более транзисторов, и *сверхбольших интегральных схем (СБИС).*

Термин «сверхбольшие интегральные схемы», появившийся в 1986 г., относился к схемам с памятью 1 Мбайт. Каждый такой чип содержал более миллиона транзисторов. Изготовлять интегральные схемы можно только на автоматизированном производстве, поэтому стоимость чипов была невелика. Массовое производство больших интегральных схем началось в 70-х гг. с выпуска памяти ЭВМ и микросхем для карманных калькуляторов.

Современный компьютер объединяет множество интегральных схем, имеющих металлические контакты для соединений, источник питания и другие мелкие элементы (резисторы, конденсаторы и др.). Интегральные схемы заключаются в корпуса, которые отличаются формой друг от друга.

Наиболее «старые» и поныне популярные интегральные схемы *(микросхемы)* — это ДИПы (*англ.* dual in-line package, DIP), в них выводы («ноги») организованы в два параллельных ряда с интервалом 2,54 мм между ближайшими контактами. Иногда выводы расположены только с одной стороны микросхемы, такой чип носит название СИПа (*англ.* single in-line package, SIP). Сложный, современный, требующий сотню или более контактов чип имеет иную конфигурацию, где выводы микросхемы расположены по всем её краям, он именуется ПГА (*англ.* pin-grid array).

Интегральные схемы можно соединять как при помощи пучков проводов, так и гораздо проще. Для этого на тонком листе изолятора (например, на стекловолокне) располагают интегральные схемы и другие элементы, чтобы контакты через подготовленные отверстия (сверления) выходили на другую сторону листа (платы), которая одновременно выполняет и функцию механического крепежа. Далее торчащие «ножки» необходимо соединить между собой по заранее разработанной электрической схеме. Соединения можно просто «нарисовать» каким-нибудь токопроводящим материалом, например медью. Такой «рисунок» и будет заменой соединительных проводов. При промышленном изготовлении подобная *печатная плата* (*англ.* printed-circuit board, PCB) намного дешевле и технологичнее, чем простое соединение проводниками, так как «рисование» можно произвести за один этап, если наложить трафарет, в котором будет вырезан рисунок соединений для всей платы целиком.

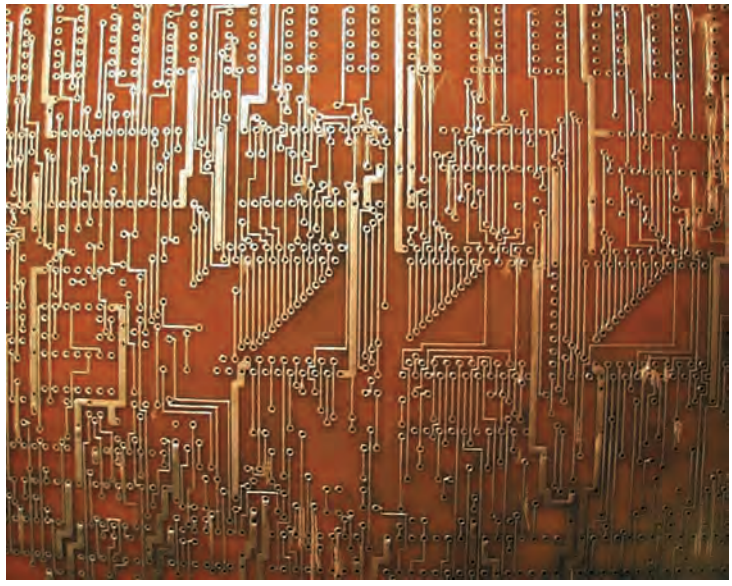


Плата ЭВМ VAX.

Иногда достаточно сложно без пересечений произвести все нужные соединения на плате, в таком случае используется многослойный монтаж. В мини-компьютерах были платы с семи- и даже десятислойным монтажом. Как правило, пара слоёв выделялась для подводки питания к микросхемам на плате, а остальные выполняли роль соединителей. Производство таких плат более трудоёмкое. Чем больше интегральных схем и элементов на плате, тем она сложнее.

Кроме того, существует и иная проблема. Чипы, как и транзисторы, при работе излучают тепло. Поэтому, чтобы избежать перегрева и выхода из строя микросхемы и даже целой платы, разработчики вынуждены помещать микросхемы на безопасном расстоянии друг от друга. Иногда для соблюдения температурного режима на интегральную схему устанавливается металлический радиатор или даже вентилятор.

Джек Килби, «отец» интегральной схемы, лишь в 2000 г. был удостоен Нобелевской премии за своё гениальное изобретение.



Печатная плата.

МИКРОПРОЦЕССОРЫ

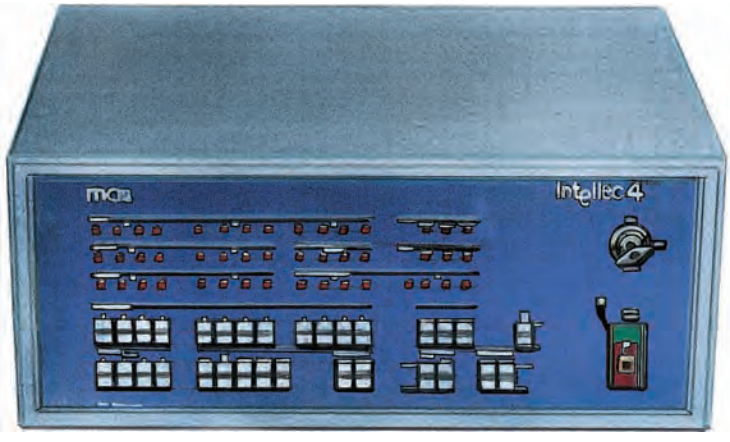
С возникновением интегральных схем впервые появилась возможность изготовить процессор (или даже целый компьютер) в одном-единственном чипе, т. е. создать микропроцессор.

Первым настоящим микропроцессором был Intel 4004, произведённый фирмой Intel (США) в 1971 г. Первоначально он использовался в калькуляторах в составе четырёх чипов набора MSC-4 (большие интегральные МОП-схемы). Длина машинного слова Intel 4004 всего 4 бит. «Уложить» все команды в одно такое слово просто невозможно: разнообразные комбинации битов 4-битного слова дают $2^4 = 16$, т. е. потенциально у этого «малыша» имелось 16 команд, причём со всеми методами адресации. Поэтому инструкции были длиной 8 бит: $2^8 = 256$. Данные и программа хранились в разной памяти. Всего 1 кбайт памяти отводился под данные и 4 кбайт — под программу. В Intel 4004 предусмотрен 12-битный счётчик команд (PC) и малюсенький стек глубиной всего в 4 вызова — для выполнения подпрограмм (CALL) и возврата из подпрограмм (RET), а также шестнадцать 4-битных регистров, которые могли группироваться в восемь 8-битных. Всего Intel 4004 имел 46 инструкций (команд).

После MSC-4 появилось большое количество микропроцессоров различных производителей. С увеличением плотности интегральных схем возрос до 32 бит размер машинного слова, расширился набор команд микропроцессоров, хотя по сравнению с «большими» процессорами он был невелик (например, отсутствовали команды для вычислений с плавающей запятой), увеличилась производительность микропроцессоров, т. е. число выполняемых команд в секунду.

Микропроцессоры уже могли управлять, например, компонентами подачи топлива («компьютеры» в автомобильных двигателях) и сложными бытовыми приборами. Они постепенно приобретали черты процессоров «больших» ЭВМ. С увеличением выпуска микропроцессоров появились их индустриальные стандарты.

Скромные возможности микропроцессора Intel 4004 объясняются ограничениями, наложенными плотностью интегральных схем (размер чипа 1971 г. — 1 см^2).



Машина Intellec 4 на микропроцессоре Intel 4004.

Z-80

В апреле 1974 г. фирма Intel выпустила 8-битный микропроцессор Intel 8080, превосходивший по всем показателям предыдущие процессоры. Но настоящую нетленную славу снискал себе микропроцессор Z-80, который разработала и произвела небольшая, по сравнению с Intel фирма Zilog. Он был полностью совместим с Intel 8080, т. е. умел всё, что Intel 8080, но и ещё многое... С этих пор фирма Intel перешла к разработке 16-битных, а затем 32- и 64-битных процессоров,

ставших основными для современных персональных компьютеров.

Z-80 применяется и в настоящее время, а в 80-х гг. он произвёл фурор. 8-битный микропроцессор использовал 16-битную адресацию, т. е. мог напрямую обращаться к памяти в 64 кбайт. Было добавлено более 80 новых инструкций по сравнению с Intel 8080: операции с 1-, 4-, 8- и 16-битными данными, а также блоковые (для пересылки за одну команду до 64 кбайт данных) или специальные команды для поиска нужного байта в памяти. Процессор имел два набора регистров и операции переключения между ними, что упрощало организацию обработки прерываний (не требовалось запоминать состояние регистров в программе). В Z-80 также было добавлено два 8-битных индексных регистра (IX и IY).

Быстродействие микропроцессора определялось особенностями интегральной схемы и при функционировании задавалось кварцевым генератором — часами. Это называлось тактовой частотой процессора.

Первый Z-80 имел тактовую частоту 2,5 МГц, более поздний Z-80C — уже 8 МГц. Самая «короткая» команда пересылки из одного регистра в другой выполнялась за четыре такта процессора. Первый Z-80 осуществлял более 600 тыс. команд в секунду — весьма впечатлительно для «малыша».

«КИСКИ» И «РИСКИ»

Большинство микропроцессоров относится к стандарту архитектуры CISC (*англ.* Complex Instruction Set Computer) с полным набором команд. Её прародителем принято считать IBM.

Фирма Intel — лидер в разработке CISC-микропроцессоров — знаменита своими сериями x86 и Pentium.

Для CISC-процессоров характерно небольшое число регистров, но при этом количество команд, методов адресации с большим уровнем косвенности, форматов команд остаётся большим.

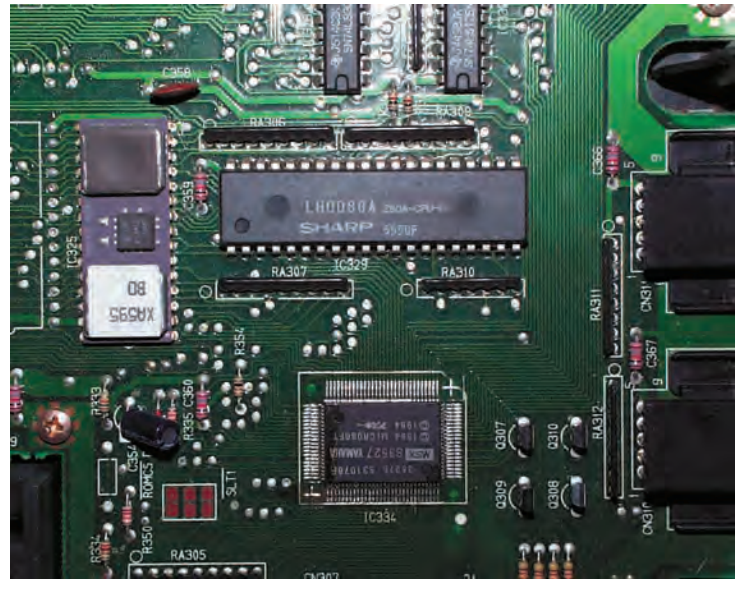
Данная архитектура является стандартом для микрокомпьютеров.

RISC (Reduced Instruction Set Computer) — процессоры с сокращённым набором команд. Они имеют упрощённую логику-схему — очень простые инструкции и методы адресации без многоуровневой косвенности. Благодаря этому процессор становится меньше, проще и соответственно быстрее.

История RISC восходит к компьютерам CDC 6600, когда Сеймур Крей предложил упростить набор команд для ускорения вычислений, с успехом применив эту архитектуру при создании широко известной серии суперкомпьютеров Cray. Однако окончательно RISC-архитектура сформировалась после трёх исследовательских проектов: процессора 801 компании IBM, процессора RISC Калифорнийского университета (Беркли) и процессора MIPS Станфордского университета.

Технология RISC, которую впервые представила IBM в 1970 г., никогда не публиковалась — компьютер на её основе серийно не изготовлялся.

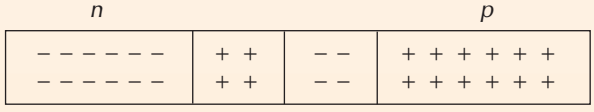
В 1980 г. Дэвид Паттерсон с коллективом Калифорнийского университета (Беркли) создал две машины, получившие названия RISC-I и RISC-II. Главной идеей этих ЭВМ было отделение медленной памяти от высокоскоростных регистров, создание отдельных «быстрых» команд для



Фрагмент платы компьютера MSX на процессоре Z-80.

ТИПЫ ИНТЕГРАЛЬНЫХ СХЕМ

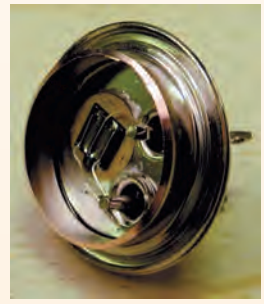
Полупроводники, из которых изготовляют транзисторы и диоды, разделяются на полупроводники с электронной — *n*- (от *англ.* negative — «отрицательный») и дырочной — *p*- (от *англ.* positive — «положительный») проводимостью. Принцип действия полупроводниковых диодов основан на свойствах *p* — *n*- перехода, когда в контакте находятся два полупроводника *p*- и *n*- типа. В месте контакта происходит диффузия положительных зарядов (дырок) из области *p* в область *n*, а электронов обратно, из *n* в *p*. Однако без внешнего воздействия (электрического тока) процесс стабилизируется, потому что образуется так называемый *запирающий слой*.



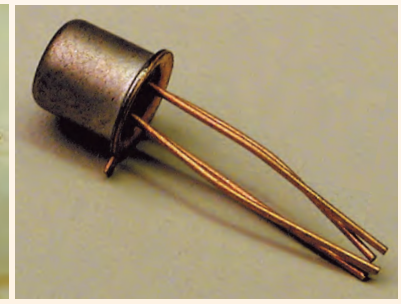
При подключении к области *p* «плюса» источника электрического тока, а к *n* соответственно «минуса» запирающий слой разрушится, и такой диод будет проводить ток. Если осуществить подключение источника питания наоборот, т. е. к *p* — «минус», а к *n* — «плюс», то ток будет фактически равен нулю, запирающий слой расширится. Это основное свойство полупроводниковых диодов позволяет применять их в качестве выпрямителей тока. Большинство полупроводников делается из кремния и германия с различными добавками, из оксидов некоторых металлов и из собственно металлов. В зависимости от добавок они имеют *n*- или *p*-тип.

Транзистор представляет собой трёхслойную структуру из таких же полупроводниковых материалов, однако в основе его работы лежит не один, а два *p* — *n*- перехода. Внешние слои называют *эмиттером* и *коллектором*, а средний (обычно очень тонкий, порядка нескольких микрон) слой — *базой*. При подключении к транзистору источника тока малые изменения тока, протекающего от базы к коллектору, вызовут аналогичные, но бóльшие изменения тока от эмиттера к коллектору. Это свойство транзистора позволяет использовать его в качестве усилителя тока.

Простейший трёхслойный транзистор называется биполярным. Если база относится к полупроводникам *p*-типа, а эмиттер и коллектор — *n*-типа, то транзистор имеет тип *n* — *p* — *n*, а если, наоборот, внутри находится материал типа *n*, а по краям — типа *p*, то тип транзистора *p* — *n* — *p*. Основной недостаток биполярного транзистора — большое потребление энергии и выделение тепла.



Транзистор в разрезе.

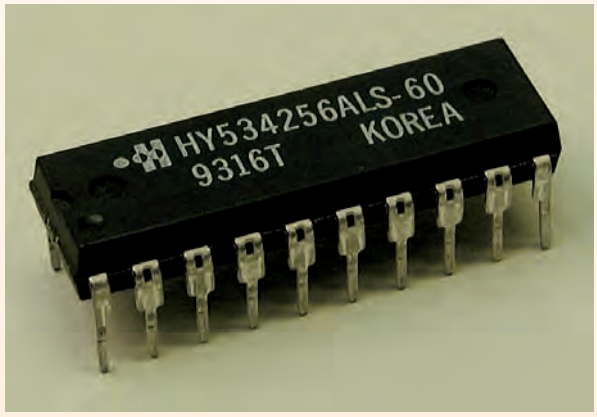


Полевой транзистор.

В качестве альтернативы был разработан полевой транзистор. Он представляет собой однополярный полупроводниковый прибор, выводы которого называются *исток*, *сток*, *затвор*. При подаче напряжения на затвор и сток (или соответственно исток) носители заряда, электроны в областях с проводимостью *n*-типа (или дырки в областях с проводимостью *p*-типа), проходят через возникающий под затвором тонкий проводящий канал. Как правило, ток в канале изменяется посредством электрического поля, создаваемого напряжением, приложенным к переходу (между затвором и стоком). Если сток и исток изготовлены из полупроводника *n*-типа, то транзистор называют *n*-канальным, если *p*-типа — *p*-канальным.

Существуют полевые транзисторы, у которых затвор отделён от канала слоем диэлектрика. Такие транзисторы состоят из пластины полупроводника (подложки) с высоким удельным сопротивлением, имеющей две области с противоположным типом электропроводности, куда нанесены электроды — исток и сток. Поверхность полупроводника между ними покрыта тонким слоем диэлектрика (обычно оксид кремния SiO_2), на который нанесён металлический электрод — затвор. Поэтому полевые транзисторы с изолированным затвором — МДМ (металл — диэлектрик — полупроводник) часто называют МОП-транзисторами (металл — оксид — полупроводник). Используют ещё слоистые диэлектрики, например SiO_2 — Al_2O_3, которые соответственно называются МАОП-транзисторы (металл — алюминий — оксид — полупроводник). В отличие от биполярных транзисторов МОП-транзисторы более экономичны.

Транзистор, изобретённый в 1948 г., лежит в основе всех современных микросхем и микропроцессоров. Недаром его авторы — Уильям Шокли (1910—1989), Уолтер Браттейн (1902—1987) и Джон Бардин (1908—1991) получили Нобелевскую премию по физике в 1956 г. В вычислительной технике состояние транзи-



Микросхема.

стора, когда через коллектор течёт большой ток, можно условно принять, например, за 1, а малый — за 0. Но условия использования транзистора не ограничиваются только потреблением им энергии при хранении информации (0 или 1), важно также и время, которое потребуется для перевода его из одного состояния в другое: из 0 в 1 или из 1 в 0.

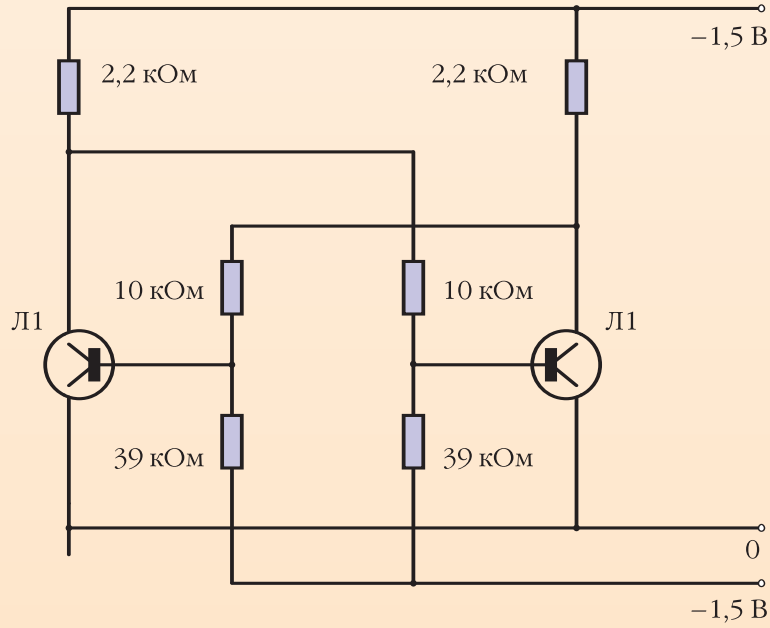
Вначале транзисторы изготовлялись как отдельные элементы и представляли собой цилиндры диаметром в десяток миллиметров с несколькими проволочными выводами.

Однако в вычислительной технике используются не отдельные транзисторы, а их комбинации. Обычно соединяют два МОП-транзистора: *p*-канальный и *n*-канальный. Такую пару называют КМОП-парой (комплиментарные МОП). КМОП-пара хороша тем, что практически не потребляет энергии в интервалах между переключениями, т. е. пока находится в одном из двух устойчивых состояний, которые соответствуют 0 или 1. Применение КМОП-технологии позволяет существенно повысить производительность, понизить энергопотребление схем и соответственно уменьшить выделение тепла. Эта технология является основой так называемых TTL-схем (транзисторно-транзисторная логика). Чем меньше размеры МОП-устройств, тем больше их можно расположить на единице площади, или, как говорят, увеличить плотность упаковки. Малое выделение тепла КМОП-парами позволяет объединить на одном кристалле микросхемы многие миллионы элементов. При этом можно избежать опасного перегрева схем, что приводит к их выходу из строя.

Из транзисторов составляют вентиль — электрическую схему, преобразующую два и более входных сигнала в один выходной (например, логическое «И», «ИЛИ» и пр.), или триггер, схему для запоминания одного бита, т. е. находящуюся в одном из двух устойчивых состояний: 0 или 1. Триггер можно составить и из пары транзисторов, а на составление логического вентиля средней сложности обычно идёт четыре — восемь транзисторов.

Интегральные схемы (чипы) классифицируются по плотности, которая определяется числом транзисторов, включённых в них, или числом логических вентилей. Если говорят, что на кристалле 100 тыс. транзисторов, значит, можно сказать, что схема содержит 50 тыс. КМОП-пар, или 12,5 тыс. логических вентилей.

Хотя большинство интегральных схем традиционно производится из кремния, может также использоваться полупроводник из другого материала — арсенида галлия. Такие интегральные схемы сложнее в производстве, но зато имеют более высокие скоростные показатели (примерно в пять раз выше, чем у кремниевых схем).



Так можно собирать триггеры.

обработки данных (вычислений) и отдельных «медленных» команд для загрузки и выгрузки данных из регистров в память и обратно.

В 1981 г. Джон Хеннесси со своими коллегами представил описание «станфордской машины MIPS», разработка которой основывалась на конвейерной обработке данных с предварительным, тщательным планированием «загрузки» процессора.

Эти машины имели много общего. Система команд процессоров была устроена так, чтобы выполнение команды занимало минимальное количество тактов (в идеале один). При этом с целью повышения производительности команды реализованы аппаратно, без применения техники микропрограммирования, а чтобы максимально упростить реализацию и увеличить скорость выполнения, использовались команды только фиксированной длины.

Среди других особенностей RISC-процессоров следует отметить наличие 32 и более регистров (по сравнению с 8—16 регистрами в CISC-процессорах), что позволяет хранить больший объём данных. Для обработки информации, как правило, используются трёхадресные команды, что даёт возможность сохранять большее число переменных в регистрах без перезагрузки.

Простота и быстродействие RISC-процессоров, подтверждённые университетскими проектами, вызвали к ним интерес производителей вычислительной техники, и с 1986 г. начался выпуск RISC-процессоров.

Одним из таких процессоров был Motorola PowerPC 601, разработанный совместно фирмами Motorola и IBM. Это 32-битный RISC-процессор с тактовой частотой 80 МГц. Следующий, PowerPC 603 (КМОП-чип), в два-три раза более экономичный по сравнению с 601, состоящий всего из 1,6 млн транзисторов, имел также тактовую частоту 80 МГц.

Дальнейшие разработки Motorola и IBM возникали одна за другой: PowerPC 604 (100 МГц) появился спустя несколько месяцев. PowerPC 615 имел аппаратную совместимость с популярным процессором 8086 фирмы Intel, создав ей весьма солидную конкуренцию, а PowerPC 620 был уже 64-битным процессором. Развитие RISC-процессоров в значительной степени определялось совершенствованием специальных оптимизирующих компиляторов. «Умный» компилятор обеспечивал эффективный программный код, распределяя множество данных по большому количеству регистров RISC-процессора, стараясь как можно меньше использовать команды выгрузки и загрузки данных из «медленной» памяти компьютера. Подобно конвейерным системам, компилятор должен был так преобразовать программу, чтобы

задать код для одновременного выполнения различных вычислений на конвейере процессора.

Для совместимости с распространёнными CISC-процессорами RISC-процессоры имели специальный режим эмуляции CISC-процессора, при котором скомпилированная программа даже не «подозревает», что выполняется не на том процессоре.

Интересно отметить, что в разработках компании Intel (Pentium и Pentium Pro), а также её последователей-конкурентов (AMD R5, Cyrix M1, NexGen NX586 и др.) используются идеи конвейерной обработки, реализованные в RISC-микропроцессорах, так что многие различия между CISC и RISC постепенно стираются.

МИКРОКОМПЬЮТЕРЫ

Типичный микрокомпьютер с достаточной памятью и развитой периферией (множеством портов для устройств ввода-вывода) содержит порядка 50 интегральных схем и может быть реализован на единственной печатной плате. Такие машины называются одноплатными микроЭВМ.

Микрокомпьютер с маленькой памятью и небольшим числом портов ввода-вывода может быть изготовлен в одном-единственном СБИС-чипе. Одночиповые микроЭВМ получили широкое распространение благодаря компактным размерам, низкой цене и простоте ремонта (выбросить неисправный, установить новый). Их производство началось в середине 70-х гг., вскоре после появления первых микропроцессоров. Эти микроЭВМ использовались в основном в качестве программируемых контроллеров для различных машин и механизмов (например, для станков с числовым программным управлением, микроволновых печей и т. д.).

Появился новый класс ЭВМ широкого применения — персональные компьютеры. Это маленькие и дешёвые микрокомпьютеры, которые предназначаются для установки в офисе, в кабинете учёного, на парте школьника. Некоторые из них имеют

Компьютер Sinclair ZX Spectrum.

питание от аккумуляторов и настолько компактны, что их можно брать с собой в путешествие.

Типичный персональный компьютер включает клавиатуру, видеомонитор и системный блок (где и размещается плата с микропроцессором). Для связи с внешним миром компьютер использует телефонные линии, для хранения данных — различные магнитные диски, для ввода графической информации — сканеры и манипуляторы, для создания твёрдых копий — принтеры и графопостроители.

В Великобритании в 80-х гг. была очень популярна модель, разработанная Клайвом Синклером, — Sinclair ZX Spectrum. Машина размещалась в одном корпусе с клавиатурой, а в качестве монитора использовался домашний телевизор. Она стоила всего несколько сотен долларов, что делало её доступной по цене практически каждому. Домашний компьютер имел

Одна из первых микроЭВМ, Intellec 8 (базировавшаяся на процессоре Intel 8008), стоила около 2,5 тыс. долларов.

Компьютер Apple I.

48 кбайт памяти и базировался на микропроцессоре Z-80.

Другая известная машина на том же микропроцессоре — MSX-компьютер, появившийся в Японии в 1982 г. Эта ЭВМ была широко распространена в Японии, Корее, Южной Америке, Нидерландах и Франции. Четыре тысячи MSX, поставленных в СССР, стали базой для школьной информатики с 1986 г. В MSX-компьютерах использовался процессор Z-80 с тактовой частотой 3,58 МГц. Видеопроцессор (процессор, формирующий изображение на экране), разработанный фирмой Texas Instruments (чипы TI9918 или TI9928), намного превосходил видеопроцессоры IBM PC XT и AT (1983—1988 гг.) по количеству воспроизводимых цветов. Эта машина была снабжена аудиосистемой — трёхканальным музыкальным синтезатором, что считалось редкостью для ЭВМ. MSX-1 имел 64 кбайт памяти, а MSX-2 — 128 кбайт (максимальный объём памяти MSX-2 — 4 Мбайт). Вероятно, MSX-машины — лучшие компьютеры, построенные на 8-битном процессоре.

Машина серии IBM PC, впервые представленная в 1981 г., в настоящее время фактически стала стандартом для персональных компьютеров. Однако видеопроцессоры первых моделей IBM PC XT (базирующейся на процессоре Intel 8086) и IBM PC AT (базирующейся на процессоре Intel 80286) сильно проигрывали по сравнению с другими. Правда, после того как машины, подобные IBM PC, стали производить на микропроцессорах фирмы AMD (*англ.* American Micro Device), машины заметно улучшили свои характеристики.

Другой знаменитый персональный компьютер — Macintosh. В США родоначальниками персональных компьютеров считаются Стив Джобс и Стефан (Стив) Возняк. 1 апреля 1976 г. они основали фирму Apple в Пало-Альто (штат Калифорния). Первый компьютер, собранный в их знаменитом гараже, назывался Apple I и имел всего 8 кбайт памяти. В 1977 г. они разработали, собрали и продали первый настоящий компьютер Apple II. Он подключался прямо к телевизору. Процессор марки 6502, используемый в этой машине, стоил всего 25 долларов. Цена являлась определяющим фактором в его выборе. Возможности процессора были достаточно слабыми: пять регистров, из них два служебных (PC и SP), остальные 8-битные.

В 1983 г. фирмой создан компьютер Lisa со странным устройством-манипулятором для ввода графической информации. Большинство учёных и разработчиков скептически отнеслись к новинке. Название этого устройства — «мышь». Стоил Lisa очень дорого — около 10 тыс. долларов. В 1984 г. появился менее дорогой компьютер — Macintosh. Он имел чёрно-белый монитор, аудиосистему и операционную систему с графическим интерфейсом. В 1994 г. был представлен первый Macintosh на базе RISC-процессора Motorola PowerPC.

▶ Электронная записная книжка Xircom также использует процессор Z-80.

Персональные компьютеры всего за несколько десятков лет полностью захватили мир, они теперь столь же привычны в офисах, как некогда пишущие машинки. Возможности персональных компьютеров помимо текстообработки в последние годы расширились за счёт мультимедиаприложений и средств мировой коммуникации. Персональные компьютеры становятся практически такими же мощными, как и большие машины. Круг задач, решаемых ими, расширяется.

СБИС-технология позволила изготовлять центральные процессоры, память и прочие устройства отдельными СБИС; это делает возможным массовое производство компьютеров, доступных по цене практически каждому. В результате развития СБИС-технологий появились как новый класс ЭВМ — недорогие персональные компьютеры, так и высокоскоростные параллельные вычислительные системы, содержащие тысячи процессоров. Термин «четвёртое поколение» часто считается синонимом машин, основанных на СБИС-технологиях.

Компьютер Lisa.

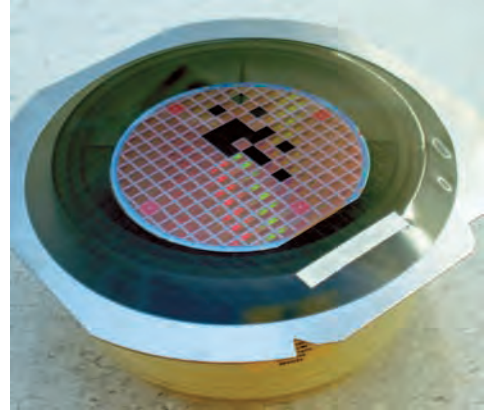
ПРОИЗВОДСТВО МИКРОПРОЦЕССОРОВ

Производство интегральных схем очень сложный и дорогостоящий процесс. Микропроцессоры изготовляются поэтапно. На каждом этапе производства обрабатываются сразу сотни интегральных схем — целая кремниевая пластина, состоящая из многих одинаковых заготовок — *кристаллов*. Любой кристалл образован миллионами элементов — транзисторов, т. е. за одну технологическую операцию обрабатываются миллиарды элементов.

Перед началом производства разработчик интегральной схемы представляет электронную схему, скомпонованную из транзисторов, их соединений и других элементов в виде плоской многослойной конструкции. Каждый слой конструкции — сложный рисунок, выполненный из определённого материала (полупроводник, металл, изолятор и др.).

Интегральная технология не только дала жизнь новому массовому производству и сборке компьютеров из стандартных чипов и микропроцессоров, но и позволила осуществлять быстрое и недорогое проектирование и производство чипов для компьютеров, в том числе нестандартных (заказные, специализированные).

Разработка каждого нового микропроцессора требует огромных вычислительных мощностей и применения самой современной техники. Производство микросхем нового поколения возможно лишь с помощью микросхем предыдущего поколения и, в свою очередь, открывает дорогу для разработки следующего поколения.



В наши дни технологии полиграфического производства переносятся в мир информационных технологий. Почтовые марки печатаются не по одной, а блоками: несколько рядов по нескольку марок в ряду. Так и микропроцессоры (интегральные схемы): их делают не поштучно, а группами, расположенными на одной пластине полупроводника, обычно кремния.

Кремниевые пластины перед обработкой.

Фотошаблоны представляют собой тонкие стеклянные пластины с нанесённой схемой микропроцессора. Они изготовляются на высокоточном оптическом оборудовании, рисунок наносится электронным лучом.

Сложным путём интегральная схема изготовляется на специализированном производстве, слой за слоем повторяя разработанную схему. Предварительный этап начинается с заготовки. К кремниевой пластине предъявляются высокие требования. Свойства полупроводника сильно различаются в зависимости от ориентации его кристаллической решётки. При изготовлении пластины, представляющей собой тонкий срез, выполняемый алмазной пилой из искусственно выращенного монокристалла кремния диаметром 200 или 300 мм, важно её правильно ориентировать. Чем больше размер пластины, тем большее количество кристаллов (микросхем) умещается на ней. Однако это требует применения высокотехнологичного оборудования, так как на этапах обработки пластины важно сохранять равномерность воздействия (например, обеспечивать постоянную температуру) по всей её поверхности.

Перед тем как пластина примет участие в первой технологической операции, её готовят: моют, полируют, сушат и пр. Заготовка должна быть идеально чистой — это обязательное условие во всём процессе производства. Любая пылинка, попавшая в кристалл, может привести к браку, потому что расстояния между элементами чрезвычайно малы.

На первом этапе пластину, состоящую из чистого кремния, покрывают тонкой плёнкой диэлектрика — диоксида кремния (SiO_2). Для этого заготовки окисляют в камере при давлении 25 атм и температуре около 1000 °C или используют другой метод — окисление водяными парами. Он состоит в том, что кислород под давлением обдувает поверхность пластины парами воды. Чтобы полученная тонкая плёнка изолятора была равномерной (без дефектов), надо чётко выдержать одинаковую температуру по всему объёму камеры.

ФОТОЛИТОГРАФИЯ

Это одна из основных операций планарной технологии, которая позволяет обеспечить выборочность воздействия на пластину при технологических операциях.

ПРОЕКТИРОВАНИЕ ИНТЕГРАЛЬНЫХ СХЕМ

Разумеется, кристалл сложностью в несколько миллионов элементов нельзя разработать вручную. Для этого нужны мощные компьютеры и специализированные программные системы автоматизированного проектирования интегральных схем — САПР (*англ.* Computer-Aided Design, CAD).

В процессе проектирования выделяют три этапа.

Задание интегральной схемы. Проектируется схема на CAD-компьютере. Разработчик рисует её на экране. Как простая схема, так и сложный микропроцессор задаются поэлементно. Элементы схемы либо спроектированы человеком самостоятельно, либо для этого выбраны некоторые стандартные элементы из базы данных. Или, например, интегральная схема просто берётся из файла со своим прообразом и корректируется. При проектировании используется специальный язык описания схемы микропроцессора.

Верификация интегральной схемы. Это процесс сверки интегральной схемы. Созданная схема тестируется на соответствие проекту с помощью специализированного программного обеспечения. Причём производится не только проверка на соответствие проекту, но проверяются и временные показатели так, чтобы операции компонентов будущего микропроцессора были согласованы между собой. На этом этапе человек может вносить поправки в проект, а затем снова проверять интегральную схему, улучшать или полностью менять её части. Вручную, без применения CAD-компьютера, данный процесс выполнить практически невозможно. Результатом второго этапа будет файл, поэлементно описывающий интегральную схему.

Подготовка производства интегральной схемы. Как правило, база данных CAD-компьютера содержит описания различных производств интегральных схем. Файл с проектируемой схемой преобразуется в специальный формат автоматизированного процесса производства интегральной схемы. Остаётся только, используя этот файл, заказать изготовление на заводе *фотошаблонов* спроектированной интегральной схемы.

Для изготовления на пластине транзисторов определённого типа необходимо внести примеси в кремний, чтобы получить полупроводниковые приборы с *p*- или *n*-типом проводимости. Такое внедрение должно быть точечным, при этом области, в которые не следует вносить примеси, надо прикрыть, например, изолятором. Тогда можно закрыть целиком всю пластину и вытравить в ней окошки перед операцией *травления* (процесс удаления с поверхности пластины ненужных участков полупроводника, диэлектрика, алюминиевых или медных проводников). Это и происходит при *фотолитографии* — переносе рисунка с эталонного фотошаблона на поверхность пластины.

После обработки на пластине останется замысловатый рисунок из плёнки специального светочувствительного материала, восприимчивого к действию ультрафиолетовых лучей. Следовательно, если участки пластины облучить (проэкспонировать) ультрафиолетовой лампой, то покрывающий их материал станет устойчивым к травлению.

Фотолитография — сложный процесс, состоящий из нескольких этапов. Поверхность пластины целиком покрывается специальным материалом с использованием центрифуги: пластина помещается на вращающийся горизонтальный диск, а сверху из форсунок распыляется покрывающий материал; под действием центробежной силы жидкий раствор растекается от центра к краям пластины.

Сушка этого слоя выполняется в печи под воздействием точно рассчитанных доз инфракрасного излучения. Затем пластину совмещают с фотошаблоном. Данную операцию раньше проводили вручную с помощью микроскопа, ориентируясь на метки на пластине и фотошаблоне. Однако сейчас её уже не доверяют человеку.

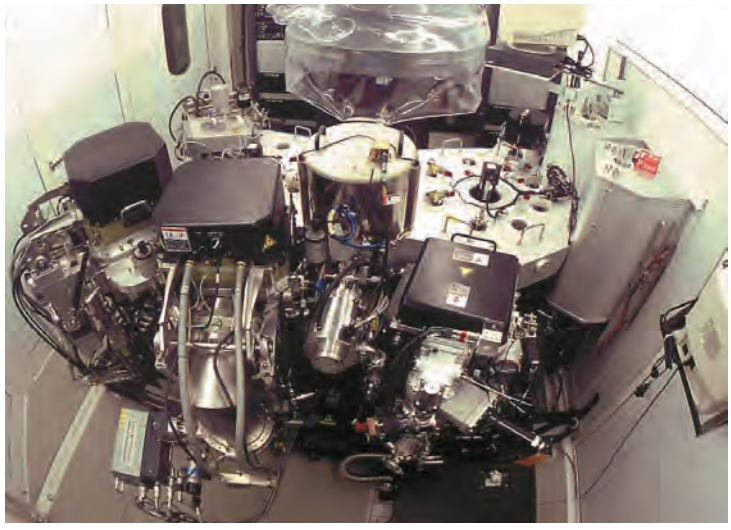
После того как шаблон строго совмещён с пластиной, поверхность облучают ультрафиолетом, и рисунок будущего слоя остаётся запечатлённым. Участки, на которые попал ультрафиолетовый свет, стали устойчивыми к травлению, они будут защищать слой диоксида кремния во время следующего этапа.

Два главных химических свойства пары кремний — диоксид кремния ($Si - SiO_2$) являются основой кремниевой технологии. Первое: возможность выборочного воздействия на один из этих материалов (так, плавиковая кислота растворяет диоксид кремния, но не действует на сам кремний); второе: диоксид кремния можно использовать для защиты участков кремния при легировании (внесении примесей).

Современная технология при фотолитографии может использовать защитный материал, который становится устойчивым к травлению при облучении или теряет защитные свойства.

Кроме фотолитографии существуют другие методы подготовки пластины к травлению, например *электронно-лучевая литография* и *лазерная литография*. При этом в отличие от оптической литографии не производится одновременное экспонирование всего рисунка интегральной схемы, а электронный (или лазерный) луч последовательно проходит весь рисунок под управлением компьютера, т. е. не требуется использование никакого фотошаблона. Из-за небольшой скорости обработки (на одной

В производственных помещениях должна быть абсолютная чистота.



Один из этапов производства интерактивных схем.

установке можно обработать около 60 пластин в год в отличие от фотолитографии при которой скорость обработки порядка десятков тысяч пластин в год) такая техника в основном используется для изготовления фотошаблонов.

ТРАВЛЕНИЕ

Может использоваться жидкостное травление, например, азотной кислотой. Пластину обрабатывают так же, как радиолюбители травят свои печатные платы. На покрытый токопроводящим слоем (фольгированный) изолятор (текстолит) наносят рисунок дорожек будущей платы, затем плату опускают, например, в азотную кислоту. Ненужные участки фольги стравливают, обнажая текстолит.

При этом на результат травления влияет много факторов. Это и концентрация кислоты, и температура. При неточном соблюдении параметров кислота постепенно проникает под край маски из защитного слоя, как бы подтачивая нужные соединения.

Сухое (плазменное) травление с применением ионизированного газа позволяет вести процесс в одном направлении, строго сверху вниз. Газ вступает в реакцию с обрабатываемой поверхностью, а летучие побочные продукты поглощаются камерой, в которой происходит травление.

Ионное легирование — ещё один современный способ добавления примеси в кремний. Он не требует высокой температуры и позволяет осуществлять поверхностное (слоем тоньше 1 мкм) распределение легирующих веществ в полупроводнике.

Травление повторяется многократно в процессе производства, причём не обязательно в указанной последовательности. Путём комбинирования операций внедрения примесей, напыления слоёв проводника, травления ненужных участков на кремниевой поверхности создаётся сложнейший, многослойный узор. Он является точным повторением разработанного проекта интегральной схемы, состоящей из переплетения полупроводников, изоляторов и проводников.

В результате травления диоксида кремния в диэлектрике появляются отверстия, обнажающие участки кремния. Затем удаляются остатки фотозащитного слоя, и кремний готов для внедрения примесей.

ДИФФУЗИЯ (ВНЕДРЕНИЕ ПРИМЕСЕЙ)

Цель внедрения примесей — изменение электрических характеристик кремния, чтобы чистый полупроводник стал полупроводником *p*- или *n*-типа. Для получения полупроводника *n*-типа можно использовать фосфор, а для *p*-типа — атомы бора. Диффузия состоит из двух шагов: *предосаждения* и собственно диффузии.

На первом этапе в поверхность кремния внедряется небольшое количество *легирующего вещества*. Как и на многих других этапах обработки пластины, её поверхность должна быть тщательно очищена, чтобы нежелательные примеси не попали в кристалл. При температуре около 1000 °C легирующее вещество проникает сквозь кристаллическую решётку кремния. На втором этапе происходит диффузия — «разгонка», при которой примеси глубже проникают и равномернее распределяются в кремнии.

СОЗДАНИЕ СОЕДИНЕНИЙ

Отдельные приборы (транзисторы, КМОП-пары), полученные при планарной технологии, необходимо соединить между собой и, возможно, с внешними выводами.

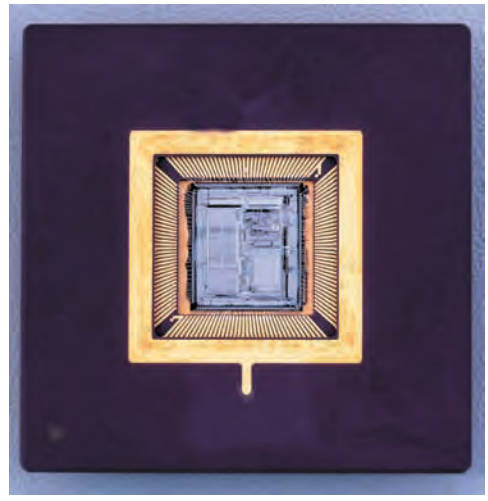
Для этого с участков, где должен получиться контакт с кремнием, удаляют диоксид, затем на поверхность пластины наносится осаждением слой металла. Металл нагревают до высокой температуры, при этом он испаряется. Затем, охлаждаясь, газообразный металл конденсируется на кремниевой пластине в виде однородной плёнки. Потом металл (обычно алюминий) с помощью фотолитографии и травления удаляется со всех участков, где его не должно быть. Вся операция носит название *металлизация*.

Так как плотность транзисторов, размещённых в интегральных схемах, всё время увеличивается, то возрастает число соединений между элементами, которые не всегда удаётся разместить на одном металлическом слое. Может потребоваться второй, третий и большее число слоёв. Указанные технологические операции многократно повторяются, при этом слои металлизации изолируются друг от друга диэлектриком.

После проведённых циклов технологических операций получается пластина с сотнями одинаковых маленьких прямоугольников (каждый прямоугольник — будущий микропроцессор или интегральная схема другого назначения).

Пластине предстоит пройти ещё несколько операций: резку на отдельные чипы, сборку интегральных схем, тестирование. Последнее покажет, сколько микропроцессоров пойдёт в брак, а сколько будет использовано в производстве компьютеров, автомобилей, микроволновых печей и т. д.

При травлении окон в кремнии на поверхности образуются ступеньки, которые влияют на качество контактов с пластиной. Новый метод позволяет проводить металлизацию по так называемой дамасской технологии, получившей своё название из-за определённого сходства с гравировкой металла. На поверхность диоксида с предварительно протравленными дорожками осаждается слой металла, при этом он заполняет углубления, создавая контактные соединения. Затем поверхность полируется и лишний металл удаляется. Этот метод позволяет увеличить выход годных микросхем. В начале XXI в. стала использоваться более прогрессивная разновидность этой технологии.

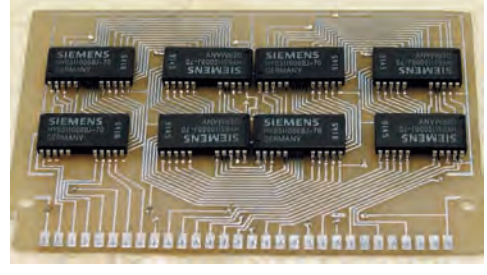


Микропроцессор.

ЭВОЛЮЦИЯ ПАМЯТИ ЭВМ

Перефразируя известное выражение, можно сказать, что «лучше купить дешёвый компьютер с большой и быстрой памятью, чем дорогой компьютер с маленькой и медленной памятью». Но большая память стоит дороже, чем маленькая, а быстрая — дороже, чем медленная. Поэтому производителям и потребителям приходится искать компромисс между ценой, объёмом и скоростью памяти, стараясь скомбинировать разные виды памяти так, чтобы компьютер работал оптимально.

Компоненты компьютера, связанные проводами, обмениваются



Память.

Шина обслуживает не только процессор и память, но и другие компоненты ЭВМ. Кроме передачи данных шина предназначена для согласования скоростей работы подключённых к ней устройств. Для этого она вырабатывает собственные сигналы, позволяющие понять, готова ли шина передать или принять информацию. Важно, чтобы производительности шины и всех подсоединённых к ней устройств были согласованы. Неразумно иметь быстрый процессор и медленную память или быстрые процессор и память, но медленную шину.

Ферромагнетики — вещества (например, железо, никель, кобальт) с большой магнитной проницаемостью, способные к намагничиванию под действием внешнего поля. Французский учёный Андре Мари Ампер высказал гипотезу, что внутри молекул и атомов циркулируют элементарные токи. При намагниченном состоянии вещества элементарные токи ориентированы так, что их действия складываются.

информацией между собой, используя электрические сигналы. Если принять за аксиому, что процесс передачи информации по проводам *(шине)* между процессором и памятью не зависит от собственных характеристик шины (числа проводов, скорости прохождения сигнала и пр.), что, конечно же, не так, то память оценивают по трём основным критериям: стоимость, объём и время доступа.

Из всего многообразия принципиально разных видов памяти можно выделить две основные группы: *оперативную память* и *внешнюю память*. К оперативной памяти процессор компьютера обращается часто, считывание и запись информации в неё происходят достаточно быстро. Во внешней, долговременной, памяти компьютер хранит большие объёмы информации и, можно сказать, обращается к ней эпизодически. По экономическим соображениям оперативная память всегда меньше по объёму, чем внешняя.

На протяжении всей истории вычислительной техники как оперативная, так и внешняя память непрерывно эволюционируют: повышаются производительность и объём, уменьшается стоимость. Применяемые технологии сменяют друг друга, как кадры в фильме. Память на электронно-лучевых трубках и ультразвуковых линиях задержки уступает место памяти на ферритовых сердечниках. Вместо ферритовой памяти приходит память на полупроводниках и интегральных КМОП-схемах. То же происходит и с долговременной памятью.

В 60—70-х гг. полупроводниковая память была дорогостоящей, и, кроме того, её содержимое очищалось при выключении питания (не сохранялись ни данные, ни программы). Поэтому искали новое решение проблемы. Множество разработок приводили к созданию, казалось, быстрой и надёжной памяти, но она тут же вытеснялась новой технологией. Однако случались и неудачи, которые не оправдывали ожиданий.

Так выглядят домены пузырьковой памяти.

ПУЗЫРЬКОВАЯ ПАМЯТЬ

К одной из этих неудач относят так называемую пузырьковую память (*англ.* bubble memory), разработанную в Bell Labs (США) в 1966 г. Устройство запоминало данные в виде магнитных доменов на тонком магнитном материале, например таком, как феррогранат.

Если пластину предварительно намагнитить, а затем приложить к ней определённого уровня магнитное поле в направлении намагниченности, то домены из кривых линий примут форму пузырьков, отсюда и

название памяти. На самом деле домен имеет форму цилиндра, его ось перпендикулярна пластине. Наличие домена в определённой точке материала принималось за 1, а отсутствие — за 0. На 1 см^2 плёнки могло помещаться до 16 млн доменов, иначе говоря, ёмкость пузырьковой памяти была до 16 Мбайт.

Такая память обладала ещё одним интересным свойством. Под действием слабого внешнего магнитного поля домены перемещались в пластине ферромагнитного материала по заранее заданным направлениям с большой скоростью. Это позволило создать запоминающее устройство без механически движущихся частей — так, для чтения из памяти нужно было последовательно переместить считываемые домены к считывающему устройству. Ещё одно важное свойство пузырьковой памяти — умение сохранять свою намагниченность при отсутствии электрического тока. То есть на базе данной памяти легко получалось энергонезависимое перепрограммируемое постоянное запоминающее устройство ПЗУ, или ROM (*англ.* read only memory).

Оригинальная научная идея уже к началу 90-х гг. была полностью вытеснена более дешёвой и быстрой электронно-перепрограммируемой постоянной памятью EEPROM (*англ.* electrically erasable programmable read-only memory).

Домены (*фр.* domaine — «область») — довольно большие группы атомов, внутри которых весь металл намагничивается только в одном направлении, обычно вдоль одной из главных кристаллических осей. При намагничивании металла домены, намагниченные в направлении приложенного поля, «растут» за счёт атомов соседей. Если поле слабое, то эти изменения обратимы и невелики, напротив, в очень сильном магнитном поле некоторые домены могут принять направление поля, даже если оно не совпадает ни с одной осью кристалла. Домены по физическим размерам очень малы, их можно увидеть только с помощью мощного микроскопа, однако по сравнению с атомами эти элементарные магнитики кажутся колоссальными скоплениями.

ПРИБОР С ЗАРЯДОВОЙ СВЯЗЬЮ

Другое гениальное изобретение — CCD (*англ.* charge coupled device), или ПЗС — прибор с зарядовой связью. Этот тип полупроводника, чувствительный к свету, был придуман в Bell Labs американцами Уильямом Бойлем и Джорджем Смитом в конце 1969 г. CCD также предназначался для хранения информации в компьютере, но прибор оказался слишком медленным для использования в качестве памяти, зато для сенсора он идеален.

CCD представлял собой матрицу светочувствительных элементов, «запоминающих» поток света в виде зарядов при экспонировании. После этого прибор с зарядовой связью осуществлял передачу информации о картинке: за один шаг заряды сдвигались по рядам матрицы вниз на одну строку. Нижняя строка попадала в так называемый сдвиговый регистр, «покинув» который, сигнал усиливался и преобразовывался в цифровую форму. Процесс прекращался только тогда, когда вся матрица CCD была полностью прочитана.

Прибор с зарядовой связью назван так из-за способа передачи заряда между элементами матрицы.

CCD нашёл своё применение в 1975 г. — в видеокамерах и сканерах, в 1980 г. — в первых цифровых фотоаппаратах.

КМОП-полупроводники, основные компоненты микросхем, также восприимчивы к свету и используются в качестве светочувствительных матриц в камерах. В начале 90-х гг. существенно повысилась чувствительность КМОП-сенсоров и уменьшилась потребляемая ими мощность.

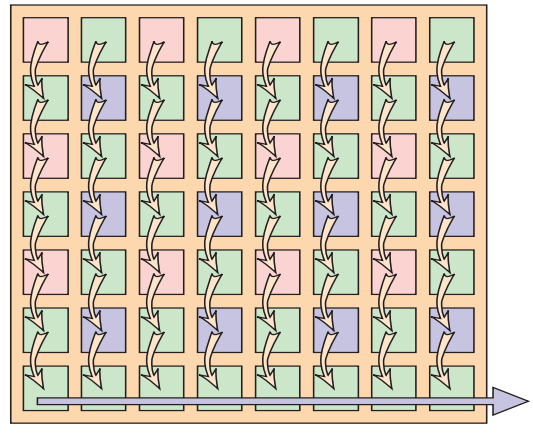


Схема работы ПЗС (CCD).

▶ Лентопротяжный механизм.

Считывание заряда с активных элементов КМОП-сенсора производится по параллельной схеме, что позволяет организовать произвольный доступ, считывая не всю матрицу целиком, а лишь некоторые области. Скорость считывания матрицы более высокая, чем в CCD, где весь заряд выходит через единственный сдвиговый регистр. КМОП-сенсор способен также уменьшить размер изображения. По экономическим соображениям становится нерентабельным содержать отдельное производство CCD-схем, когда можно организовать выпуск аналогичных КМОП-сенсоров на обычном заводе, где изготовляют КМОП-микропроцессоры.

Вполне вероятно, что в начале XXI в. на смену CCD-матрицам придут КМОП-матрицы и технология CCD будет забыта. Не потому ли CCD и пузырьковая память, в отличие от транзистора, не вошли в список десяти великих изобретений Bell Labs, которые «изменили мир»?

Первое устройство несъёмного магнитного носителя состояло из двух дисков, по 30 Мбайт в каждом. Суммарная ёмкость памяти обозначалась 30/30, что совпало с названием популярной винтовки фирмы «Винчестер». Считается, что именно так получила свое название важная деталь компьютера.

ВНЕШНИЕ ЗАПОМИНАЮЩИЕ УСТРОЙСТВА

Оперативная память компьютера должна работать быстро, но даже в XXI в. это по-прежнему достаточно дорого. Теоретически можно предположить ситуацию, когда стоимость оперативной памяти упадёт настолько, что она останется единственной памятью в ЭВМ. Однако пока она не вмещает громадных объёмов информации. Кроме того, содержимое оперативной памяти всё ещё теряется при выключении ЭВМ. Это доказывает, что внешняя память компьютера должна быть дешевле и иметь больший объём.

В одну группу внешней памяти попадают так называемые устройства с *произвольным доступом*, в другую — устройства с *последовательным доступом*. Типичный пример сейчас уже редко используемых устройств с последовательным доступом — магнитная лента. Данные можно читать или записывать только последовательно, при нарушении порядка приходится долго ждать, пока лента будет перемотана на нужное место.

Магнитные ленты — достаточно медленные устройства с последовательным доступом, хотя и с большой ёмкостью. Современные устройства для работы с магнитными лентами — *стримеры* имеют увеличенную скорость записи (несколько мегабайтов в минуту). Дешевизна магнитных лент позволяет использовать стримеры для организации долгосрочного хранения информации. Ёмкость одной кассеты стримера измеряется сотнями и тысячами мегабайтов.

Устройства с произвольным доступом позволяют любую часть данных получить примерно за одно и то же *время доступа*. Наиболее популярные устройства с произвольным доступом — *дискеты*, маленькие магнит-

ные диски, упакованные в пластиковый «конверт» размером 3,5 дюйма (1 дюйм = 2,54 см). Они защищены от механических повреждений и свободно помещаются в карман пиджака, поэтому их легко носить с собой. Данные хранятся на концентрических магнитных дорожках, расположенных на поверхности диска с двух сторон. Устройство, которое читает и пишет на дискеты, называется *дисководом для гибких магнитных дисков*. Внутри него находятся две магнитные головки, перемещающиеся радиально, от границы диска к центру шаговым двигателем. Сама дискета вращается с небольшой скоростью. В итоге чтение или запись можно произвести в любом месте дискеты, подождав не более продолжительности одного оборота.

Максимальная ёмкость дискеты — 2,88 Мбайт. Скорость передачи данных при работе с дискетой невелика, примерно несколько килобайтов в секунду, среднее время доступа — 250 мс.

До 3,5-дюймовых дискет существовали 5,25- и даже 8-дюймовые дискеты. В отличие от маленькой дискеты её большие собратья заклеивались в мягкий картонный конверт. 8-дюймовая дискета вмещала всего 0,3 Мбайт информации, а 5-дюймовая дискета до 1,2 Мбайт. В конце 80-х гг. выпускались очень похожие на 3,5-дюймовые дискеты размером ровно 3 дюйма (в пластиковом кожухе). Однако объём помещавшейся на них информации был меньше, и дискеты быстро вышли из моды.

Съёмные магнитные диски применялись на больших ЭВМ давно. Переход к несъёмному магнитному носителю — *винчестеру* (диск конструктивно объединён в едином блоке с приводом) — дал возможность существенно увеличить скорость вращения диска (до нескольких тысяч оборотов в минуту) и плотность записи. В компьютерах 2000 г. ёмкость винчестеров измерялась десятками гигабайтов, скорость передачи данных составляла сотни мегабайтов в секунду, а среднее время доступа — миллисекунды. В компьютерах начала XXI в. это основной вид внешней памяти.

МАГНИТНАЯ ЗАПИСЬ

Вальдемар Поульсен.

Магнитная запись существует уже более 100 лет, тем не менее она остаётся одной из самых передовых технологий современного общества.

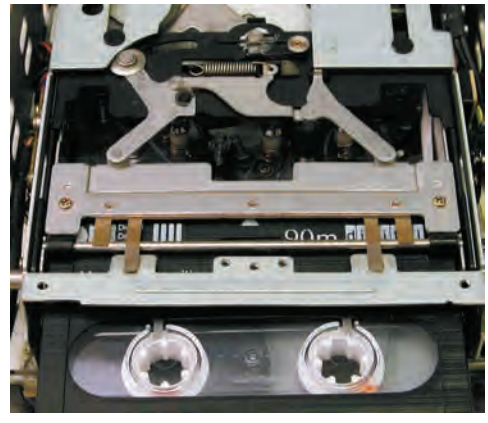
Первый прибор для записи был создан в 1898 г. датским инженером и изобретателем Вальдемаром Поульсеном (1869—1942), работавшим в то время в Датской телефонной компании.

Поульсен использовал стальную рояльную струну, протянутую через лабораторию. Запись и считывание осуществлялись с помощью винтового механизма, который перемещал вдоль струны каретку с микрофоном и тонким стержнем из специального магнитно-мягкого материала (практически не обладающего остаточным намагничиванием). Обмотка превращала стержень в соленоид (от *греч.* «солен»— «трубка» и «эйдос» — «вид») — возбудитель магнитного поля. При подаче напряжения к соленоиду на конце стержня формировалось магнитное поле, меняющееся в полном соответствии с колебаниями тока в обмотке. Это магнитное поле и создавало остаточные магнитики — элементы записи в стальной магнитно-жёсткой проволоке (сохраняющей остаточную намагниченность) — пропорционально напряжённости магнитного поля соленоида.

При воспроизведении записи применялся тот же самый стержень, работающий как магнитопровод, воспринимая изменение записанного магнитного поля при перемещении вдоль проволоки. В обмотке возникал электрический ток, который потом преобразовывался в звуковые колебания.

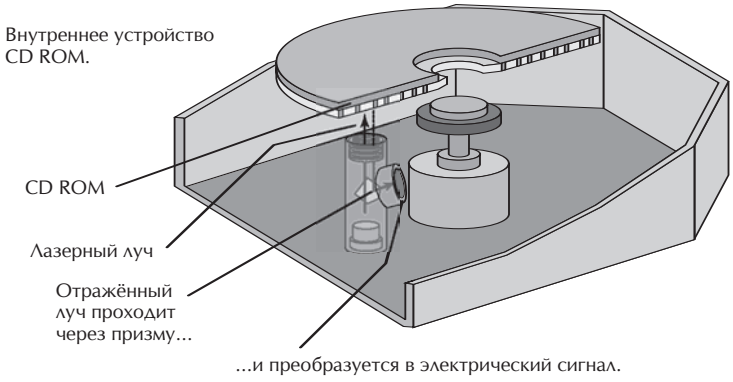
Поульсен первым создал универсальную стержневую магнитную головку. Сейчас она используется в устройствах магнитной записи в компьютерах.

Магнитная лента, лента на ацетатной основе с порошковым ферромагнитным материалом (гамма-окись железа), была разработана в 30-х гг. XX в. на фирме BASF.



Стример.

CD был разработан фирмой Philips для записи музыки. Полная ёмкость диска — около 650 Мбайт, а длительность проигрывания — 74 мин. Для 16-битного кодирования с частотой 44 100 Гц стерео требовалась скорость около 170 кбайт/с.



Оптические диски получили распространение вслед за бытовыми оптическими аудиокомпакт-дисками (лазерными дисками, или CD). Устройство, умеющее читать CD на компьютере, называется CD ROM. Оно очень похоже на проигрыватель компакт-дисков и механически аналогично устройству дисковода для гибких дисков, однако в нём единственная спиральная дорожка с информацией (как на виниловой пластинке), а при чтении используется инфракрасный лазер с длиной волны 780 нм. CD ROM позволяет также воспроизводить на компьютере аудиокомпакт-диск. Компакт-диск изготовляется из прозрачного полимера (например, поликарбоната) диаметром 120 мм и толщиной 1,2 мм.

CD ROM хранит данные в форме микроскопических углублений на отражающей поверхности, обозначающих двоичные нули и единицы, расположенные на спиральной дорожке с шагом в 1,6 мкм. Длина углубления — 0,13 мкм, ширина — 0,5—0,8 мкм, а минимальная длина углубления вдоль канавки — 0,834 мм. Скорость передачи данных составляет несколько мегабайтов в секунду и более, а среднее время доступа — десятки миллисекунд.

Достоинства CD ROM заключаются в приемлемой стоимости (меньше 1 доллара США) и большом объёме записанной информации (600—700 Мбайт). Практически всё современное программное обеспечение распространяется на CD. На них записывают большие мультимедийные обучающие энциклопедии, игры, видеофильмы и т. п.

На дисках CD RW (*англ.* ReWritable) можно многократно повторять процесс стирания и записи. Их отличает ёмкость в 700 Мбайт и более высокая надёжность хранения.

Многие персональные компьютеры, выпущенные в XXI в., уже не имеют дисковода для дискет. Магнитооптические диски, магнитные съёмные карты, DVD ROM, flash-память на карточках (ёмкость в сотни мега-

Flash-карты памяти (слева); дискеты 8; 5 и 3 дюйма (справа).

байтов) — вот неполный список устройств, используемых в начале XXI столетия в компьютере.

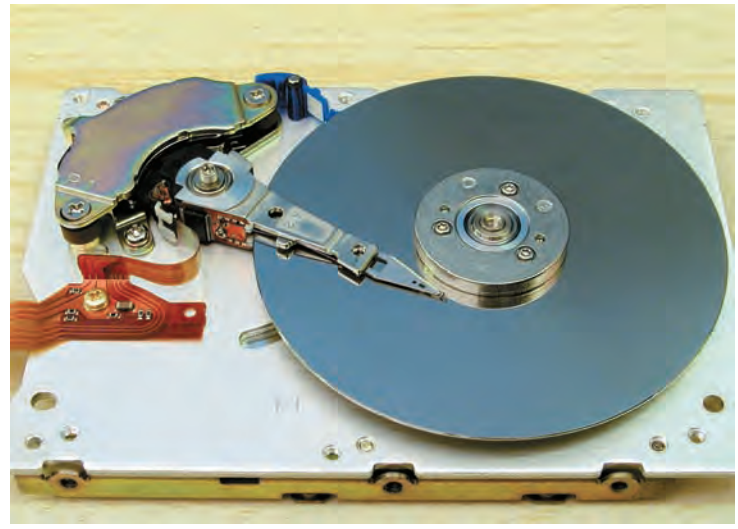
ВОЙНА СТАНДАРТОВ

В середине 80-х гг. один из специалистов фирмы JVC заметил, что если частота выхода новых стандартов VHS-кассет будет нарастать с той же скоростью, то к 2005 г. стандарты станут выпускать каждые пять минут. Однако кассеты, записанные на самом современном видеомагнитофоне стандарта VHS, могут быть просмотрены на старом видеомагнитофоне, не поддерживающем и доброй половины новомодных функций. Человеку надо бояться только того, что VHS-кассеты заменит какое-нибудь новое оборудование, например DVD — цифровой многофункциональный диск.

Нечто подобное происходило до XX в. с компьютерами. Пользователь должен был позаботиться о своевременном переносе информации со старых носителей на новые, например с 5-дюймовых дискет на 3-дюймовые, с 3-дюймовых — на CD ROM и т. д.

Однако всё изменилось с появлением DVD — носителей информации, имеющих те же размеры, что и CD ROM, но обладающие в зависимости от формата поистине гигантской ёмкостью (от 4,7 до 17 Гбайт). В DVD для достижения высокой плотности записи применяется красный лазер с длиной волны 635—650 нм, минимальная длина углубления составляет всего 0,4 мкм, а дорожка имеет шаг 0,74 мкм. DVD бывают двухслойные и двухсторонние.

В технологии DVD есть одно явно уязвимое место — различные, как правило несовместимые между собой стандарты приводов и дисков DVD. С начала 1999 г. фирмы Hitachi, Toshiba и Panasonic выпускали так называемый записываемый DVD-ROM, плохо совместимый с обычными, компьютерными DVD-ROM-приводами и бытовыми проигрывателями DVD-дисков. Формат однократно и многократно записываемых дисков DVD+RW поддерживали Hewlett-Packard, Ricoh, Philips и Yamaha. Записанные диски подчас воспроизводились на DVD-проигрывателях. Третий, но не последний, вариант предлагала технология DVD-RW, разработанная фирмой Pioneer. В записывающих устройствах можно использовать диски с однократной записью, которые легко воспроизводить практически на любом DVD-плеере фирмы Pioner и часто на DVD-проигрывателях других фирм.



Винчестер со снятой крышкой.

Как же разобраться пользователю в таком многообразии стандартов,

КЭШ НАМ ПОМОЖЕТ

Быстрее всего процессор записывает или читает данные из своей сверхбыстрой памяти (регистров). Далее, в порядке убывания скорости доступа, идут основная память и внешняя память. Если бы вся память была сверхбыстрой, то значительно повысилась бы производительность компьютера. Но реализовать это очень дорого. На помощь приходит кэш-память — быстрая по сравнению с оперативной память небольшого объёма.

Когда процессор записывает в основную память байт, то он и его адрес в основной памяти одновременно заносятся и в кэш-память. Всякий раз, когда процессор намерен прочитать некоторый байт, сначала проводится анализ: есть ли байт с этим адресом в быстрой кэш-памяти? Если есть, то происходит чтение из кэш-памяти. Если нет, то байт копируется из основной в кэш-память и одновременно передаётся процессору.

В результате в каждый момент времени наиболее часто используемая часть основной памяти уже оказывается скопированной в кэш-память и становится быстродоступной. Принцип кэш-памяти используется в винчестере, CD ROM и др.

В 1976 г. компания JVC (Victor Company of Japan) разработала формат видеозаписи VHS (Video Home System) и кассеты VHS. В 1978 г. был выпущен первый портативный VHS-видеомагнитофон.

что выбрать? Может, стоит подождать, когда один из них всё-таки станет лидером? А пока, придя домой и вставив диск в DVD-плеер или компьютер, ещё можно обнаружить, что кино не показывается, а диск не читается.

«БЛЕСК И НИЩЕТА» IBM

ГОЛУБОЙ ГИГАНТ ПРИХОДИТ НА РЫНОК

В начале 80-х гг. XX в. всё громче заявляла о себе новая отрасль компьютеростроения — персональные ЭВМ. В 1981 г. сбыт составил около 700 тыс. персональных ЭВМ, планы на следующий год прогнозировали увеличение продаж как минимум в два раза. Эту часть бизнеса не упустила и крупнейшая электронная корпорация США IBM, лидер в производстве больших ЭВМ.

12 августа 1981 г. IBM представила свой первый персональный компьютер, названный IBM PC (Personal Computer). Он использовал процессор Intel 8088 (8-битную версию процессора Intel 8086), память объёмом 64 кбайт, дисковод (или даже два) для гибких дисков ёмкостью 160 кбайт. В ПЗУ компьютера размещался интерпретатор — язык BASIC. Одновременно были выпущены программные и технические спецификации, которые позволили разработчикам других фирм создавать платы для расширения функций компьютера.

Компьютер Mainframe IBM. 80-е гг. XX в.

В том же году число проданных машин IBM PC составило 4 % от общего объёма сбыта. Лидерами на рынке персональных ЭВМ являлись:

- Radio Shack — 26 %;
- Apple — 25 %;
- Commodore — 10 %;
- Nippon Electric —5 %.

В 1983 г. IBM произвела новую модель — PC XT (eXtended Technology). Машина уже была оснащена винчестером ёмкостью 10 Мбайт, тремя разъёмами под дополнительные платы (слоты расширения). Оперативная память занимала 128 кбайт, но могла расширяться до 640 кбайт. На дискеты записывалось 360 кбайт информации. В качестве операционной системы использовалась диалоговая система MS DOS компании Microsoft.

Компьютеры IBM по сравнению с машинами фирмы Apple проигрывали по всем показателям. Liza рядом с IBM PC выглядел как гостья из будущего. А выпущенный Macintosh, использующий 32-битный процессор Motorola 68000 с тактовой частотой 8 МГц, оснащённый оконной операционной системой, затмил IBM.

К основным причинам, которые помогли IBM в борьбе за рынок персональных компьютеров, можно отнести колоссальные финансовые ресурсы корпорации, позволяющие производить заведомо проигрышный компьютер в течение долгого времени, и открытость архитектуры IBM PC.

◄ Винчестер.

Это привело к выпуску в 1982 г. первого клона IBM PC — Compaq I, совместимого с IBM PC. В дальнейшем подобные машины стали называть IBM PC-compatible (совместимыми), а IBM продолжала выпускать далеко не самые лучшие компьютеры в ряду IBM-compatible.

Часто причиной выживания IBM PC называют процессоры Intel.

INTEL. «ДОВЕДЁМ — ДОДЕЛАЕМ»

Как правило, когда новую архитектуру создаёт небольшая группа специалистов, то отдельные части очень хорошо подогнаны друг к другу, чего нельзя сказать о линейке процессоров 80×86. Это продукт нескольких независимых групп разработчиков, развивавших данную архитектуру десятки лет, добавляя всё новые и новые возможности к исходному набору команд.

В 1978 г. был представлен процессор Intel 8086 как расширение своего 8-битного собрата — микропроцессора 8080. Процессор стал 16-битным, построенным на базе одного аккумулятора, но дополненный новыми регистрами. Практически каждый регистр этой архитектуры имел чётко определённое назначение; Intel 8086 по удобству программирования проигрывал классу более прогрессивных процессоров с регистрами общего назначения. Версия 8088 (с 8-битной внешней шиной данных) — основа серии компьютеров IBM PC, завоевавшая впоследствии весь мир. В 1980 г. выпустили сопроцессор плавающей точки 8087, что добавило к архитектуре 8086 несколько десятков команд для работы с действительными числами.

Появление в 1982 г. микропроцессора 80286 определило дальнейшее расширение старой архитектуры 8086. Адресное пространство расширили до 24 разрядов (16 Мбайт), увеличили количество команд. Для совместимости с 8086 в 80286 был предусмотрен так называемый режим реальных адресов.

В 1987 г. создан микропроцессор 80386. В дополнение к 32-битной архитектуре с 32-битными регистрами и 32-битным адресным пространством в микропроцессоре 80386 появились поддержка страничной организации памяти и ряд дополнительных команд. Наконец-то бесконечные доработки практически превратили 80386 в процессор с регистрами общего назначения. Как и все предыдущие модели, он имел режим выполнения программ, написанных для 8086. Страничная организация памяти и 32-разрядная архитектура позволили установить на этой платформе полноценную операционную систему UNIX. Последующие процессоры: 80486 (1989 г.) и Pentium (1993 г.) были нацелены больше на увеличение производительности и добавили лишь несколько новых команд для обеспечения многопроцессорной работы.

Тяжёлое наследие в виде совместимости с ранними версиями 80x86 процессоров, как ни странно, не отразилось на распространённости процессоров Intel. Четыре из пяти процессоров персональных компьютеров базируются именно на этой архитектуре.

Процессоры, начиная с i486, имели некоторые черты RISC-процессоров. Например, наиболее употребительные команды выполнялись за один такт. Процессор Pentium ознаменовал новый этап в развитии 80x86, связанный с использованием элементов RISC-архитектуры. В отличие от процессоров i486 в нём использовалась раздельная кэш-память команд и данных ёмкостью по 8 кбайт, что повышало производительность. Основные

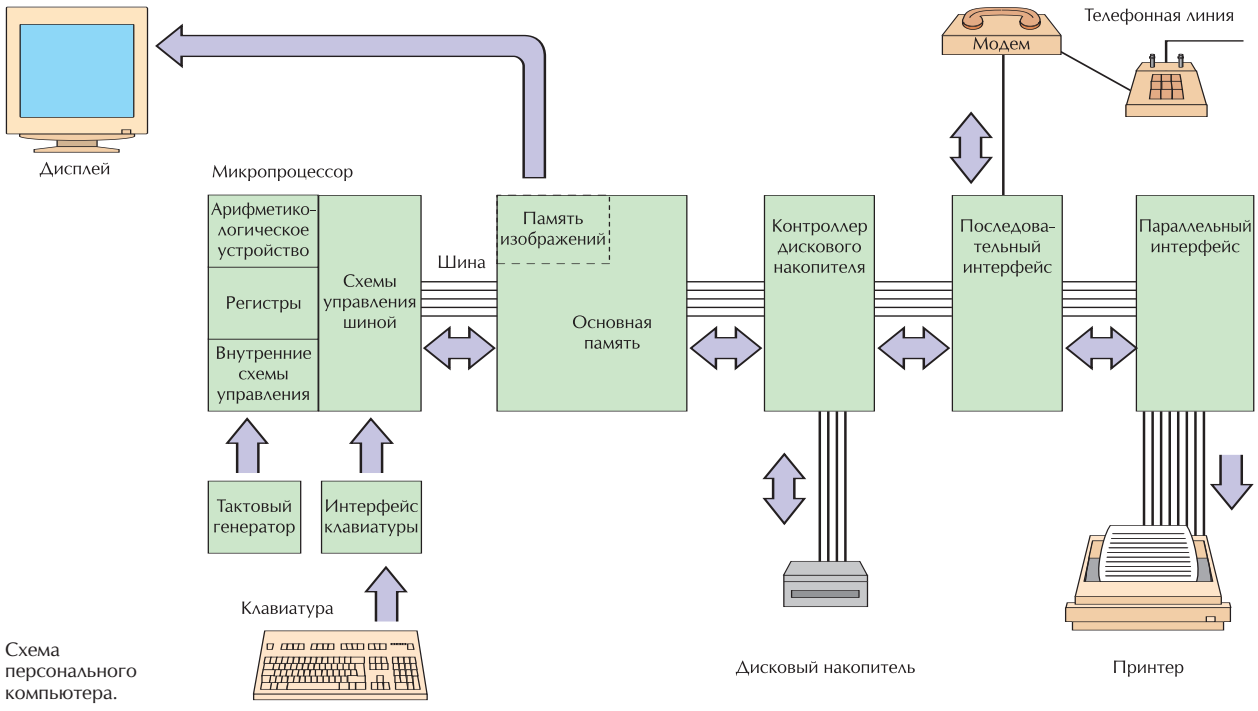


Схема персонального компьютера.

Процессоры i486SX, i486DX, i486DX2 и i486DX4 — это 32-битные процессоры с внутренней кэш-памятью ёмкостью 8 кбайт и 32-битной шиной данных. В процессоре i486SX отсутствует сопроцессор плавающей точки. В процессорах с маркой OverDrive и i486DX2 внутренняя тактовая частота удваивается, а в процессоре i486DX4 — утраивается, что увеличивает производительность процессора почти в два раза.

MMX-инструкции могли выполнять простейшие команды с 8-битной целочисленной матрицей размером 8 x 8. Это ускоряет выполнение дискретного косинусного преобразования, используемого при воспроизведении MPEG-кинофильмов, музыки в формате MP-3 и просмотра картинок, сжатых алгоритмами JPEG.

команды процессора распределялись по двум независимым исполнительным устройствам — конвейерам. Первый конвейер выполнял любые команды, а второй — простые целочисленные команды и некоторые команды с плавающей запятой. Команды направлялись на конвейеры одновременно, более сложная команда поступала в первый конвейер, а менее сложная — во второй. К сожалению, такая работа была возможна на ограниченном множестве целочисленных команд.

Нельзя утверждать, что от технологии «доведём — доделаем» фирма Intel отказалась в 80-х гг. XX в. В 90-х гг. вышла модификация процессоров Pentium — MMX (*англ.* Multi-Media eXtension) — мультимедийное расширение, добавившее несколько десятков новых команд с восемью 64-разрядными регистрами. Для доступа к MMX-регистрам использовался тот же механизм, что и для доступа к регистрам при выполнении команд с плавающей запятой, что исключало одновременное выполнение команд.

Справедливости ради можно заметить, что в следующем процессоре — Pentium II подобную оплошность исправили. Также была исправлена ошибка в процессоре Pentium, которая в середине 90-х гг. имела настолько большой общественный резонанс, что фирма, устранив её, все вновь выпущенные процессоры помечала надписью «Bug free» — «Ошибок нет».

АТАКА КЛОНОВ. ГДЕ IBM?

Открытость компьютера IBM PC привела к тому, что вместе с ростом объёма продаж мир наполнился многочисленными аналогами третьих производителей. В 1984 г. их выпускали более 50 компаний, а через два года объём продаж клонов превысил объём собственных поставок IBM. После этого IBM так больше и не смогла вернуть

себе лидерство. С тех пор название «IBM PC» стало нарицательным, так же как «ксерокс».

Вероятно, большой ошибкой можно считать и производство в 1984 г. IBM PCjr — домашнего персонального компьютера. Машина имела улучшенную графику (по сравнению с IBM PC), звуковую плату и беспроводную клавиатуру. Клавиатура, работавшая из рук вон плохо, сыграла злую шутку с домашней машиной. Улучшение (добавление новых плат) PCjr не оправдалось, и компьютер был снят с производства в 1986 г.

В 1984 г. компания IBM объявила о выпуске новой серии персональных компьютеров — IBM PC/AT. Первые машины работали на частоте 6 МГц (однако появились процессоры на частоте 12 и даже 20 МГц). Intel 80286 стали использовать в AT-клонах. Существует легенда, что BIOS (место в ROM, где содержатся основные функции устройств) в те годы поставляла главным образом фирма IBM, его коды держались в секрете, и он просто устанавливался на каждом компьютере без переделки. Чтобы в клонах не применяли процессоры на более высокой тактовой частоте, чем настоящие IBM PC/AT, при старте проверялась скорость процессора: если она была выше положенных 6 МГц, то компьютер останавливался. Как рассказывают, для этого на клонах появилась кнопка «Turbo» («Ускоритель»). При старте 12-мегагерцевый процессор запускался на 6 МГц, проходил проверку, затем человек нажимал кнопку, и процессор работал в два раза быстрее. Кнопка «Turbo» исчезла с корпусов после появления процессоров Pentium, хотя BIOS давно уже выпускали другие фирмы и, конечно же, не проводили никаких проверок.

Первый компьютер на процессоре 80386 был изготовлен компанией Compaq Computers в 1987 г. В апреле того же года IBM попыталась вернуть себе лидерство и создала семейство персональных компьютеров PS/2 (Personal System). Эти компьютеры использовали другой стандарт шины и соединений, поэтому их производителей ждала неудача. Девять компаний-клонмейкеров во главе с Compaq (AST, Epson, Hewlett- Packard, NEC, Olivetti, Tandy, Wyse и Zenith) выпустили свой стандарт шины, полностью совместимой с предшественницей в IBM PC/AT. Не отставали и производители микропроцессоров. Advanced MicroDevices (AMD) и Cyrix разработали процессоры, совместимые с i386 и i 486, с улучшенными скоростными показателями при более низкой цене.

Клон IBM. PC XT.

Разочарование пережила фирма IBM при выпуске операционных систем. Полностью 32-разрядная многозадачная операционная система OS/2 Warp 3.0 вышла в 1994 г., ещё до системы Windows 95 фирмы Microsoft. Её презентация прошла весьма торжественно. Тем не менее ниша уже была занята Windows 3.11, привычной для пользователей, которые

Компьютер PS/2.

К одной из самых досадных неудач фирмы Intel относят выпуск в конце 1982 г. iAPX 432 — набора из четырёх микросхем с общей производительностью 2 млн команд в секунду. Это был объектно-ориентированный процессор, включавший поддержку многих функций операционных систем, таких, как планировка процессов и обмен сообщениями. iAPX 432 имел производительность в 10 раз меньшую, чем Motorola 68000 и Intel 80286.

с нетерпением ожидали появления Windows 95. IBM попала в затруднительное положение. С одной стороны, OS/2 создавалась под будущие персональные ЭВМ на базе своих процессоров Power PC, с другой — IBM сотрудничала с Microsoft для поддержания собственной линейки компьютеров IBM PC. OS/2 — классический пример, как замечательная идея умирает, натыкаясь на неудачную внутреннюю политику фирмы и бизнес-стратегию.

А где сейчас IBM? Голубой гигант по-прежнему остаётся суперкорпорацией и неплохо себя чувствует, шагнув в XXI век.

О ЧЁМ НЕ ЗНАЛ СТИВ ДЖОБС

Стивен Пол Джобс родился в феврале 1956 г. в Маунтейн-Вью (штат Калифорния, США), районе, впоследствии ставшем сердцем Силиконовой долины (центр микрокомпьютерной промышленности). Здесь же спустя несколько лет Роберт Нойс создал свою первую интегральную микросхему.

Первооткрыватели часто действуют одержимо, словно наделены божественным правом вершить дела. Для достижения цели они не жалеют ни себя, ни коллег, ни подчинённых, ни начальников. Они обладают огромной силой воли. Всё сказанное можно отнести к Стиву Джобсу, одному из основателей компьютерной фирмы Apple. Его девиз: «Это мой путь — и это лучший путь».

Стефан Возняк фактически был изобретателем первого персонального компьютера Apple. В его производство Стив Джобс вложил средства, полученные от продажи принадлежавшему ему автобуса «фольксваген», а Возняк для того же продал за 1,3 тыс. долларов свой калькулятор.

Фирма Apple была создана 1 апреля 1976 г. Президент фирмы Майк Скотт считал главным поиск средств на производство компьютеров. Он составлял бизнес-планы, проводил маркетинговые исследования. Однако финансовое положение компании долго оставалось тяжёлым. Но в 1980 г. было продано уже более 130 тыс. экземпляров Apple-II. (В 1980 г. акционерный доход Джобса составил 256 млн долларов.)

Стивен Джобс и Стефан Возняк.

Стефан Возняк занимался разработкой компьютеров, а Джобс выступал как катализатор, который не просто дал толчок к возникновению нового рынка, но и организовал это рискованное дело. Он сам определял для себя правила и сам же нарушал их, совершенно не считаясь с людьми. Во имя избранной цели Стивен готов был пожертвовать всем и не терпел другого отношения к делу ни от кого.

Но не все решения Джобса приносили прибыль. Он выступал против создания совместимых компьютеров. Apple-II, Lisa, Macintosh и Apple-III были несовместимы даже между собой. Но странно, совершив такую крупную ошибку, компания приобрела целую армию фанатов — пользователей машин фирмы Apple. Удобная техника с самого начала стала доминировать на рынке учебных компьютеров.

Стивена Джобса в 1981 г. назначили руководителем компании. Итогом 1982 г. явилось вручение премии «Человек года»... персональному компьютеру. (Доход Стива тогда составил около 500 млн долларов.)

В 1983 г. фирма Apple выпустила новую машину — Lisa. Это первый компьютер, использующий оконную систему GUI (graphic user interface). Другой новацией явилась иерархическая файловая система с директориями и поддиректориями и выпадающими меню, к которой сейчас так привыкли.

Однако высокая стоимость (10 тыс. долларов) и небольшая производительность привели к тому, что машина была вытеснена с рынка более успешной разработкой Apple — компьютером Macintosh. Lisa считается одной из самых больших производственных ошибок фирмы Apple, так как за три года производства компьютер не принёс компании финансовой выгоды.

В 1983 г. Apple возглавил бывший исполнительный директор компании PepsiCo Джон Скалли, а через два года Джобса отстранили от управления.

1984 год — время триумфа. Компания подготовила для рынка новый продукт — компьютер Macintosh. Чтобы победить IBM, требовалось правильно провести маркетинговую подготовку.

Был сделан новый шаг для рекламодателей. Прямо в середине трансляции Суперкубка по американскому футболу показали рекламный ролик с нехитрым на первый взгляд сюжетом. Зал. Люди. Взгляды прикованы к экрану, с которого вещает «Большой Брат», олицетворяющий фирму IBM. Всё в унылых тонах. Внезапно вбегает девушка и кидает в экран молот. Экран разбивается вдребезги, и появляется надпись: «В этом году фирма Apple выпускает компьютер Macintosh. Вот почему 1984 год не станет “1984”!» . Этот клип, обыгрывавший темы знаменитого романа-антиутопии Джорджа Оруэлла «1984», признан лучшим за всю историю рекламы XX века. Macintosh стал продаваться «на ура».

Лишь в 1988 г. компания Apple значительно отстала от других фирм, производящих компьютеры.

Вот итоги продаж продукции за год (в миллиардах долларов США):

- IBM — 59,681;
- Digital Equipment — 11,475;
- Hewlett-Packard — 9,831;
- Apple Computer — 4,071.

Но к 90-м гг. XX в. компания уже обошла IBM по объёмам продаж персональных компьютеров (к тому времени IBM практически ушла с этого рынка). К 1992 г. ежегодный доход компании достиг 7 млрд долларов, хотя Apple охватывала лишь 10 % мирового рынка персональных компьютеров.

Дизайн первых моделей выполнялся в стиле all-in-one, т. е. системный блок и 9-дюймовый монитор в одном корпусе весом около 9 кг. Небольшой размер экрана компенсировался чрезвычайно высокой чёткостью изображения. В качестве центрального процессора использовался Motorola 68000, а затем и более мощный процессор Motorola 68020.

Тот самый Macintosh. Теперь он называется Classic.

История 90-х гг. для Apple — история убытков и несбывшихся надежд. В 1991 г. компания уволила 10 % сотрудников, в 1993 г. Джон Скалли покинул Apple. Фактически для компании закончился очередной этап истории, возможно самый лучший. Через два года компания терпит серьёзные убытки — около 70 млн долларов. Через год, несмотря на ротацию в руководстве, убытки за один квартал достигли 700 млн долларов.

В 1993 г. Apple выпустила один из первых PDA (*англ.* personal digital assistant — «персональный цифровой помощник») — Apple Newton. Основной рекламный слоган: «Newton может читать ваши рукописные тексты!». Эта возможность PDA стала притчей во языцех. Newton плохо распознавал рукописный текст, возникало немало забавных, смешных ошибок. Тем не менее Newton проложил дорогу более совершенным программам распознавания текста.

Один из своих принципов Джобс обозначил аббревиатурой KISS (Keep It Simple, Stupid! — «Делай это проще, дурачок!»).

MACINTOSH НА РУБЕЖЕ ВЕКОВ

Компьютер, неизвестный для большинства российских пользователей, является сильным помощником для узкого круга профессионалов. Macintosh сочетает в себе, казалось бы, несовместимое: надёжность и высокую эффективность, с одной стороны, и подкупающую простоту в общении — с другой. Компьютеры, на которых красуется логотип, изображающий надкусанное яблоко, можно часто увидеть в домах и учебных организациях Америки. Среди этих машин есть переносные «чемоданчики» — PowerBook и оригинальный Cube в форме куба — и более привычные «башни» PowerMac G3 и PowerMac G4. Существуют и мощные производительные модели с двумя процессорами G4 (один ум хорошо, а два — лучше!), особенно популярные у художников и дизайнеров, занимающихся компьютерной графикой. Есть iMac и eMac, где монитор и системный блок совмещены в одном корпусе. В 2002 г. появился необычный iMac-2 — серая полусфера с жидкокристаллическим монитором на кронштейне, как у настольной лампы. Целенаправленное удешевление iMac лишь увеличило число поклонников этих «подсолнухов».

Macintosh (Мас-компьютер) изначально создан для людей творческих, по принципу «сел — и работай». Фирма Apple производит не только компьютеры Mac, но и операционную систему для них — MacOS (или System), которая ведёт себя так дружественно, что порой её присутствие даже незаметно. Программы настраиваются легко, независимо от квалификации пользователя. Новые версии программ и драйверов меняются нечасто, и именно поэтому на рынок попадает продуманный до мелочей продукт. Приложения реже конфликтуют в борьбе «не на жизнь, а на смерть» за ресурсы компьютера, отсюда меньше «зависаний». Конечно, для опытных пользователей в MacOS существует возможность покопаться и поглубже, например автоматизировать рутинную работу с помощью языка управления заданиями AppleScript. Этот язык описывает *скрипты* (интерпретируемые программы), управляющие действиями компьютера, которые задаёт пользователь.

Операционная система MacOS X основана на ядре хорошо себя зарекомендовавшей ОС FreeBSD (свободно распространяемый UNIX), отлично использует многопроцессорность Macintosh. Она более устойчива, чем её младшие сёстры: имеет защиту памяти — при «зависании» какой-либо из программ не нужно перезагружать компьютер, достаточно выгрузить эту программу из памяти и работать дальше. Новый графический интерфейс Aqua (*лат.* «вода») имеет ряд приятных усовершенствований, как внешних, так и для более удобной работы.

Считается, что компьютеры Macintosh не могут решать столь широкий круг задач, как PC-компьютеры. Это заблуждение. Даже производители компьютерных игр не оставляют Macintosh без работы: практически все игрушки можно найти и для этих машин. Да и в Сети «соперники» отлично чувствуют себя вместе. Конечно, Macintosh несовместим с PC, но если запустить специальную программу-эмулятор PC — Virtual PC, то станут доступны все приложения и игры, написанные для PC. В последнее время (начиная с модели компьютера Power Mac G4) к Macintosh можно подключать многие устройства, созданные изначально для PC: оперативную память и др.

Когда-то Стив Джобс выдвинул лозунг Apple: «Думай иначе!». Детище компании — Macintosh, постоянно радующий пользователей своей нестандартностью, — лучшая его реализация.

В 1985 г. Стив Джобс, уйдя из Apple, основал компанию NeXT Computer. Джобс хотел создать совершенно новый компьютер, он был полон идей. Во исполнение цели NeXT взял всё лучшее из разработок других компаний: операционную систему UNIX, объектно-ориентированный язык Objective-C и новую систему визуализации, основанную на Adobe PostScript.

В течение пяти лет компания выпускала компьютеры собственной разработки, но уже в 1993 г. их производство прекратилось. Операционные системы по-прежнему реализовывались: сначала появилась NextStep для процессоров Intel, затем она была портирована на Hewlett-Packard PA-Risc и рабочие станции Sun SPARC.

В период работы в NeXT Computer Стив Джобс показал себя профессиональным финансистом. Он привлёк инвесторов личным примером, вложив 15 млн долларов в свою новую компанию, заключил 20-миллионную сделку с миллиардером Россом Перо, получил инвестиции от японского концерна Cannon на сумму 100 млн долларов. Содружество с IBM принесло компании десятки миллионов долларов. Джобс провёл классическую сделку с магазинами Businessland Computer, реализовав будущую продукцию NeXT на сумму 100 млн долларов.

В сентябре 1997 г., после 12-летнего перерыва, Стивен Джобс снова возглавил компанию Apple. Но это был уже не прежний Джобс. Многие его решения вызвали протест фанатов компьютеров Apple. Принцип Джобса — простота и прямота ведут к успеху — воплотился в новой стратегии компании. Сократилась линейка выпускаемой техники, а основные усилия в области программного обеспечения сфокусировались на современной операционной системе.

Стив Джобс.

Результатом деятельности вернувшегося президента стал рост объёма продаж. Похоже, Apple вышел из затяжного кризиса. Правда, не обошлось без финансовой помощи от Microsoft. К сожалению, мир устроен так, что научный и финансовый успех редко идут рука об руку. Стив Джобс так и не осознал, что все его гениальные идеи и начинания, словно по воле рока, изначально обречены. А если так, то Apple снова произведёт фурор своими компьютерами и операционными системами, за которым последует неминуемый финансовый крах.

Вы, скорее всего, не фанат Microsoft, если на вашей любимой футболке написано «WINDOWS MUST DIE!».

УИЛЬЯМ (БИЛЛ) ГЕНРИ ГЕЙТС III

Основатель и председатель совета директоров корпорации Microsoft, ведущего мирового производителя программного обеспечения для персональных компьютеров, Билл Гейтс родился 28 октября 1955 г. в Сиэтле. В семье он был третьим ребёнком. Его полное имя — Уильям Генри Гейтс III. (Билл Гейтс I — дед, владелец мебельного магазина, Уильям Гейтс II — отец, адвокат.) Его мать, Мэри Гейтс (урождённая Максвелл), была школьной учительницей, членом правления в университете штата Вашингтон и председателем благотворительной организации United Way International, а дед со стороны матери являлся основателем банка National City.

Билл Гейтс начал своё образование в муниципальной школе, продолжив его в элитарной частной школе Lakeside School. Там с 13 лет он увлёкся программированием.

В 1973 г. Билл Гейтс поступил на первый курс Гарвардского университета, где разработал язык программирования BASIC для первого компьютера — Altair 8800. Но Гарвард Билл так и не закончил. Он покинул стены учебного заведения на третьем году обучения, решив полностью посвятить себя компании по разработке программного обеспечения для персональных компьютеров Microsoft, которую основал вместе с Полом Алленом в 1975 г., предвидя развитие PC.

По ошибке Биллу Гейтсу приписывается авторство операционной системы DOS, но на деле это не что иное, как удачное вложение денег. Придумал DOS Тим Патерсон, который в то время в Microsoft даже не работал. В 1981 г. Гейтс и Аллен, выполнявшие работу для IBM, купили за не слишком крупную сумму у Патерсона из Seattle Compute Products созданную им QDOS (Quick Disk Operating System — «быстрая дисковая операционная система») для 16-разрядных процессоров Intel. Конечно, операционная система требовала доработки, но это не изготовление с нуля абсолютно новой ОС. Учтя все требования со стороны IBM, компаньоны переделали её для 8-разрядных процессоров, которыми оснащались IBM PC. Неизвестная тогда компания Microsoft смогла заключить с IBM контракт на поставку новой операционной системы. IBM оплатила её производство и обязалась продавать компьютеры только с версией PC DOS (MS DOS). Microsoft получала проценты с каждой проданной машины. Теперь данная сделка изучается молодыми менеджерами как классическая — ведь стоило продать право на поставку за некоторую круглую сумму, и не был бы сейчас Билл Гейтс мультимиллиардером.

1 января 1994 г. Билл Гейтс женился на Мелинде Френч, работавшей в Microsoft. Свадьба проходила на Гавайях, на острове Ланаи. Чтобы избежать появления журналистов, были оплачены все гостиничные номера и зарезервированы все доступные автомобили и самолёты. Шафером стал Стив Баллмер, занявший в начале 2000 г. пост генерального директора Microsoft. 26 апреля 1996 г. у Гейтсов родилась дочь Дженнифер Кэтрин Гейтс. В сентябре 1997 г. на берегу озера Вашингтон Билл Гейтс построил для своей семьи полностью компьютеризованный дом. 23 мая 1999 г. в семье Гейтсов появился сын Роури Джон Гейтс.

В 1995 г. Гейтс вместе с Натаном Мирволдом, в то время вице-президентом Microsoft, и журналистом Питером Райнарсоном написал книгу «Дорога в будущее», опубликованную издательством Viking в США и 18 недель находившуюся в списке бестселлеров. Вторая книга — «Бизнес со скоростью мысли», изданная в 1999 г., — была написана в соавторстве с Коллинз Хемингуэй и впоследствии издана на 25 языках. Прибыль от обеих публикаций Билл Гейтс пожертвовал в благотворительный фонд, который осуществляет поддержку учителей мира, использующих компьютерные технологии в учебном процессе.

Microsoft постоянно разрабатывает и совершенствует информационные продукты и технологии. Её успех — стремление сделать работу с компьютером более простой и удобной. Компания расходует огромные средства на научно-исследовательские работы.

Однако идея использования оконных систем принадлежит не Биллу Гейтсу или кому-либо из компании Microsoft. Её уже внедрял Стив Джобс, тот самый, который вместе со Стивом Возняком создал первый персональный компьютер. Компьютеры Apple — Lisa и Mac уже были на рынке и превосходили IBM оконными системами. Игровой компьютер Atari в конце 80-х гг. оказался на порядок лучше IBM AT — по возможностям оконной операционной системы и по 32-битному процессору Motorola. Тогда в 1985 г. Гейтс выпустил Windows. Первые версии (вплоть до 3.0) были почти неработоспособны. Требовалось вводить не только оконный интерфейс, но и многозадачность — возможность выполнять сразу несколько программ. В надстройке над MS DOS Windows 3.0 это уже удавалось сделать. Открывалось сразу несколько окон и оставалось лишь переключаться между задачами. Такой механизм был встроен в DOS, только не MS (фирмы Microsoft), а DR DOS (фирмы Digital Reserch — успешного разработчика ОС для 8-битных ЭВМ (CP/M); DR DOS получился весьма удачным изделием. И дни MS DOS были бы сочтены, если бы не умение Билла Гейтса бороться с конкурентами. Дело в том, что Windows работала исключительно в MS DOS, а с другими операционными системами была несовместима.

В 1997 г. Билл Гейтс претворил в жизнь свою идею — кредитование 150 млн долларов главному конкуренту — компании Apple, которую предварительно он поставил перед угрозой банкротства и ликвидации. Что принесла эта сделка Биллу Гейтсу и его компании кроме удачной возможности избежать антимонопольного расследования и принудительного раздела компании?

Билл Гейтс.

Пол Аллен.

Тим Патерсон.

Скорее всего, вы работаете с продукцией Microsoft, если на вопрос трёхлетнего малыша: «Почему небо синее?» — начинаете бормотать что-то о Windows.

Microsoft нанесла удар по конкурирующим компаниям Netscape, Sun, Oracle, объявив, что доступ в Internet будет встроен в операционную систему Apple. Одновременно Microsoft гарантировала уничтожение как конкурента операционной системы Apple, созданной на основе NeXT (машины, как бы говорящей своим названием, что она не IBM PC XT). Не по своей воле, но Apple сыграла определённую роль в фактической ликвидации языка Java (общая основа для разработки платформонезависимых приложений, выпущенная компанией Sun).

А чего стоит заключение соглашения между Microsoft и провайдерами о снятии логотипа Netscape в обмен на включение последних в список информационных каналов продуктов Microsoft! Это ли не победа Explorer над Netscape? Не вкладывать труд и средства в доработку и победить ввиду явного превосходства (как в 1995 г. Netscape победил NSCA Mosaic, Cello и другие web-навигаторы), а просто не дать возможности сравнить два продукта.

Каждая новая версия Windows требует у несчастных пользователей серьёзного усовершенствования их компьютеров. На память приходит знаменитая шутка: «На каждый наперёд заданный высокопроизводительный компьютер найдется Word, работающий на нём медленно». Однако при этом Microsoft не обманывает покупателей: ничего им не обещает, фактически никакой поддержки; неработающая ОС обменивается (с доплатой) на новую, снова неработающую. И так до бесконечности... А ведь никому из автомобилистов не придёт в голову купить машину, которая поедет только через месяц и тут же сломается.

Борьба с пиратами, распространяющими и использующими программные продукты Microsoft, не так уж и выгодна, как может показаться на словах. Дешёвые незаконные копии, как наркотик, втягивают миллионы граждан в использование программ фирмы Microsoft. А остальная, честная часть человечества вынуждена пользоваться именно этими продуктами, чтобы читать письма и документы, получаемые из разных концов света, написанные в тех же MS Word. Как только разработки третьих фирм начинают бить Word по всем параметрам, обеспечивая пользователю возможность работы с документами, созданными в MS Word, Microsoft выпускает новую версию, которая имеет всего одно отличие от предыдущей — другой формат документа. И гонка продолжается.

Высокая эффективность программ Microsoft — это миф о суперпрограммистах, работающих в фирме. Они просто пишут программы, используя недокументированные функции системы Windows и специальные библиотеки, как бы обходясь без «программ-посредников» между их продуктами и Windows. А для всех остальных программистов это тайна за семью печатями. Поэтому характеристики почтовой программы Outlook ничуть не выше, чем у её конкурентов, просто она использует внутренние функции Windows эффективнее.

Кроме корпорации Microsoft Билл Гейтс основал компанию Corbis Corporation, занимающуюся разработкой цифрового архива произведений искусства и фотографий из государственных и частных коллекций всего мира. Также он входит в правление фармацевтической компании Icos Corporation, владеет акциями компании Darwin Molecular, занимающейся генной инженерией, — подразделением другой британской фармацевтической корпорации Chiroscience. Он вложил средства в компанию Teledesic, работающую над осуществлением проекта по запуску на низкую орбиту вокруг земного шара нескольких сотен спутников, способных обеспечить всемирную двухстороннюю высокоскоростную передачу данных.

Билл Гейтс увлекается не только компьютерными технологиями, он интересуется биотехнологиями, много читает, а также любит играть в гольф и бридж. Гейтс III — великий специалист по маркетингу. Своё состояние (более 60 млрд долларов) он собрал сам и вошёл в историю как самый молодой миллиардер.

КОМПЬЮТЕР XXI ВЕКА

КОМПЬЮТЕР И ЗДОРОВЬЕ

Компьютеры прочно заняли своё место и в доме, и в офисе. Они сделались настолько обыденными, что невольно забываются связанные с их использованием опасности. Как и любое достижение прогресса, будь то автомобиль, самолёт, телевизор, компьютер является источником негативных воздействий на человека, приносит вред его здоровью при неправильном применении, может вызвать профессиональные заболевания. Однако компьютер становится добрым, умным, неутомимым помощником, если человек продумывает все моменты общения с ним, соблюдает правила техники безопасности, размышляет над научной организацией труда, не нарушает санитарные нормы, составленные специалистами.

Исследует эти проблемы *эргономика* (от *греч.* «эргон» — «работа» и «номос» — «закон») — наука о взаимодействии человека и машины. Эргономика занимается комплексным изучением трудовой деятельности человека и поэтому объединяет многие научные дисциплины: физиологию, гигиену труда, психологию и др. Учёные стремятся найти пути снижения нагрузки на организм человека, связанной с работой на компьютере, участвуют в создании совершенной и безопасной техники. Среди их задач — орга-

низация рабочего места, профилактика вредного воздействия компьютера на человека.

ВРЕДНЫЕ ИЗЛУЧЕНИЯ ПРИ РАБОТЕ ЗА КОМПЬЮТЕРОМ

Компьютер — источник нескольких видов излучений и полей. Электронно-лучевая трубка монитора создаёт ионизирующее (рентгеновское) излучение. Однако в современных мониторах оно незначительно, так как надёжно экранируется и сравнимо с естественным радиационным фоном, в чём можно легко убедиться с помощью бытового радиометра.

Как и любой другой электроприбор, компьютер создаёт также электромагнитное излучение. Когда компьютер включён, все входящие в него устройства (монитор, системный блок, клавиатура, принтер, сканер и т. д.) и вспомогательное электрооборудование (сетевые фильтры, блоки бесперебойного питания) формируют сложное электромагнитное поле.

Большинство исследований по влиянию электромагнитного излучения свидетельствуют о его вреде для здоровья.

Кроме того, компьютер создаёт электростатическое поле. Оно возникает в результате облучения экрана монитора потоком заряженных частиц. Электростатическое поле способствует оседанию пыли и аэрозольных частиц на лице, шее, руках, что может вызвать у людей, особо чувствительных к подобному воздействию, негативные кожные реакции — сухость, аллергию.

Электростатическое поле влияет на ионный состав воздуха. На поверхности кинескопа монитора возникает положительный заряд, который нейтрализует отрицательно заряженные полезные ионы воздуха, что ухудшает среду в помещении с компьютерами.

КОМПЬЮТЕР И ЗРЕНИЕ

Уже в первые годы компьютеризации было отмечено специфическое зрительное утомление у пользователей дисплеев, получившее общее название «компьютерный зрительный синдром» (*англ.* Computer Vision Syndrome, или CVS). Причин его возникновения несколько, и прежде всего сформировавшаяся за миллионы лет эволюции зрительная система человека, которая приспособлена для восприятия объектов (природы, рисунков, печатных текстов и т. п.) в отражённом свете, а не для работы с дисплеем. Изображение на дисплее принципиально отличается от привычных глазу объектов наблюдения: оно светится, состоит из отдельных точек — пикселей, мерцает (точки с определённой частотой зажигаются и гаснут), не соответствует естественным цветам.

При длительной беспрерывной работе за компьютером у глаз не бывает необходимых фаз расслабления, они напрягаются, работоспособность снижается. Большую нагрузку орган

Наиболее чувствительными к воздействию компьютера являются центральная и сердечно-сосудистая системы. Наблюдаются нарушения условно-рефлекторной деятельности, снижение биоэлектрической активности мозга, изменения межнейронных связей. Возможны отклонения в работе эндокринной системы. Первым симптомом недомогания может стать повышенная возбудимость, затем появляется физическая и нервно-психическая слабость. Для общей клинической картины хронического воздействия электромагнитного поля характерны головная боль, утомляемость, ухудшение самочувствия, гипотония, брадикардия, изменение проводимости сердечной мышцы. Указанные явления могут быть слабо, умеренно или явно выражены. Порой отмечаются незначительные и, как правило, нестойкие изменения в крови.

Защитный экран.

зрения испытывает при вводе информации, так как пользователь вынужден часто переводить взгляд с экрана на текст и клавиатуру, находящиеся на разном расстоянии и по-разному освещённые.

Признаками CVS являются снижение остроты зрения, замедленная перефокусировка с ближних предметов на дальние и обратно, двоение предметов, быстрая утомляемость при чтении, чувство жжения в глазах, ощущение «песка» под веками, покраснение глаз, боли в области глазниц и лба при движении глаз.

У части пользователей симптомы CVS обнаруживаются через 2 ч непрерывной работы перед экраном, у большинства — через 4 ч и практически у всех — через 6 ч. При этом считывание информации с экрана считается менее нагрузочным, чем её ввод.

ЗАБОЛЕВАНИЯ МЫШЦ И СУСТАВОВ

Врачи различают несколько синдромов, которые встречаются у пользователей ЭВМ.

Один из них — синдром длительной статической нагрузки, его симптомами являются боли в руках, шее, пояснице. При неудобной рабочей позе мышцы ног, плеч, шеи и рук длительно пребывают в состоянии сокращения. Поскольку мышечные ткани подолгу не имеют возможности расслабиться, то в них ухудшается кровоснабжение, нарушается обмен веществ, накапливаются продукты распада. В результате мышцы находятся в состоянии постоянной усталости и со временем ослабевают. Это может привести к искривлению позвоночника и другим изменениям скелета, а также перерождению мышечных тканей.

Другой синдром — так называемый запястный туннельный синдром, или синдром канала запястья. Его возникновение связано со сдавливанием срединного нерва руки сухожилиями мышц, сгибающих пальцы, при длительной и неудобной для пользователя работе на клавиатуре. Во время частых, повторяющихся движений кистей рук в неудобном положении сухожилия трутся о кости запястья и связки. В результате сдавливания нервов и сухожилий развивается серьёзное недомогание. В начальной стадии болезни её симптомы: дрожь, зуд и покалывание в пальцах — появляются только через несколько часов после окончания работы на компьютере. Как правило, большинство людей не связывают это со своей работой, что приводит к запущенным случаям синдрома. Постепенно присоединяются онемение, боль и тяжесть в руках. В наиболее тяжёлой форме запястный туннельный синдром диагностируется по мучительным болям, лишающим человека трудоспособности и требующим хирургического вмешательства.

КАК СОХРАНИТЬ ЗДОРОВЬЕ

Профилактика профессиональных заболеваний, связанных с работой за компьютером, основана на соответствии оборудования выработанным

стандартам и на соблюдении правильного режима труда.

Как правило, наибольший вред здоровью пользователя наносят устройства ввода-вывода: монитор, клавиатура, мышь. Появляется множество различных стандартов на экологическую безопасность оборудования персонального компьютера. Современный монитор должен соответствовать общепринятым нормам безопасности и эргономике.

FCC Class B — этот стандарт разработан Канадской федеральной комиссией по коммуникациям для обеспечения приемлемой защиты окружающей среды от воздействия радиопомех в замкнутом пространстве. Оборудование, соответствующее требованиям FCC Class B, не должно мешать работе теле- и радиоаппаратуры.

MPR-II, выпущенный Шведским национальным департаментом стандартов, налагает ограничения на излучения от компьютерных мониторов и промышленной техники, используемой в офисе.

TCO-03 — стандарт безопасности компьютерных мониторов, выработанный группой TCO Development (принадлежащей шведской Федерации профсоюзов), распространяется на электронно-лучевые и жидкокристаллические дисплеи. Для ЭЛТ-мониторов TCO-03 устанавливает следующие требования: максимальная яркость — не менее 120 кд/м^2; частота 85 Гц для 22-дюймовых мониторов должна поддерживаться в разрешении 1600×1200. Для ЖК-мониторов разрешение должно составлять не меньше 1024×768 для 15- и 16-дюймовых дисплеев, не меньше 1280×1024 для дисплеев с диагональю 17—19 дюймов и 1600×1200 для 21-дюймовых моделей. Максимальное значение яркости должно достигать 150 кд/м^2.

EPA Energy Star VESA DPMS (США) — согласно этому стандарту, монитор должен поддерживать три энергосберегающих режима: ожидание (standby), приостановку (suspend) и «сон» (off). Такой монитор при долгом простое компьютера переводится в соответствующий режим с низким энергопотреблением.

Эргономичная клавиатура.

Монитор должен иметь возможность регулировки параметров изображения (яркость, контраст и т. д.). Рекомендуется, чтобы при работе с компьютером частота вертикальной развёртки монитора была не ниже 85 Гц. При этом пользователь перестаёт замечать мерцание изображения, которое вызывает быстрое зрительное переутомление.

Для компьютерных устройств ввода (клавиатура и мышь) до сих пор не существует общепринятых и широко распространённых стандартов. Между тем многие производители данного оборудования, рекламируя свою продукцию, описывают различные конструктивные решения, повышающие эргономичность её использования: клавиатура с возможностью регулировать расположение клавиш, мышь такой формы, которая уменьшает усталость кисти при длительной работе, и другие устройства.

Логотип TCO-03.

В России также действуют санитарные правила и нормы «Гигиенические

Монитор с TCO.

УПРАЖНЕНИЯ ДЛЯ РАЗМИНКИ

Положите руку на край стола ладонью вниз. Взявшись за пальцы другой рукой, отведите кисть назад и удерживайте в таком положении в течение 5 с. Повторите упражнение для другой руки. Слегка упритесь рукой в стол, на 5 с напрягите пальцы и запястье. То же проделайте другой рукой. Сильно сожмите пальцы в кулаки, затем распрямите их. Сядьте на стул прямо, ноги твёрдо поставьте на пол (если стул на колёсиках, позаботьтесь о том, чтобы он оставался неподвижным). Наклонитесь как можно ниже, чтобы достать головой колени. Оставайтесь в таком положении 10 с, затем распрямитесь, напрягая при этом мышцы ног. Повторите упражнение три раза. Многие держат на столе резиновую эластичную игрушку или кольцо-эспандер и с их помощью время от времени разминают кисти рук.

требования к видеодисплейным терминалам, персональным электронно-вычислительным машинам и организации работ» СанПиН 2.2.2.542-96.

В части электромагнитных полей они соответствуют международным стандартам. Кроме того, этот документ определяет режим труда и отдыха при работе за компьютером и требования к организации рабочего места. В частности, объём считываемой с монитора информации не должен превышать 60 000 знаков за смену, объём вводимой информации — 40 000 знаков, а суммарное время, проведённое за монитором, — не более 6 ч. На протяжении рабочего дня через каждые 45—60 мин следует устраивать перерывы продолжительностью 10—20 мин, во время которых рекомендуется выполнять комплексы физических упражнений.

Очень важна правильная организация рабочего места. В комнате, предназначенной для работы за компьютером, должно быть как естественное, так и искусственное освещение. Поэтому расположение рабочих мест в подвальных помещениях не допускается. Лучше всего, если окна выходят на север или северо-восток. Офисы необходимо оборудовать не только отопительными приборами, но и системами кондиционирования воздуха или эффективной вентиляцией. Стены и потолки следует окрашивать матовой краской (блестящие и тем более зеркальные поверхности утомляют глаза и отвлекают от работы). В помещениях ежедневно должна проводиться влажная уборка.

Желательно, чтобы площадь рабочего места составляла не менее 6 м^2, а объём — 20 м^3. Стол следует поставить сбоку от окна так, чтобы свет падал слева. Наилучшее освещение для работы с компьютером — рассеянный непрямой свет, который не даёт бликов на экране.

В поле зрения пользователя не должно быть резких перепадов яркости, поэтому окна желательно закрывать шторами либо жалюзи. Искусственное же освещение должно быть общим и равномерным, однако использование одних только настольных ламп недопустимо.

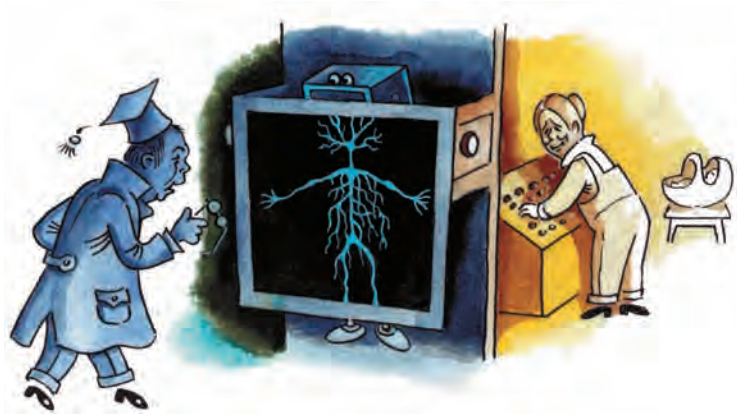
Правильно организованное рабочее место — это лишь первый шаг к профилактике возможных заболеваний. Чтобы работа за компьютером не вредила здоровью, в процессе её необходимо постоянно следить за положением тела. Правильная осанка максимально разгружает мышцы и позволяет работать дольше, меньше уставая. Считается, что при правильной осанке уши располагаются точно в плоскости плеч, а плечи — над бёдрами. Голову следует держать ровно по отношению к плечам. Когда вы смотрите вниз, голова не должна наклоняться вперёд. Сутулость — положение, при котором линия плеч располагается не точно над линией бёдер и под линией ушей, — вызывает чрезмерную нагрузку на плечевые сухожилия и мышцы плеча. Длительная работа в такой позе может стать причиной заболевания.

НЕЙРОСЕТИ

Процесс мышления человека чрезвычайно сложен и лишь частично понятен современным учёным. Однако наука достигла определённых успехов в познании устройства нервной системы человека. Структурной единицей нервной системы человека является нейрон — нервная клетка. Такая клетка состоит из тела (сома с ядром) и множества древовидных отростков — дендритов и длинного аксона. Дендриты служат в качестве входных каналов для нервных импульсов от других нейронов. Импульсы поступают в сому, вызывая её специфическое возбуждение, распространяющееся затем по выводному отростку — аксону. Соединяются нейроны с помощью специальных контактов — синапсов, в которых разветвления аксона одного нейрона подходят очень близко (на расстояние нескольких десятков микронов) к соме или дендритам другого нейрона.

Нейроны, размещающиеся в рецепторах, воспринимают внешние раздражения, в сером веществе ствола головного и спинного мозга — управляют движениями человека (мышцами и железами), в мозге — связывают чувствительные и двигательные нейроны. Последние образуют различные мозговые центры, где происходит преобразование информации, поступившей от внешних раздражителей, в двигательные сигналы.

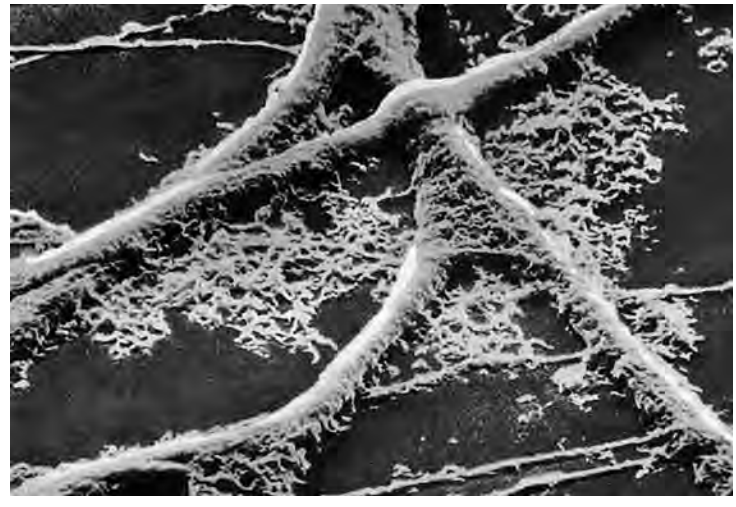
Как же работает эта система? В нейронах происходят три основных процесса: синаптическое возбуждение, синаптическое торможение и возникновение нервных импульсов. Синаптические процессы обеспечиваются особыми химическими веществами, которые выделяются окончаниями одного нейрона и взаимодействуют с поверхностью другого. Синаптическое возбуждение вызывает ответную реакцию нейрона и при достижении определённого порога переходит в нервный импульс, быстро распространяющийся по отросткам. Торможение, напротив, уменьшает общий уровень возбудимости нейрона.



Сотни миллиардов нейронов образуют сложную систему, превосходящую пока смелые мечты о супер-ЭВМ.

Первые опыты по моделированию таких систем, т. е. созданию искусственных нейросетей, проводились ещё в середине XX в. Целью было не только изучать функции мозга, но и попытаться их частично продублировать, используя искусственные нейросети. Схемы соединения нейронов в сети различны, но все сети представляют собой многослойные пространственные структуры.

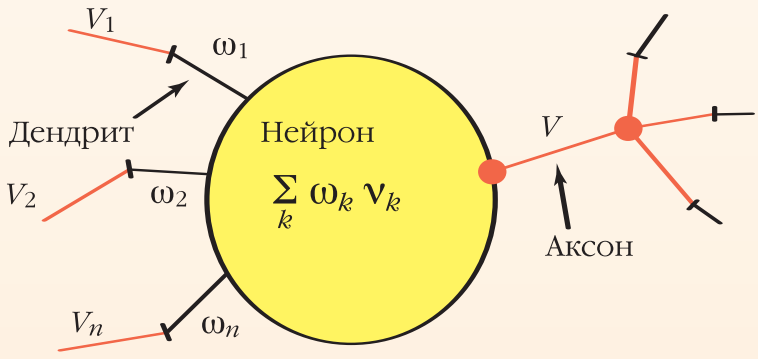
В 1949 г. Д. Хэбб продемонстрировал обучение сети нейронов. Позднее, в 60-х гг., группа исследователей, среди которых был Марвин



Часть нейросистемы человека.

ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

Нейросеть состоит из узлов (аналоги нейронов) и соединений (синаптические связи).



V_i — импульсы, вырабатываемые другими нейронами и поступившие на дендриты i, ω_i — вес дендритов. Импульс V распределяется между дендритами других нейронов с помощью ветвящегося аксона.

В нейросети можно задавать и менять вес синаптических связей и значения порогов для обучения (и самообучения) сети. Такая схема определяет пути прохождения возбуждения через сеть, формируя связи посылка — следствие. В сети существует входной (рецепторный) слой, воспринимающий внешнее возбуждение, и выходной слой, определяющий результат решения задачи. Работа сети тактируется импульсами для имитации и управления проходящим возбуждением. Сеть функционирует или в режиме обучения, или в режиме распознавания (рабочий режим).

Как работает искусственная нейросеть?

Пусть создана сеть для распознавания текста. Входной слой воспринимает изображение, возбуждающее в некой конфигурации множество нейронов — рецепторов. Через некоторое количество тактов окажется максимально возбуждён один из нейронов выходного слоя. Если на вход подан эталон знака, например буква А, то максимально возбуждённый нейрон выходного слоя можно отметить как образ буквы А, словно получен ответ: «Это буква А, я узнал её!». Если когда-нибудь потом снова будет введена буква А, возможно с искажениями, то, по логике, должен возбудиться тот же нейрон, но может и другой, что явится ошибочным поведением системы. Чтобы «научить» систему, необходимо добиться максимального возбуждения того же нейрона выходного слоя, по определённому алгоритму изменив вес или пороги в сети на пути прохождения возбуждения так, чтобы заставить возбудиться нужный нейрон.

Путём последовательного предъявления множества эталонов и регулирования параметров сети и осуществляется её обучение. Оно заканчивается тогда, когда вероятность ошибки становится мала, что соответствовало заданным условиям. Предъявляя «обученной» сети различные буквы, можно быть уверенным, что при появлении искажённой в определёной мере буквы она будет распознана. Сеть фактически имитирует ассоциативное мышление.

Минский, разработала сети, состоящие из одного слоя искусственных нейронов. Сети использовали для решения разных задач, таких, как предсказание погоды, анализ электрокардиограмм и создание искусственного зрения. Впоследствии в работе «Персептроны» Минский вместе с Пайпертом доказывал, что однослойные сети в принципе не способны решить многие простые задачи, в том числе моделировать исключающее ИЛИ. Подобные оценки оказались излишне пессимистичными, немало из перечисленных в книге задач решаются сейчас многослойными сетями с помощью набора стандартных процедур.

В конце 80-х гг. интерес к нейронным сетям снова вырос. Существовало уже много примеров обучения нейронных сетей: одну сеть научили переводить текст в фонетическое представление, которое затем превращалось в речь; другую — распознавать рукописные буквы. С использованием искусственных нейросетей была сконструирована система сжатия изображений.

Основным методом обучения многослойных сетей является так называемое обратное распространение. Однако даже этот метод не даёт никаких гарантий, что сеть будет обучена за конечное время и окажется лучше при повторном обучении.

Люди привыкли, что компьютеры не допускают ошибок, но искусственные нейросети, смоделированные на компьютере, иногда будут их совершать даже при правильном функционировании. Нейронные сети не способны «объяснить», как они решают задачу. Внутреннее представление, складывающееся в результате обучения, часто настолько сложно, что его нельзя проанализировать. Человек также подчас не в силах объяснить, как он узнал в толпе друга, которого не видел многие годы. Экспертная система прослеживает процесс рассуждений в обратном порядке, так что может проверить логичность.

Считают, что искусственные нейронные сети заменят собой искусственный интеллект. Однако есть вероятность, что они будут существовать

совместно в системах, где каждый подход используется для решения тех задач, которым он лучше соответствует. Распознавание образов часто требует быстрой реакции. Когда нервная система не в состоянии распознать новый объект, вопрос передаётся в высшие отделы мозга. Если потребуется дополнительная информация, то процесс поиска решения займёт больше времени, но качество результата повысится. Так, нейронная сеть реагировала бы в большинстве случаев на внешнюю среду, а сложные случаи, когда вероятность ошибки велика, передавались бы для решения экспертной системе. Комбинация этих двух систем была бы более эффективной и следовала модели, полученной в результате биологической эволюции.

В известной игре «Creatures» (*англ.* «создания»), в которой необходимо выращивать, воспитывать и обучать поколения смешных существ норнов, для моделирования поведения персонажей, реакции на внешние раздражители и обучения использовались нейронные сети.

НАНОТЕХНОЛОГИИ

ЧТО ТАКОЕ «НАНО»

В современном научном языке приставка «нано» служит для образования наименования дольных единиц физических величин, равных одной миллиардной исходных единиц. Например, нанометр — это одна миллиардная часть метра, т. е.

$$1 \text{ нанометр (нм)} = 0{,}000\,000\,001 \text{ м} = 10^{-9} \text{ м}.$$

Учёные и инженеры договорились о точном значении этой приставки в 1960 г., произведя её от слова «нанос» (в переводе с греческого «карлик»), которое употребляется уже тысячи лет.

Соответственно под нанотехнологиями понимают технологии, имеющие дело с чем-то карликовым, очень маленьким, составляющим всего лишь миллиардные доли чего-то, скажем миллиардные доли метра.

С размером в 0,001 м, или миллиметром, часто встречаются в повседневной жизни. Размер булавочной головки чуть больше 1 мм. Но по сравнению с нанометром миллиметр огромен, он равен целому миллиону нанометров. Инженеры уже давно привыкли иметь дело с тысячными долями миллиметра — микрометрами (мкм), или микронами:

$$1 \text{ мкм} = 0{,}000\,001 \text{ м} = 10^{-6} \text{ м} = 1000 \text{ нм}.$$

Диаметр человеческого волоса — 1/20 мм, т. е. 50 мкм. Цветочная пыльца состоит из пылинок размером несколько микронов, которые незаметны невооруженным глазом.

При анализе крови человека лаборанты рассматривают образцы под микроскопом и в первую очередь подсчитывают количество красных кровяных клеток, имеющих форму дисков диаметром несколько микронов.

А как часто сталкиваются с размерами меньше микрона? Все видели радугу на небе или в брызгах воды фонтана. Цвета в радуге меняются от красного до фиолетового. Физика учит, что они различаются длиной световой волны — от 0,44 мкм (красный) до 0,7 мкм (фиолетовый).

Как же представить себе один нанометр? Надо изготовить миниатюрный металлический стерженёк длиной 1 мкм, поместить под микроскоп

Единица измерения микрон обозначается μm. Из-за того, что вводить греческие буквы при наборе текстов на русском или английском языке неудобно, наиболее употребительно неофициальное обозначение um. Если у греческой буквы μ («мю») стереть нижний хвостик, получится как раз буква латинского алфавита u.

Длина световой волны не так уж мала. На расстоянии, равном диаметру человеческого волоса, умещается всего лишь 100 колебаний световой волны.

Сканирующий туннельный микроскоп.

и с помощью какого-нибудь хитроумного приспособления разрезать пополам. Одну половинку выкинуть, другую снова разрезать пополам. Возможно ли, хотя бы теоретически, повторить процедуру десять раз подряд? Исходный стерженёк состоит из атомов, и после разрезания его размер никак не станет меньше расстояния между атомами. Для большинства металлов расстояние между атомами составляет 0,1—0,7 нм.

Таким образом, на стерженьке длиной 1 мкм уместится всего несколько тысяч слоёв атомов металла. После первого деления стерженька число слоёв уменьшится примерно вдвое, после второго — примерно вчетверо, после десятого — примерно в 1 тыс. раз, т. е. после десяти делений (если бы удалось их проделать) длина стерженька стала бы около 1 нм и стерженёк состоял бы всего из нескольких (меньше десяти) слоёв металла. Следовательно, нанометр — единица, удобная для измерения объектов, состоящих из небольшого числа атомов.

Теперь вновь вернёмся к слову «нанотехнологии». На самом деле приставка «нано» в этом слове подчёркивает не размеры объектов, а их тип. Нанотехнологии нацелены на индивидуальную работу с отдельными атомами.

Обычные технологии, например обработка детали на токарном станке, варка стекла или изготовление бетонных конструкций, создают предметы, т. е. собирают атомы в нужные скопления, манипулируя огромными плохо управляемыми группами атомов.

Нанотехнологии тесно связаны с информатикой. Во-первых, к необходимости создания нанотехнологий пришли после постановки Р. Фейнманом в своём докладе задачи компактного хранения информации. Во-вторых, нанотехнологии обещают радикально изменить как инструменты обработки информации (компьютеры), так и методы их использования. В-третьих, развитие нанотехнологий невозможно без применения методов информатики.

ПЕРВЫЕ ЛАБОРАТОРНЫЕ УСПЕХИ

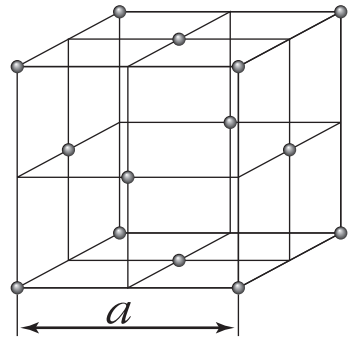
Ещё философы Древней Греции догадались, что вещество состоит из атомов. Затем это было доказано теоретически и экспериментально, но ни один человек не мог поклясться: «Я видел атом!». Различить отдельный атом с помощью какого-либо оптического прибора в принципе невозможно. Оптические приборы

работают со световыми колебаниями. Их разрешающей способностью является расстояние между двумя объектами, на котором они ещё отличимы друг от друга (расстояние не превышает длины волны световых колебаний). Длина волны видимого света около 500 нм, что в тысячи раз превышает расстояние между атомами. Для «подглядывания» за атомами нужны более частые колебания. С точки зрения современной физики электрон не только частица, но и волна, колебание. Длина этой волны меньше расстояния между атомами, посредством электронных волн в микроскоп с «электронным светом» можно увидеть отдельный атом (отличить его от других атомов). В 80-х гг. создали такой микроскоп, названный *сканирующим туннельным микроскопом* (СТМ).

Основная идея состоит в том, чтобы в вакууме перемещать над поверхностью твёрдого тела кончик острой иглы, к которой приложено напряжение. Если расстояние между образцом и кончиком иглы достаточно мало, то электроны туннелируют (перескакивают) с острия иглы на образец, образуя ток туннелирования. Водя иглой по образцу и измеряя ток, исследователи получают возможность «нанести на карту» расположение микроскопических (атомных размеров) «холмов» и «долин» на поверхности образца.

В 1986 г. изобретатели СТМ были удостоены Нобелевской премии. Сканирующий туннельный микроскоп, умещающийся (если без вакуумной камеры) на ладони, имеет разрешение по вертикали детали размером в 0,1 Å, или, иначе говоря, одну десятую диаметра атома водорода. Разрешающая способность сканирующего острия шириной всего в несколько атомов допускает разрешение детали горизонтальной плоскости размером не более 2 Å. Учёным уже удалось изготовить остриё шириной в один атом. Наконечник иглы делается в форме пирамиды, предпоследний и последний слои состоят из трёх и одного атомов соответственно.

СТМ позволяет не только видеть, но и перемещать атомы. Например, в 1980 г. сотрудники фирмы IBM на-



a — расстояние между атомами.

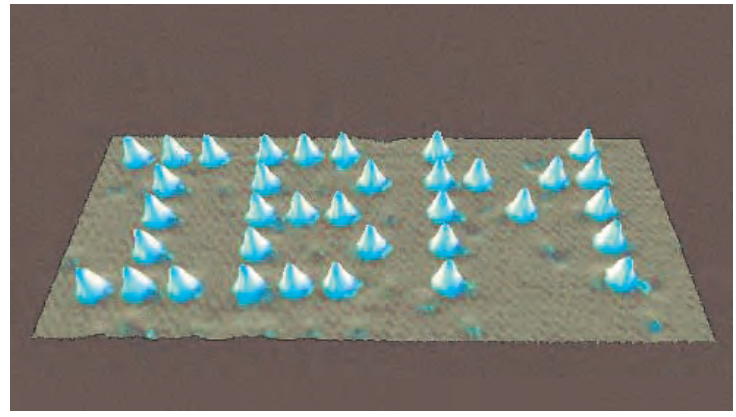
несли на никелевую подложку 35 атомов ксенона, выложив из них название своей компании.

НЕИЗБЕЖНОСТЬ «РАЗДЕЛЕНИЯ ТРУДА» В НАНОТЕХНОЛОГИЯХ

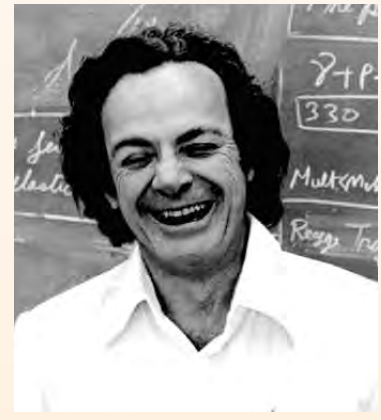
Манипулируя отдельными атомами и молекулами, в принципе можно создавать новые устройства разных размеров. От микроскопических, неразличимых для невооружённого глаза, до устройств планетарного масштаба, по величине превосходящих Землю. Даже самые крошечные устройства будут содержать огромное количество атомов. Чтобы выполнить устройство с помощью нанотехнологии, каждый атом придётся переместить на отведённое ему место. Мысленно проведём эксперимент. Для этого потребуется простейшая фигура — кубик из атомов одного типа. Сколько атомов и на какое расстояние придётся переместить, чтобы собрать кубик

«Любой материальный предмет — это всего лишь скопление атомов в пространстве. То, как эти атомы собраны в структуру, определяет, что́ это будет за предмет».

Станислав Лем



«ТАМ, В ГЛУБИНЕ...»



Ричард Фейнман.

В 1959 г. американский физик, будущий лауреат Нобелевской премии Ричард Филлипс Фейнман выступил перед Американским физическим обществом с докладом «Там, в глубине, полным-полно свободного места (Приглашение начать новое направление в физике)».

Нанотехнологии тесно связаны с информатикой по нескольким причинам. Исторически обсуждение нанотехнологий началось с задачи компактного хранения информации. Применение нанотехнологий обещает радикально изменить как компьютеры, так и методы их применения. Развитие нанотехнологий невозможно без применения методов информатики.

Начинался доклад с неожиданного вопроса: можно ли записать все 24 тома Британской энциклопедии на булавочной головке?

Фейнман ответил на этот вопрос положительно. Чтобы понять его рассуждения, помимо сведений, входящих в школьную программу, нужно знать только, что расстояние между атомами булавочной головки, сделанной из обычного металла, составляет доли ангстрема (Å — единица измерения длины, равная 0,1 нм).

«Размер булавочной головки составляет примерно 1,5 мм. Если увеличить диаметр головки в 25 тыс. раз, то её площадь станет примерно равна суммарной площади всех страниц энциклопедии. Значит, всё, что нужно сделать, это научиться писать в 25 тыс. раз мельче. Возможно ли такое? Разрешающая способность глаза около 0,2 мм, что примерно равно диаметру точек, из которых складываются полутоновые картинки. Если уменьшить подобную точку в 25 тыс. раз, то получится область 80 Å в диаметре, т. е. равная по размеру 32 расстояниям между атомами обычного металла. Иными словами, после уменьшения наша точечка будет всё ещё содержать около тысячи атомов. Значит, ей можно легко придать нужную форму в процессе фотогравировки и, без всякого сомнения, на булавочной головке найдётся достаточно места, чтобы нанести на неё содержимое всей Британской энциклопедии».

Далее Фейнман заявил, что накопленную человечеством информацию вполне можно уместить в небольшой брошюре, и притом не в закодированном виде, а просто уменьшив в размере исходный материал (со всеми рисунками, репродукциями, таблицами и т. п.). Если же, продолжил учёный, хранить информацию не в виде пропорционально уменьшенной копии на поверхности вещества (булавочной головке), а в закодированном виде внутри вещества, то результаты будут ещё более поразительны. Представим, что информация закодирована точками и тире, причём точка изображена в виде атома одного металла, а тире — в виде атома другого. Если хранить каждую такую порцию информации внутри маленького куба размером 5 x 5 x 5 атомов, то после несложного подсчёта получится, что «вся информация, которую человечество собирало в книгах за всю свою историю, может быть помещена в кубик размером чуть меньше 0,1 мм — пылинку, едва различимую человеческим взглядом».

В заключение доклада Фейнман сказал:«Физические принципы, насколько я могу видеть, не содержат запрета на возможность манипулировать предметами атом за атомом. Такое манипулирование не нарушает никаких законов, в принципе оно возможно; однако на практике это ещё не делалось, поскольку мы слишком велики».

с ребром 5 мкм? 1 мкм — 1000 нанометров, или 10 тыс. ангстрем. Если принять, что расстояние между соседними атомами в кубике 0,5 Å, то число атомов на ребре кубика будет равно 100 001, а общее число атомов в кубике — $100\,001 \cdot 100\,001 \cdot 100\,001 \approx \approx 10^{15}$.

Попробуем собрать кубик, используя сканирующий туннельный микроскоп. Предположим, что на перемещение каждого атома понадобится одна секунда, тогда на всю сборку уйдёт около 30 млн лет. Такую технологию трудно назвать практически применимой. Два СТМ позволили бы сделать эту работу вдвое быстрее, а десять СТМ, к сожалению, не удалось бы задействовать одновременно — вокруг собираемого кубика они не разместятся!

Чтобы сложить кубик за разумное время (например, за несколько часов, минут или секунд), нужно либо повысить скорость перемещения атомов, либо использовать большое количество одновременно работающих «сборщиков» (универсальные наномашины, или самоусовершенствующиеся микророботы), которые настолько миниатюрны, что не помешают друг другу.

Во сколько раз надо ускорить перемещение атомов, чтобы один «сборщик» выполнил всю работу за 10 с? В этом случае он за 1 с должен перемещать одну десятую от общего числа атомов. Пусть каждый атом необходимо передвинуть на расстояние 10 мкм, тогда за 1 с «сборщик» произведёт 10^{14} перемещений по 10 мкм, т. е. общее перемещение составит 10^{15} мкм = 10^9 м = 10^6 км > 300 000 км.

Таким образом, за 1 с мельчайшие сдвиги атомов, производимые «сборщиком», в целом составят расстояние 1 млн километров . Чтобы подобного достичь, он должен временами двигаться в несколько раз быстрее скорости света, что противоречит принципу относительности.

При создании устройств размерами в миллиметры или метры по принципам нанотехнологии трудности усугубятся. Так, расстояния, на которые придётся перемещать атомы, при переходе от микронов к миллиметрам и метрам увеличатся в тысячу и миллион раз соответственно, а число атомов возрастёт в 10^9 и 10^{18} раз.

Следовательно, создание продукции в нанотехнологии возможно, только когда «пункты сборки» распределены по всей поверхности (объёму) создаваемого устройства и работают параллельно.

Чтобы представить такой параллелизм, вообразим повреждённый муравейник и его обитателей, занятых починкой. Толпы муравьёв работают не мешая друг другу. По сравнению с размером муравейника размер любого муравья крайне мал. Хотя отдельные муравьи трудятся относительно независимо, их совместная работа подчинена единой цели.

ОСУЩЕСТВИМЫ ЛИ НАНОТЕХНОЛОГИИ

Пример со сборкой кубика наводит на грустные мысли. Неужели основные физические принципы и законы доказывают практическую неосуществимость нанотехнологий, подобно тому как принцип относительности говорит о невозможности полёта к ближайшей звезде на выходные, а принцип сохранения энергии — о невозможности постройки вечного двигателя?

Один из теоретиков нанотехнологий, американский учёный Эрик Дрекслер, приводит следующий пример. Если бы человечество владело нанотехнологиями, медики могли бы создать устройство размером в несколько микрон, состоящее из мешка, лап

Эрик Дрекслер.

с присосками и хобота. Его вводили бы в кровь человека, и устройство отыскивало бы микробы, присасывалось к ним и через хобот впрыскивало антибиотик, запасённый в мешке. Лечить многие болезни стало бы легче.

Но откуда известно, что нанотехнологии, в отличие от вечного двигателя, вполне возможны? Реально ли создание «карликовых» устройств, не противоречит ли это каким-либо физическим законам? Ответим на вопрос, рассмотрев капельку крови человека под микроскопом. В крови плавают и охотятся за микробами по сути точно такие же «устройства», которые называются антителами. Они появляются в организме по генетическому коду (точный план), шаг за шагом, молекула за молекулой. Всё живое на Земле, от бактерии и простейшего гриба до человека и секвойи, создано с помощью процессов, манипулирующих небольшими группами атомов — аминокислотами и белками. То есть в каком-то смысле можно считать, что нанотехнологии уже работают в живой природе. К настоящему времени на принципах нанотехнологий разработаны конструкции из сотен и даже тысяч атомов, но среди них пока ещё нет ни одной, сравнимой по сложности с живой клеткой. Однако само существование жизни и биологических процессов доказывает практическую осуществимость нанотехнологий.

Любые самовоспроизводящиеся объекты, будь то бактерии в человеческом теле, водоросли в пруду, сорняки на вспаханном поле, закваска в тесте или вирусы в компьютерной сети, быстро размножаются, преобразуя окружающую среду часто нежелательным для человека или даже катастрофическим образом. Одна из таких катастроф описана в известной немецкой сказке «Горшок каши», когда горшочек всё варил и варил кашу, покрыв в конце концов ею весь город. Но действительность может оказаться страшнее любой сказки. Учёные, разрабатывающие подходы к нанотехнологиям, уже сейчас задумываются над опасностью выхода из-под контроля самовоспроизводящихся (или даже самоусовершенствующихся) микророботов. Гипотетическая катастрофа получила название «серая слизь» — так обозначают неконтролируемый процесс переработки почвы, воды, воздуха микроскопическими роботами, при котором Земля покрывается неисчислимой массой «сборщиков».

Подобное может произойти буквально за несколько дней. Представьте себе выброшенного волной на песчаный морской берег самовоспроизводящегося робота размером 10 мкм. Пусть он состоит из атомов кремния, кислорода, водорода, азота, углерода (и некоторых металлов, соли которых растворены в морской воде). При попадании на солнечный свет микроробот начинает самовоспроизводиться (при условии, что рядом есть запас нужных атомов). Если процесс «клонирования» занимает полчаса, то к концу первого дня на пляже вырастет 20 поколений роботов общей численностью около миллиона штук и общим объёмом 1 мм^3. К концу второго дня объём их составит 1 дм^3, к концу третьего — 1000 м^3, а к концу четвёртого дня, если хватит песка, — 1 км^3. Весь пляж превратится в «серую слизь».

Теоретики нанотехнологий уже сейчас начали вырабатывать принципы устройства микророботов-«сборщиков», позволяющие избежать ката-

строфы «серой слизи». Эти принципы похожи на знаменитые законы робототехники, сочинённые писателем-фантастом А. Азимовым. Только они не выдуманы писателями, а изложены учёными для практических целей.

Принцип 1. «Сборщик» должен начинать самовоспроизводство только по команде извне.

Принцип 2. Запрещается разрабатывать процессы сборки, идущие с выделением энергии.

Принцип 3. Для воспроизводства должны быть необходимы вещества, не встречающиеся в природе.

СМЕЛЫЕ ПРОГНОЗЫ

Уже найдено много интереснейших способов применения нанотехнологий, и количество прогнозов увеличивается с каждым днём. Приведём только три описания технических устройств завтрашнего дня.

Механический нанокомпьютер, подобный компьютеру Бэббиджа, способный работать с частотами в сотни гигагерц. Этот компьютер будет состоять из мельчайших деталей, каждая из которых образована всего несколькими тысячами атомов.

Сверхпрочные саморемонтируемые материалы, позволяющие построить башню с лифтом для подъёма полезных грузов в космос. По соединённым углеродным трубкам в стенах такой башни должны ползать «ремонтники», обнаруживая и устраняя повреждения.

Микроскопическая «подводная лодка», плавающая в крови человека и способная транспортировать кислород из лёгких в ткани и углекислый газ — обратно. Если полстакана таких искусственных кровяных телец ввести человеку в кровь, то он сможет обходиться без воздуха несколько часов.

Нанотехнологии немыслимы без компьютеров, коммуникаций, программирования и других элементов информатики. Роботы-«сборщики» будут получать и обрабатывать информацию извне, а также обмениваться информацией друг с другом. Мощности передатчиков и чувствительности приёмников микроробототов хватит только для связи на короткие расстояния, так что роботы, скорее всего, станут ретранслировать информацию, образуя коммуникационные сети. Микроробототов и программы для них придётся создавать на самой современной технике: сначала на обычных электронных компьютерах, а впоследствии и на нанокомпьютерах.

Айзек Азимов.

УВИДИТ ЛИ СЕГОДНЯШНЕЕ ПОКОЛЕНИЕ РЕЗУЛЬТАТЫ ВНЕДРЕНИЯ НАНОТЕХНОЛОГИЙ

Практическое применение нанотехнологий не за горами. Сегодня за 4 тыс. долларов продаётся 500-страничный отчёт, в котором для бизнесменов

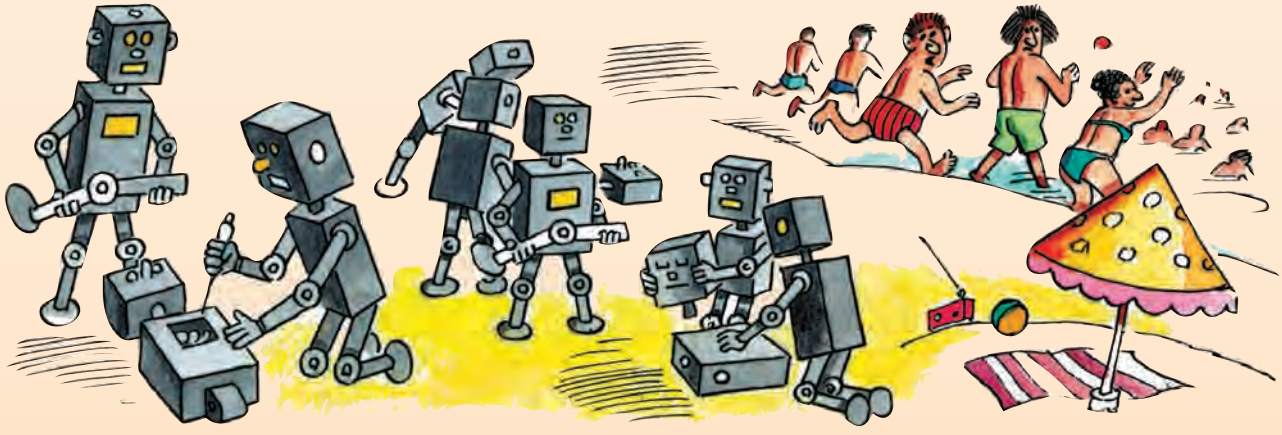
НАНОРОБОТЫ

Систематическое изложение идей нанотехнологии было опубликовано в середине 80-х гг. XX в. Э. Дрекслером в книге «Машины творения». Основная идея заключалась в разработке и массовом производстве специальных роботов-«сборщиков», способных по заданной программе собирать новые конструкции, в том числе и самих себя. Человек сначала использовал простейшие инструменты для создания более сложных, а затем после их усовершенствования возникали технологии, позволяющие выпускать автомобили и самолёты, компьютеры и телевизоры и многие другие полезные вещи. Так и в нанотехнологиях придётся создать несколько поколений универсальных наноинструментов, пока, наконец, с их помощью можно будет произвести что-то полезное.

Дрекслер считает, что первые поколения наноинструментов станут копировать конструкции живой природы: «...гибкая, программируемая белковая машина схватит большую молекулу (объект работы), в то время как маленькая молекула будет установлена именно напротив правильного места. Подобно ферменту, она тогда свяжет молекулы вместе. Привязывая молекулу за молекулой к собираемому куску, машина собирает всё большую и большую структуру, в то же время ведётся полный контроль за тем, как упорядочены атомы. Это есть ключевое умение, которым не обладают химики. Подобно рибосомам, такие наномашины смогут работать под управлением молекулярных лент».

После того как с помощью простейших белковых машин будут построены сложные, а затем ещё более сложные, удастся создать первые универсальные «сборщики».

«...Это второе поколение наномашин, построенное из чего-то большего, чем только белки. Некоторые новые наномашины смогут служить как усовершенствованные устройства для сборки молекулярных структур. Устойчивые к кислоте или вакууму, замораживанию или нагреву, в зависимости от цели использования ферментоподобные машины второго поколения будут применять в качестве инструментов почти каждую из химических молекул, они смогут связать атомы для получения практически любой устойчивой структуры. Поскольку "сборщики" позволят нам размещать атомы почти любым разумным образом, то мы сможем построить почти всё, что угодно, что не противоречит законам природы. В частности, и новые "сборщики".



и государственных деятелей даются рекомендации по инвестициям в нанотехнологии. Согласно этому отчёту, первое промышленное применение нанотехнологий произойдёт уже в 10-х гг. XXI в., а спустя ещё десятилетие нанотехнологии образуют заметный сектор в мировой экономике.

БУДУЩЕЕ ВЫЧИСЛИТЕЛЬНЫХ МАШИН

Аббревиатура ЭВМ расшифровывается как электронная вычислительная машина, а английское слово computer (давно вошедшее в русскую речь как синоним ЭВМ) произошло от глагола to compute, что означает «вычислять». Для решения вычислительных задач разрабатывались специализированные ЭВМ. Самыми успешными из них в течение многих лет остава-

лись суперкомпьютеры Cray, которые выпускала американская компания Cray Research.

Первый суперкомпьютер Cray-1 в 1976 г. был установлен в лаборатории Лос-Аламос Министерства энергетики США (ответственного за атомное оружие). Этот суперкомпьютер обошёлся налогоплательщикам в 8,8 млн долларов, мог выполнять 160 млн операций в секунду с 64-разрядными плавающими числами и хранить в памяти миллион таких чисел. Интегральные схемы в компьютере делались по специальной технологии (с использованием дорогого арсенида галлия). Cray достигал пика производительности при выполнении операций не с отдельными числами, а с наборами чисел — векторами. Всё программное обеспечение на этом компьютере, даже сам язык программирования, было модифицировано для достижения рекордной производительности в вычислительных задачах.

Производительность 160 Мфлопс (миллион операций с плавающей запятой) в 1976 г. считалась феноменальным техническим достижением. Но с точки зрения отношения стоимость/производительность Cray-1 не был чем-то выдающимся. Продуктивность 1 млн операций в секунду обходилась покупателям Cray-1 в 55 тыс. долларов. Мини-компьютеры эпохи Cray в пересчёте на 1 млн операций имели примерно такую же стоимость. Именно это экономическое обстоятельство и определило последующее вымирание динозавров компьютерной эволюции — специализированных суперкомпьютеров.

Спустя несколько лет появились первые персональные компьютеры. Поначалу никто из математиков и инженеров, занимавшихся вычислительными задачами, не воспринимал новые игрушки для взрослых всерьёз. В 1981 г. никому и в голову не приходило сравнивать Cray-1 по вычислительной мощности с микропроцессорами Intel, которые использовались в первых моделях персональных компьютеров компании IBM. Однако тиражи суперкомпьютеров измерялись десятками и сотнями экземпляров, а тиражи персональных ЭВМ (и микропроцессоров для них) — миллионами и десятками миллионов.

Вещи, на которые возникает массовый спрос, начинают выпускать в большом количестве, они быстро дешевеют, а их качество улучшается. Одно из самых массовых технических изделий современной цивилизации — универсальный микропроцессор для персонального компьютера. Такой микропроцессор стоит несколько сотен долларов и выпускается тиражами десятки миллионов штук в год.

Специализированных интегральных схем на основе арсенида галлия выпускали намного меньше, чем традиционных микросхем на основе окислов кремния. Поэтому рост производительности микропроцессоров для суперкомпьютеров Cray значительно уступал росту, предсказанному законом Мура, и составлял около 10 % в год. Хотя в период появления первых персональных ЭВМ, в частности в 1980 г., их производительность была несравнимо меньше производительности одного процессора суперЭВМ, темп её роста был существенно выше, и к 1995 г. производительность микропроцессора персональной ЭВМ, измеренная на эталонной задаче, сравнялась с производительностью микропроцессора суперЭВМ.

Компания Cray уловила эту тенденцию и с 1993 г. начала выпускать суперЭВМ с обычными кремниевыми микропроцессорами массовых настольных ЭВМ. К 1998 г. из 500 эксплуатируемых в мире суперЭВМ в 65 % компьютеров использовали массовые микропроцессоры, ещё в 25 % — специализированные кремниевые микропроцессоры, и лишь 10 % составили ветераны, в которых применяют арсенид галлия.

Основатель фирмы Cray Research, её главный идеолог и конструктор — гениальный инженер и бизнесмен Сеймур Крэй (погиб в 1996 г. в результате автокатастрофы в возрасте 71 года).

Гордон Мур *(слева)* и Сеймур Крэй.

ЗАКОН МУРА

В апреле 1965 г. Гордон Мур, один из основателей современной микроэлектроники, опубликовал в журнале «Электроника» любопытный прогноз. Сравнив производительность трёх поколений интегральных микросхем на основе окислов кремния, он пришёл к следующему выводу: сложность и производительность интегральных микросхем будут удваиваться каждые 18 месяцев. В частности, Мур утверждал, что к 1975 г. появятся микросхемы, насчитывающие около 65 тыс. элементов.

Гордон Мур.

Десять лет спустя Мур в одной из своих статей отметил, что данный прогноз оправдался с точностью, которой он и сам не ожидал. Здесь же учёный предположил, что его предсказание будет действовать и в последующие годы, и привёл технические доводы. После выхода статьи почитатели Мура назвали его прогноз законом Мура. Хотя этот закон относится ко всем интегральным схемам на основе кремния, наибольшую известность получил частный случай, который касается микропроцессоров. Формулируется он так: производительность микропроцессора удваивается каждые 18 месяцев. Действительно, закон Мура подтверждался в течение многих лет, и специалисты считают, что он будет актуален по меньшей мере до 2012 г.

Сам Гордон Мур в 1975 г. и позже предлагал технические обоснования собственно закона. Однако эти объяснения нельзя назвать исчерпывающими. Экономисты и социологи отнесли закон Мура к так называемым самосбывающимся пророчествам и самовыполняющимся законам, т. е. к законам, которые выполняются именно потому, что все их знают и все в них верят.

Вот как экономисты объясняют, почему закон Мура продолжает действовать в течение столь долгого времени. Производительность микропроцессора персонального компьютера сама по себе не очень важна для потребителя: его интересует, на что способен компьютер в целом. Кроме микропроцессора в персональном компьютере много других компонентов. Если их производительность меньше производительности микропроцессора, то производительность последнего «пропадает», её нельзя использовать в полной мере. Если же производительность компонентов больше производительности микропроцессора, «лишняя» производительность компонентов также «пропадает», ибо не может быть использована микропроцессором.

Предположим, некий изготовитель некоего компонента решил разработать его новую версию. Будем считать, что он наметил это сделать за полтора года. В начале работы необходимо решить, какой должна быть производительность новой версии к моменту выпуска. Это крайне ответственное решение: если производительность будет слишком мала для процессоров, которые появятся на рынке через полтора года, компоненты никто не купит и изготовитель понесёт убытки. А если запланировать излишне высокую производительность, компоненты окажутся дороже, чем у конкурентов, правильно угадавших будущее процессоров, и изделия опять же не станут покупать. Но разработчик знает, что закон Мура действует уже много лет и поэтому смело его использует: он считает, что за 18 месяцев производительность доступных для потребителя микропроцессоров удвоится.

Теперь рассмотрим ситуацию с точки зрения изготовителя микропроцессора. Если запланировать рост производительности выше, чем предполагает закон Мура, лишняя производительность «пропадёт», так как изготовители компонентов не будут на неё рассчитывать.

А если поступить наоборот, то он проиграет конкурентам, которые будут руководствоваться законом Мура. Вот почему единственно правильная линия поведения и для разработчиков микропроцессоров, и для разработчиков компонентов персонального компьютера — развиваться в соответствии с законом Мура. Не быстрее и не медленнее.

Теперь стоит сделать маленькое, но важное замечание. Изготовитель микропроцессоров, который по техническим причинам не сможет более придерживаться закона Мура, непременно разорится. Эта же печальная участь ждёт и тех, кто не имеет финансовых возможностей следовать данному закону.

Теоретически максимальный рост производительности зависит от того, какие средства разработчик в состоянии вложить в создание новых моделей микропроцессора. Чем больше тираж микропроцессора и доходы от продаж, тем бóльшие суммы могут быть выделены на новые разработки. И наоборот, при небольших тиражах микропроцессора и соответственно скромных доходах изготовитель не может себе позволить серьёзные вложения, и как следствие темпы роста производительности микропроцессора данного типа невелики.

УСПЕХИ ИНТЕГРАЛЬНЫХ ТЕХНОЛОГИЙ

Качество кристалла можно охарактеризовать тремя параметрами:

• проектная норма — технологический размер измеряется в микрометрах (миллионная часть метра); этот показатель определяет плотность размещения элементов на кристалле;

• число транзисторов (или вентилей) на кристалле определяет логическую сложность кристалла;

• тактовая частота измеряется в мегагерцах и гигагерцах и определяет быстродействие кристалла, фактически максимально возможное количество переключений вентиля в секунду.

Для 1999 г. минимальные проектные нормы были равны 0,18 мк, максимальный размер кристалла интегральной схемы — 25 х 35 мм, максимальное число транзисторов на чипе — 21 млн, тактовая частота — 1,2 Ггц, максимальное число уровней металлизации — 7, минимальное напряжение питания – 1,5 В, максимальный диаметр пластины — 300 мм.

Год	1989	1992	1995	1998	2001	2004	2007
Технологические размеры (мкм)	0,7	0,5	0,35	0,25	0,13	0,10	0,08
Число транзисторов на кристалле (млн)	2,5	5	10	21	46	110	500
Рабочая частота (МГц)	100	175	300	600	1500	3500	7000

Последние 30 лет развития полупроводникового производства характеризуются увеличением:

• числа транзисторов на кристалле в 4 раза каждые 3 года;

• диаметра пластин в 2 раза каждые 15 лет;

• площади кристалла в 2,3 раза каждые 6 лет;

• стоимости производства в 2 раза каждые 3 года и уменьшением минимальных проектных норм в 2 раза каждые 5 лет.

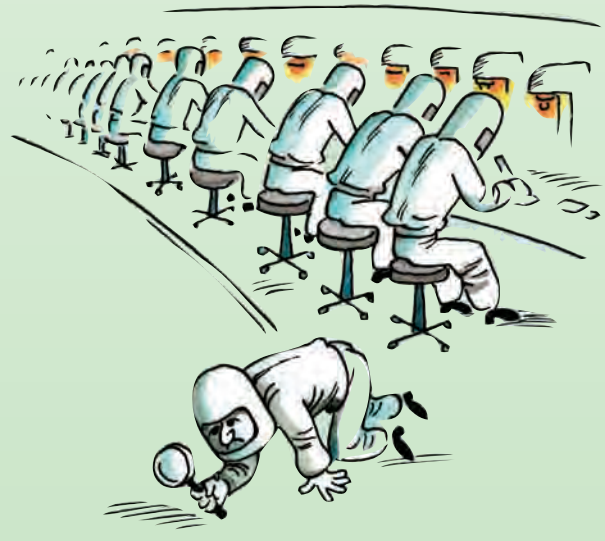
Однако все ориентиры, как правило, требуют корректировки. Так, прогноз 90-х гг., представленный в таблице, не соответствует действительному нынешнему состоянию, например, по частоте микропроцессоров.

Степень совершенства интегральной технологии определяется тем, какие кристаллы можно с её помощью изготовить. Ныне, в начале XXI столетия, уже выпускается запоминающее устройство объёмом 1 Гбит с технологическим размером 0,15 мкм.

Новое промышленное изделие — гигачип, которое принадлежит к семейству КМОП СБИС, с технологическим размером 0,07 мкм обладает памятью 8 Гбит и содержит сотню миллиардов транзисторов. По прогнозам фирмы Intel, к 2100 году число транзисторов, расположенных на одном кристалле, достигнет 10 миллиардов.

В конце прошлого века для полупроводниковой промышленности была типична вертикальная модель организации: полный цикл производства, включающий все технологические операции, проходил на одном предприятии. Высокая стоимость таких производств потребовала перейти к более гибкой горизонтальной специализации. Как самостоятельные звенья выделились автоматизированное проектирование интегральных схем, изготовление фотошаблонов, изготовление интегральных схем на пластине, тестирование и др.

Продвижение в область субмикронных проектных норм отрицательно влияет на долговечность новых продуктов: срок их жизни непрерывно сокращается. Скоро нынешний двухгодичный «жизненный цикл» сожмётся до 9 и менее месяцев. В связи с этим на первый план выходит проблема ускорения на всех стадиях — от разработки до конечного продукта. Если нынешние тенденции сохранятся и в ближайшем будущем, то новые разработки будут попросту невыгодны экономически.



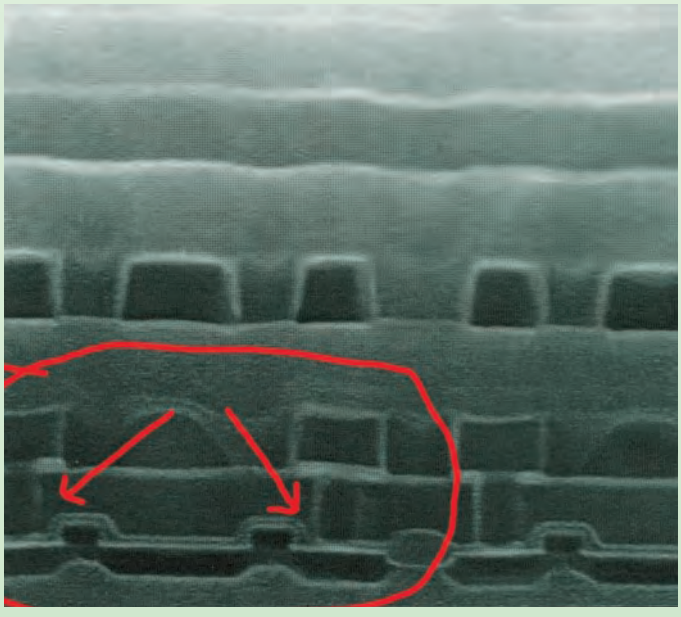
ПРОЕКТНЫЕ НОРМЫ

Один из основных показателей для микропроцессоров и производящих их предприятий — *проектные нормы*. Каждая операция технологического цикла оказывает влияние на производство, однако главные ограничения на выпускаемые интегральные схемы накладывает фотолитография. Проектными нормами при этом являются минимально допустимые размеры между элементами рисунка фотошаблона интегральной схемы. Все остальные операции не должны испортить эту «картину». Так, согласно стандарту, при совмещении фотошаблона и пластины несовпадение не может превышать половину проектной нормы.

Основной размер интегральной схемы — расстояние между истоком и стоком транзистора. Если посмотреть на фотографию, то это расстояние примерно равно ширине затвора, который, собственно, и перекрывает разрыв между истоком и стоком, и оно определяет частоту и энергопотребление интегральной схемы.

Второй по важности размер — величина контактной площадки на истоке или на стоке, образующей в проекции квадрат, своего рода дно колодца, который впоследствии заполняется вольфрамом и выходит на первый уровень металла для осуществления внешних соединений.

Оба размера примерно равны, и их называют проектными нормами.



Срез кристалла с двумя слоями металлизации. Чёрный слой на нижнем уровне образован истоками и стоками комбинированной структуры из двух транзисторов *(слева направо)*: исток 1 — сток 1, над ними затвор, далее исток 2 (совпадающий со стоком1) — сток 2, над ними ещё один затвор. На истоке 1 и стоке 2 находятся колодцы, которые выходят на первый слой металла. Выше второй слой металлизации

Разработчики американской программы ASCI (Accelerated Strategic Computing Initiative — «ускоренная стратегическая компьютерная инициатива») по развитию суперЭВМ для Министерства энергетики США использовали в суперЭВМ микропроцессоры для настольных ЭВМ. Программа успешно выполнялась, и к 2000 г. была введена в эксплуатацию самая производительная ЭВМ в мире — ASCI White (разработка компании IBM) производительностью около 7 Терафлопс (триллион операций с плавающей запятой). Эта суперЭВМ содержала около 10 тыс. микропроцессоров с пиковой производительностью около 1 Гфлопс (миллиард операций с плавающей запятой) каждый.

Тем временем в Японии работали над микропроцессором для вычислительных задач.

В 2000 г. такой микропроцессор показал на эталонной задаче производительность около 10 Гфлопс, а в 2002 г. японская суперЭВМ Earth Simulator компании NEC заняла первое место в списке суперЭВМ, показав на тесте производительность около 35 Тфлопс и «обыграв» предыдущего рекордсмена (ASCI White) более чем в пять раз.

СуперЭВМ Earth Simulator предназначена для моделирования климатических изменений на основе данных, которые поступают со спутников. Высокая производительность компьютера достигнута за счёт применения специально созданного микропроцессора NEC Vector; производительность около 8 Гфлопс. Тем самым векторные микропроцессоры снова взяли верх над микропроцессорами общего назначения.В Earth Simulator использовано свыше 5 тыс. процессоров, она занимает площадь трёх теннисных кортов (50×65 м) и использует кабели общей длиной 2900 км.

Ещё одного компьютерного «монстра» в 2004 г. выпустила компания IBM (по заказу правительства США). Специализированная суперЭВМ Blue Gene установила рекорд производительности 135,3 Терафлопс. Эта электронно-вычислительная машина стала в миллион раз производительнее, чем массовый персональный компьютер 2005 г. выпуска.

КВАНТОВЫЕ КОМПЬЮТЕРЫ

Как известно, вычислительные машины бывают не только цифровые, но и аналоговые. В аналоговых машинах вычисление напрямую моделируется некоторым физическим процессом. Например, можно составить цепь из конденсаторов, сопротивлений и катушек индуктивности и измерять зависимость напряжения на каком-то сопротивлении от времени. Таким образом, чтобы найти решение дифференциального уравнения, нужно лишь подобрать цепь, где напряжение подчиняется как раз этому уравнению. В цифровых машинах то же самое дифференциальное уравнение решается совсем по-другому: численные значения хранятся в виде набора битов, с которыми выполняются арифметические операции.

Постепенно аналоговые машины были вытеснены цифровыми. Это произошло потому, что аналоговые машины решают задачи какого-то определённого класса. Цифровые же более универсальны. Одна и та же цифровая машина может выполнять самые разные вычисления, стоит только заменить программу. При этом проигрыш в скорости компенсировался точностью вычислений. Иногда считают, что по этой же причине компакт-диски заняли место грампластинок.

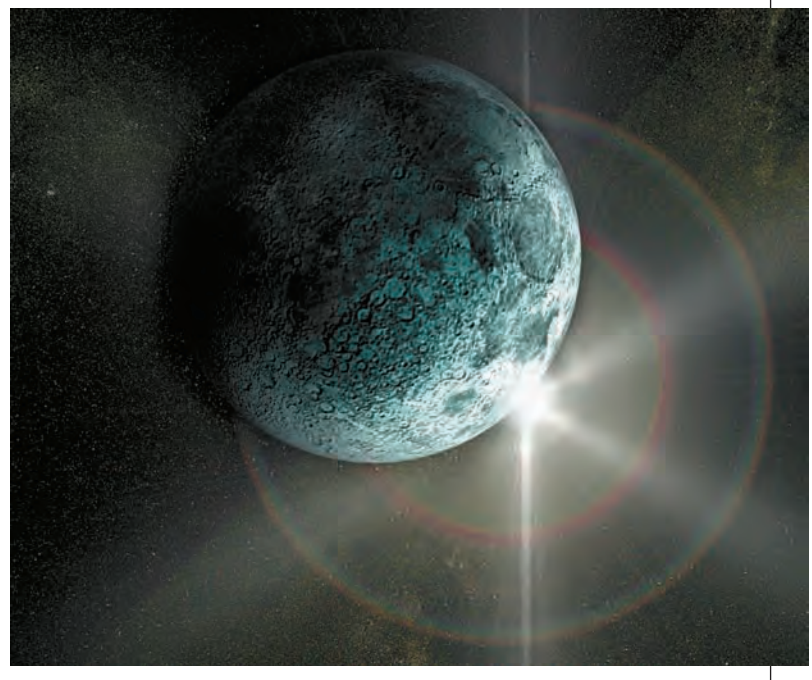
Идея квантового компьютера снова возвращает человечество к аналоговым машинам. При этом важны два обстоятельства.

Современная теория сложности вычислений показала, что некоторые задачи оказываются «универсальными» в том смысле, что к ним сводится неожиданно много других задач. Например, существуют так называемые NP-полные задачи. Одна из них (задача о гамильтоновом пути) проверка того, найдётся ли в данном графе путь, проходящий через все вершины по одному разу. В настоящее время не известно никакого быстрого алгоритма решения этой задачи. Зато доказано, что, если бы такой алгоритм существовал, многие другие задачи тоже решались бы быстро. Тогда умение находить путь в графе могло бы помочь разлагать числа на множители.

На обычной цифровой машине можно достаточно быстро моделировать законы классической механики, а моделирование квантовых процессов либо невозможно, либо требует очень долгих вычислений. Следовательно, если бы квантовый компьютер умел быстро решать какую-то универсальную задачу, то его можно было бы использовать для решения многих задач, непосильных для современных цифровых вычислительных машин.

Эту идею воплотил американский учёный Питер Шор, описав воображаемую машину, которую можно использовать для быстрого разложения больших чисел на простые множители. Про задачу разложения на множители не знают, является ли она универсальной, но она достаточно важна сама по себе. Дело в том, что различные системы шифрования, широко используемые в практике, основаны на том, что большие числа раскладывать на множители трудно. Если способ разложения будет найден, то существующие системы шифрования

Питер Шор.



станут непригодными. Неудивительно, что идея создания квантового компьютера взволновала многих.

Для алгоритма Шора необходима машина как минимум с несколькими тысячами элементов — кубитов (квантовые биты), тогда как построение одного-двух таких элементов — предел возможностей современной техники эксперимента. Кроме того, алгоритм предполагает, что все квантовые вычисления проходят с абсолютной точностью (на практике это неосуществимо), тем самым для практической реализации подобных алгоритмов нужны новые идеи: как сделать вычисления устойчивыми к ошибкам.

Но так или иначе, идея квантовых вычислений уже привела к интересным исследованиям как в теории сложности вычислений, так и в физике (теоретической и экспериментальной). В теории сложности вычислений появились новые классы задач, а в физике стали рассматривать и наблюдать состояния систем из небольшого числа индивидуально заданных квантовых частиц.

КОМПЬЮТЕР 2010 ГОДА

Ваш домашний компьютер для вас представляет смысл жизни, если его стоимость выше цены новых «жигулей».

К прогнозу, какой компьютер станет массовым в 2010 г., можно подойти с двух позиций. Одна из них предполагает акцент на технических параметрах. Надо угадать производительность (тактовую частоту) процессора, выпускающую фирму, объёмы оперативной и долговременной памяти, компоновку компьютера, тип и характеристики монитора, средства связи и скорость обмена при использовании сети Интернет и многое другое.

При бурном развитии электроники стать таким прорицателем крайне

сложно, даже если воспользоваться законом Мура (см. дополнительный очерк «Закон Мура»).

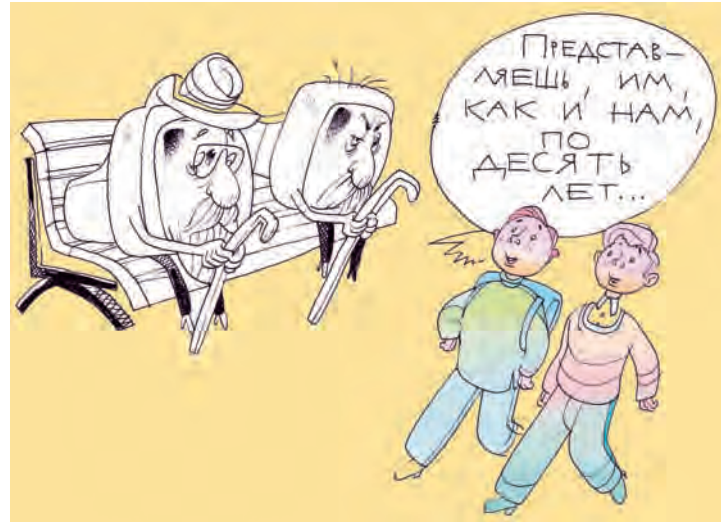
Другой подход — описать круг задач и набор функций компьютера, которых ещё нет, но которые могут появиться в ближайшее десятилетие. Сейчас уже мало говорят про поколения ЭВМ, хотя после четвёртого должно было идти пятое (в качестве шутки можно считать процессор Pentium его представителем).

Как и ранее, многие проблемы, решения которых ждали от систем искусственного интеллекта, остаются в стадии исследования.

Так, пока не получается научить компьютеры «мыслить» подобно человеку, часто принимающему быстрые решения на основании интуиции, не базируясь на жёстких логических правилах и выводах. Если будущие ЭВМ удастся вооружить «подсознательной логикой», то машины окажутся в состоянии заменить живых экспертов во многих областях человеческой деятельности. А знания, накопленные одной такой системой, легко сможет получить (например, по сети Интернет) другой компьютер, чтобы использовать их при принятии решений.

В последнее время идёт активный процесс слияния компьютера, средств связи и бытовых приборов в единый набор. Портативность, энергопотребление, надёжность, объём производства — вот характеристики, по которым будут оцениваться новые изделия электронной промышленности. Дальнейший прогресс связан с созданием новых систем, размещённых в одной интегральной схеме и включающих как сам микропроцессор и его окружение, так и программное обеспечение.

Можно предположить, что производство универсальных компьютеров начнёт сокращаться, а им на смену придут интегрированные приборы, решающие конкретный спектр задач своего владельца: сотовый телефон плюс органайзер плюс фотокамера — так называемый смартфон, к этому можно добавить web-телевизор и др. Современные компьютеры достаточно сложны, и, несмотря на героические усилия разработчиков, поддерживать работоспособность такой системы не каждому под силу. Речь идёт в первую очередь о непрофессиональных пользователях вычислительной техники. Таким людям нет смысла приобретать универсальный компьютер, им гораздо нужнее, чтобы купленный прибор был хорошо приспособлен для решения конкретных задач.



В 80-х и 90-х гг. XX в. активно велись исследования и даже появлялись более или менее успешные разработки

Противники систем искусственного интеллекта шутят, что искусственный интеллект необходим тогда, когда не хватает естественного.

Вы находитесь в авангарде информационного прогресса, если кожух вашего компьютера никогда не закрепляется винтами, чтобы не терять время на их выкручивание при очередном upgrade.

Перефразируя Шекспира, можно сказать: «Менять или не менять — вот в чём вопрос» (подразумевается компьютер и прочая техника).

ЭВМ, способных воспринимать голосовые команды человека. При этом машина должна была не только выполнять широкий спектр таких команд, но и «на слух» вводить текст, работая в качестве секретаря. Как легко стало бы тогда писать статьи для энциклопедий! Комбинация речевого ввода с уже существующей технологией синтезированной речи даёт возможность вместо утомительного нажимания на клавиши вести с компьютером увлекательный разговор.

Компьютеры наконец-то займут ведущее место в здравоохранении. Крошечные ЭВМ смогут управлять механическими органами, заменяющими повреждённые органы человека, например попавшего в аварию или просто состарившегося. Компьютеры станут таким же привычным атрибутом, каким ныне являются очки, и без этого маленького помощника больной будет так же некомфортно себя чувствовать, как и страдающий слабым зрением — без очков. Карманный компьютер проинформирует владельца о статьях в журналах, сообщит о погоде в мире, позвонит в автосервис, закажет билеты на поезд; он сумеет также покупать и продавать акции, проводить операции по финансовым счетам (вовремя платить налоги) и др.

Как показывает неумолимая статистика, время жизни персонального компьютера в среднем не превышает трёх лет. При этом компьютер остаётся ещё вполне работоспособным, однако морально устаревает, и производители оборудования и программного обеспечения предлагают пользователям всё новые и новые ЭВМ. Вложенные в домашний компьютер средства теряются наполовину примерно через два года эксплуатации. Продать за бесценок технику жалко, но, как правило, её можно менять не целиком, а по частям, каждый раз получая улучшенное качество. Этот процесс называют upgrade.

Замена в машине практически всех деталей (процессора, памяти, дисков) позволяет экономить около 150—250 долларов по сравнению с покупкой нового компьютера. При наличии у компании 2 тыс. машин такая экономия существенна. Однако не надо заблуждаться: upgrade не беспределен, и в итоге полученный в результате многих усовершенствований компьютер станет уступать новым, только что разработанным. Каждый компонент персонального компьютера имеет свои чёткие границы использования. Наиболее широки они пока у дисководов для дискет 3,5 дюйма. Фактически они претерпе-

ли изменения лишь однажды, когда формат 720 кбайт был заменён на 1,44 Мбайт. В остальном дисководы при необходимости могли кочевать с компьютера на компьютер. А к наименьшим долгожителям относится одна из самых дорогих частей персональных ЭВМ — микропроцессор.

Производители программного обеспечения вовлекают пользователей ЭВМ в бесконечную гонку за новым качеством. Часто очередные новинки, будь то операционные системы или программы для обработки изображений, практически не добавляют в свой обширный набор функций ничего существенного, однако вынуждают не только приобрести последнюю версию, но и значительно обновить компьютер. Так, при покупке самой современной версии операционной системы можно обнаружить, что часть оборудования, приобретённого всего год назад, уже не поддерживается производителем, т. е. не будет работать в данной ОС. Не исключён даже сговор между производителями программного обеспечения и аппаратуры.

Кроме того, малый срок службы современных ЭВМ существенно понизил уровень качества компьютерных компонентов, особенно имеющих механические части. Если внимательно изучить внутренности современных приводов CD ROM некоторых известных фирм, нетрудно понять, почему они не могут работать больше двух лет, а то и просто не выдерживают гарантийного срока.

ПЕРЕДАЧА ИНФОРМАЦИИ И КОМПЬЮТЕРНЫЕ СЕТИ

ПЕРЕДАЧА ИНФОРМАЦИИ НА БОЛЬШИЕ РАССТОЯНИЯ

Потребность в обмене информацией появилась в глубокой древности. О приближении врага узнавали, приложив ухо к земле; с помощью сигнальных костров производили дистанционный обмен новостями. После возникновения письменности информация на большие расстояния стала передаваться по почте. Если вначале такая связь имела разовый характер, то уже в Древней Греции, Персии, Древнем Египте, Китае и Римской империи существовала регулярная государственная почта. Как правило, сообщения передавались пешими и конными гонцами по принципу эстафеты. В Средние века в Европе получила развитие негосударственная почта: монастырская и университетская. Иногда она за плату доставляла сообщения и частным лицам. Развитие торговли было немыслимо без существования почтового сообщения между городами.

На Руси в X в. существовал *повоз* — повинность населения предоставлять лошадей с повозками для княжеских гонцов. Организованная в XIII в. специальная служба для пересылки писем — *ямская гоньба* просуществовала до второй половины XIX в. С учреждением Почтового департамента в 1782 г. ямские дворы стали называть почтовыми станциями.

Переписка велась не только внутри России, но и с зарубежными странами. Во второй половине XVII в. появились почтовые маршруты из Москвы в Ригу и Вильно. При этом письмо находилось в пути более 10 дней.

Россия одной из первых стран в мире в 1837 г. для перевозки почты стала использовать железнодорожное сообщение. В 1843 г. создана городская почта в Петербурге, в 1845 г. — в Москве.

Романтическую страницу в истории почты вписала знаменитая американская почтовая служба «Пони-экспресс», организованная в начале 60-х гг. XIX в., ещё до постройки железной дороги с запада на восток страны, когда доставка почты была не только срочным, но и опасным делом.

Первое регулярное почтовое сообщение осуществлялось между городами Сент-Джозеф (штат Миссури) и Сакраменто (штат Калифорния). Работа конной почты шла под девизом «Почта должна быть доставлена при любых обстоятельствах». Курьеры доставляли почту на расстояние более 3 тыс. километров в течение десяти дней.

Появление в начале XIX в. более быстрых средств доставки почтовой корреспонденции — паровоза и парохода, а в начале XX в. — самолёта значительно ускорило пересылку отправлений. Практически до конца XX в., независимо от изобретения телеграфа (1832 г.), телефона (1876 г.) и радио (1895 г.), почта оставалась самым массовым и дешёвым видом связи.

С созданием новых скоростных (компьютерных) средств коммуникации почта и в XXI в. не перестанет

Во второй половине XX в. в США насчитывалось 32 тыс. почтовых предприятий, в Великобритании — 25 тыс., в Японии — 20 тыс., во Франции — 18,7 тыс.

В 1878 г. заключена Всемирная почтовая конвенция, регулирующая обмен корреспонденцией между 22 развитыми странами мира.

Почтовые службы связи начала прошлого века.

пользоваться спросом, потому что за считанные секунды можно будет передать информацию практически в любую точку мира.

ПЕРЕДАЧА ИНФОРМАЦИИ МЕЖДУ КОМПЬЮТЕРАМИ

Публичная демонстрация возможностей сети Арпанет (прообраза Интернета) в октябре 1972 г. вызвала особый интерес к методу коммутации пакетов, и это способствовало его дальнейшему развитию.

Современный прогресс человечества связан в первую очередь с глобальной информатизацией всего мирового сообщества. За период с1970 г. построены сотни национальных и международных компьютерных сетей. Благодаря этому в большинстве стран обеспечивается повсеместное внедрение информационных технологий и доступ к Интернету.

Переход к межкомпьютерному взаимодействию *(компьютерным сетям)* характеризует качественный скачок в передаче информации, обеспечивающий удалённое взаимодействие компьютеров и совместное использование ресурсов компьютера всеми пользователями сети.

Для передачи данных компьютеры применяют самые разнообразные физические *каналы.* Это и традиционный электрический кабель (несколько металлических проводников, иногда экранированных); и радиосвязь, через ретрансляторы, спутники связи; и инфракрасные лучи (как в телевизионных пультах дистанционного управления); и современный оптоволоконный кабель, по которому передаются световые сигналы, генерируемые микролазером; и обычная телефонная сеть. По этим каналам пересылают целые пакеты информации.

Оборудование сети Механико-математического факультета МГУ.

Независимо от того, как соединены между собой компьютеры — с помощью проводов, радиоканала, оптоволоконного кабеля, передача информации на большие расстояния осуществляется не напрямую, а через группы компьютеров или других вспомогательных устройств.

В технологии передачи данных используют два принципиально разных метода — *коммутации каналов* и *коммутации пакетов*.

При первом методе сначала устанавливается весь путь соединения — от отправителя до получателя. После этого возвращённый пакет извещает: можно начать передачу данных, все каналы пути готовы. По завершении передачи указанный путь может быть разорван.

Достоинство метода состоит в простоте алгоритма и возможности обмениваться как данными, так и речью в цифровом виде.

При втором методе сообщение первоначально делится на меньшие пакеты, которые нумеруют, снабжают адресами (от кого и кому), и они сами прокладывают себе путь по сети независимыми маршрутами. Сеть при этом также испытывает более равномерную нагрузку. Данный метод обеспечивает бо́льшую надёжность доставки информации (при разрыве сети в пути потеряются не все пакеты). Однако реализация метода коммутации пакетов связана со значительными затратами, особенно в *узлах коммутации* (в них пакет может изменить маршрут, выбрав но-

вый), на обработку и хранение каждого пакета информации. Правда, это компенсируется удобством передачи информации между компьютерами, подключёнными к разным по скорости каналам.

ПРОТОКОЛ ПЕРЕДАЧИ

Широкое распространение различных моделей компьютеров в 70-х гг. XX в. породило не только проблему совместимости компьютеров по программному обеспечению, но и проблему их взаимодействия (передачи информации).

Решение нашлось — в середине 70-х гг. была создана «Эталонная модель взаимодействия открытых систем» (OSI). Свойство открытости обеспечивает взаимодействие разнотипных компьютеров между собой при точном выполнении в программном обеспечении компьютера универсальных соглашений, правил, называемых *протоколами*.

Модель OSI разработана Международной организацией по стандартам (International Standardization Organization — ISO), регламентирующей компьютерное оборудование и передачу информации по сети. OSI представляет собой семиуровневую сетевую модель программного и аппаратного обеспечения.

Уровень 1 — *физический* описывает, как получаются пакеты данных от высшего, канального, уровня, как их преобразуют в оптические или электрические сигналы, в поток нулей и единиц. Здесь же находятся сведения о физической среде: типах кабелей и разъёмах, разводке контактов в разъёмах, схеме кодирования сигналов и об известном стандарте IEEE 802.3 — Ethernet.

Уровень 2 — *канальный* обеспечивает создание, передачу и приём групп данных при прямом соединении по каналу связи. Он обрабатывает запросы следующего, сетевого, уровня, используя сервис физического уровня для приёма и передачи пакетов.

Уровень 3 — *сетевой* обеспечивает выбор маршрута перемещения пакетов через сеть.

Institute of Electrical and Electronics Engineers (IEEE) — профессиональная организация (США), определяющая стандарты, связанные с сетями и другими аспектами электронных коммуникаций. Группа IEEE 802.X содержит описание сетевых спецификаций, а также стандарты, рекомендациии и нформационные документы для сетей и телекоммуникаций.

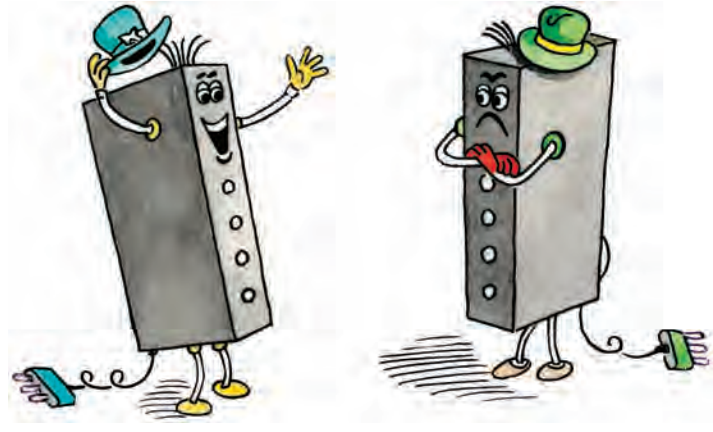
Уровень 4 — *транспортный* реализует процедуры установления соединения, поддержания и разрыва (завершения) соединения.

Уровень 5 — *сеансовый* отвечает за проведение сеансов обмена данными между машинами, в том числе за продолжение обмена после длительного сбоя связи.

Уровень 6 — *уровень представления* обеспечивает преобразование данных: кодирование, компрессию и т. п.

Уровень 7 — *прикладной* предоставляет доступ программных приложений в сеть: перенос файлов (FTP — протокол переноса файлов), обмен почтовыми сообщениями (X.400 — электронная почта, SMTP — простой протокол почтового обмена), пользовательский доступ к удалённому компьютеру (Telnet — протокол обмена с терминалом) и пр.

Уровни протоколов надстроены один над другим. Более высокий уровень использует функции более низкого и предоставляет свои возможности следующему уровню в схеме. Поэтому иногда уровни протоколов передачи данных называют *стеком* протоколов.

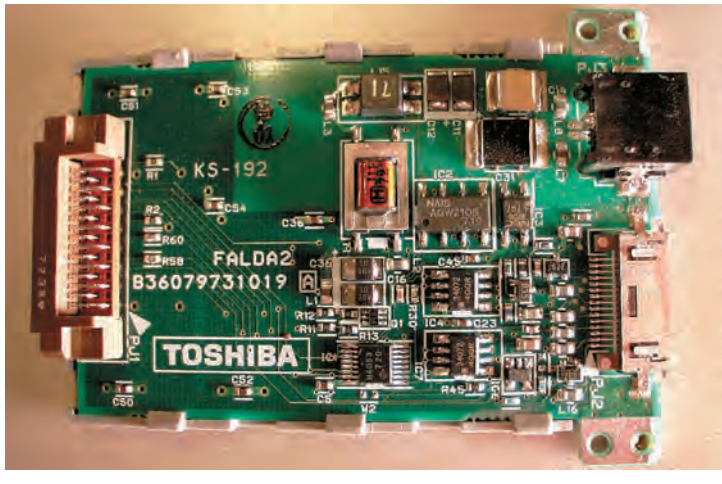


МОДЕМ

Тенденция развития современных средств телекоммуникации — представлять аналоговую информацию (звук, изображение) в цифровом виде; так появились цифровые CD, цифровая фотография и т. п.

Существующие телефонные линии не приспособлены для передачи цифровой информации. Приходится сначала преобразовывать информацию из цифровой в аналоговую форму, затем передавать её в таком виде по телефонной линии и на другом конце линии выполнять обратное преобразование. Этим и занимается *модем,* объединяющий в себе два логических устройства: *модулятор,* т. е. преобразователь из цифровой в аналоговую форму, и *демодулятор* — обратный преобразователь.

Современные модемы теоретически могут обеспечивать скорость передачи порядка 50 кбит/с. Правда, эксперименты на реальных телефонных сетях показывают, что такая скорость достижима очень редко и стабильная связь, особенно в России, осуществима обычно на скорости в два раза ниже. При более высокой скорости соединения происходят многочисленные ошибки и бесконечная повторная передача сбойных пакетов, так что реальная скорость передачи резко падает либо связь вообще рвётся. Причина этого кроется в устройстве телефонной сети.



Встроенный модем ноутбука.

ПОДКЛЮЧЕНИЕ К ВСЕМИРНОЙ СЕТИ: ПРОТОКОЛЫ UUCP И PPP

Задолго до появления Всемирной паутины люди пользовались модемами для общения друг с другом. Прежде всего, модем позволял работать с электронной почтой. Для этого применялся протокол UUCP.

Это протокол передачи данных между двумя компьютерами, который был разработан в системе UNIX. UUCP в основном используется для перекачки электронной почты с компьютера, подключённого к Интернету, на домашний компьютер пользователя и обратно. На персональных компьютерах аналогично работает программа UUPC (протокол для PC), которая с помощью модема звонит по телефону основному компьютеру, выступающему в роли почтового сервера, устанавливает UUCP-соединение и переносит электронную почту на персональный компьютер и обратно. После этого соединение разрывается, так что протокол UUCP занимает телефонную линию совсем ненадолго. При чтении и подготовке почты телефонная линия остаётся свободной.

UUCP-почта довольно удобна и не требует значительных ресурсов, позволяя сотням пользователей работать с почтой, не мешая друг другу, при всего одной телефонной линии. Тем не менее в последнее время протокол UUCP применяется всё реже и реже. Связано это с возросшей популярностью Всемирной паутины. Современные провайдеры обеспечива-

ют непосредственное подключение через модем к Интернету, что автоматически даёт возможность пользоваться электронной почтой и всеми остальными услугами, существующими в Интернете.

Для подключения к Интернету через модем имеется два протокола: SLIP и PPP. В начале XXI в. протокол SLIP устарел, и почти все провайдеры поддерживают протокол PPP.

PPP позволяет на время соединения с сервером через модем сделать персональный компьютер полноценным узлом сети: такому компьютеру даётся Интернет-адрес, он может принимать и посылать пакеты в формате используемого в сети Интернет-протокола. Недостаток один: занята телефонная линия. Платить, как правило, приходится за использованное время, а не за объём переданной информации.

Скорость и надёжность передачи информации по модему невысоки и совершенно неудовлетворительны для современных компьютерных сетей. Тем не менее огромным достоинством модема является то, что он разрешает использовать уже существующие телефонные линии, не дожидаясь создания инфраструктуры компьютерных сетей.

Возможно, «настоящие» компьютерные сети скоро станут таким же обычным явлением, как телефонные сети. Уже сейчас в крупных городах многие дома подключены к Интернету с помощью современных средств, таких, как оптоволоконные линии. Когда появятся подобные сети, модем наконец займёт своё место в ряду устаревших и достаточно курьёзных технических устройств — именно таким нам представляется сейчас устройство ввода с перфокарт, которое ещё 25 лет назад было неотъемлемой частью любой ЭВМ.

ISDN И МОБИЛЬНЫЕ ТЕЛЕФОННЫЕ СЕТИ

Чтобы избавиться от недостатков существующих телефонных линий, был разработан новый стандарт телефонных сетей — ISDN, т. е. интегрированные сервисные цифровые сети (Integrated Service Digital Networks). ISDN позволяют передавать как обычный звук, так и компьютерные данные в цифровом виде; скорость передачи равна 64 кбит/с. Во время разговора звук преобразуется в цифровую форму и в таком виде передаётся по линии, т. е. происходит преобразование, обратное тому, которое выполняет модем. К сожалению, ISDN в сфере телекоммуникаций были встречены весьма прохладно. Во-первых, для их развития требуются крупные финансовые вложения, но большинство пользователей не идут на это, работая по старинке, с помощью обычного модема.

Во-вторых, скорость передачи, которую обеспечивают ISDN, недостаточна для современных компьютеров. Она лишь незначительно превышает скорость работы современного модема в качественной телефонной сети (хотя, конечно, надёжность ISDN неизмеримо выше).

На начало XXI в. в России ISDN практически отсутствовали. Однако сотовые телефонные сети быстро распространяются. Протоколы, применяемые в мобильных сетях, весьма удобны для цифровой передачи данных между компьютерами. Поэтому подключение к Интернету через сотовый телефон становится доступнее и дешевле.

Модемы ноутбуков величиной с кредитную карточку.

КАК РАБОТАЕТ МОДЕМ

При установке соединения два модема автоматически «договариваются» между собой о максимально возможной для обоих скорости передачи и выборе *коммуникационного протокола*. Более быстрый модем может связаться с медленным, применив уменьшенную скорость передачи.

Один хакер, будучи в гостях у друга- программиста, видит любимца дома — огромного рыжего кота. И спрашивает приятеля:
— Как зовут котяру?
— Зиксель. (Zyxel — известная фирма, выпускающая модемы. — Прим.ред.)
— Опа-на! — восхищается гость. А почему?
— Смотри, — отвечает хозяин, — ATDP!..
— Шшшшшш! — шипит кот.
— Во! Законнектился! 14 400!

Если динамик модема включён, то человек в этот момент слышит шипение и свист разной частоты. Модем, устанавливающий соединение, предлагает сначала протокол, в котором указывается максимальная скорость передачи, со сжатием и коррекцией ошибок. Если второй модем не может применить этот протокол (например, из-за низкого качества телефонного соединения), то первый модем меняет его на более простой, со сниженной скоростью передачи или вообще без сжатия данных. Таким образом, первый модем перебирает все поддерживаемые им протоколы, пока не найдёт приемлемый для обоих модемов. Впрочем, обычно модему запрещено «опускаться» до самого примитивного протокола без коррекции данных, которые в обычных телефонных сетях приводят к тому, что даже одно короткое слово невозможно передать без ошибки.

Протокол описывает способы передачи байтов по телефонной линии: тип модуляции аналогового сигнала, несущую частоту, дополнительные служебные биты, а также как байты объединяются в пакеты, форматы пакетов, порядок переповтора передачи в случае искажения пакета и т. п. Каждый пакет содержит кроме самих передаваемых байтов дополнительную информацию, включающую, в частности, контрольную сумму. Контрольная сумма позволяет принимающему устройству проверить, не исказился ли пакет при передаче, и при необходимости запросить повторную передачу.

Модем может использовать коды, исправляющие ошибки, а также сжатие данных. Идея кодов состоит в том, что к каждому байту добавляются дополнительные биты, дающие возможность при искажении одного бита однозначно восстановить исходный байт. И хотя длина сообщения при этом увеличивается, в большинстве случаев не приходится повторять передачу искажённых пакетов. Это особенно важно при неустойчивой связи, а телефонные линии крайне ненадёжны.

Сжатие данных применяется для упаковки содержимого сообщений, что в ряде случаев существенно уменьшает их размер. Правда, при передаче уже упакованных данных, таких, как картинки в формате JPEG или ZIP-файл, дополнительное сжатие практически ничего не даёт. Сжатие данных может использоваться совместно с кодами, исправляющими ошибки: сначала данные упаковываются, а затем каждый байт упакованного сообщения передаётся с помощью кода, исправляющего ошибки.

В настоящее время существует множество коммуникационных протоколов, они имеют свои обозначения: MNP 4, MNP 5, V42b, V.90 и т. п. Если самые первые модемы передавали лишь отдельные байты, не используя никаких протоколов высокого уровня, то со временем скорость передачи менялась от нескольких сотен до 56 тыс. бит/с.

ПРОГРАММИРОВАНИЕ МОДЕМА

Информация передаётся модемом в виде последовательности байтов (групп по 8 бит). К каждому байту добавляются один или несколько служебных битов (старт-бит, стоп-бит). Байты могут посылаться по отдельности либо объединяться в пакеты, причём для каждого пакета сообщается контрольная сумма. При несовпадении её передача пакета повторяется.

Большинство современных модемов относится к Hayes-совместимым (по названию фирмы Hayes). Устройство такого модема в упрощённом виде выглядит следующим образом. В компьютере имеется коммуникационный порт, через который можно передавать байты в модем и принимать байты от модема. При этом передаваемые и принимаемые байты образуют два независимых потока, что позволяет в любой момент пересылать байты через порт и считывать их из порта.

Модем может находиться в одном из двух режимов — в *режиме команд* или в *режиме данных*. В режиме команд модем интерпретирует отправленные ему через порт байты как команды. Эти команды либо меняют внутреннее состояние модема (т. е. содержимое его внутренних регистров), либо заставляют модем выполнить некоторое действие. Любая команда начинается с двух символов AT (от *англ.* attention — «внимание») и заканчивается символом перевода строки. Например:

- ATL3 — «установи максимальную громкость встроенного динамика»;
- AT S0 = 0 — «запрещено отвечать на приходящие звонки»;
- AT S0 = 2 — «сними трубку после двух звонков»;
- ATZ — «инициализация модема».

Таких AT-команд несколько десятков, с их помощью происходит общение компьютера с модемом. Наиболее распространённая команда — это команда набора телефонного номера:

AT DP 839-17-86,

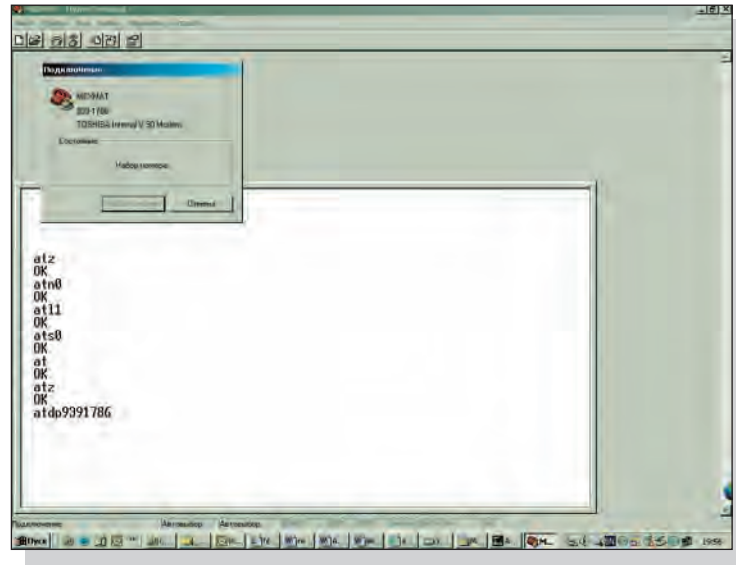
здесь D (от *англ.* Dial) — «набирать номер»; P (от *англ.* Pulse) — «импульс». В России принят импульсный набор номера, а при тональном наборе используется префикс DT.

В режиме команд модем сообщает компьютеру результат выполнения каждой команды. Коды результата передаются от модема компьютеру точно так же, как и данные, принимаемые от удалённого модема.

В режиме данных модем пересылает все байты, передаваемые ему через коммуникационный порт, удалённому модему.

В любой операционной системе имеется специальная программа «терминал», позволяющая пользователю напрямую общаться с модемом, например, с помощью AT-команд. В системе MS Windows такая программа называется Hypertrm. После установки соединения с удалённым модемом все символы, набираемые на клавиатуре, пересылаются удалённому компьютеру; обратно, все символы, выводимые удалённым компьютером на экран, пересылаются через модем терминальной программе и отображаются в её окне.

Так, терминальная программа даёт возможность превратить домашний компьютер в терминал (экран + клавиатура) и работать на удалённом компьютере, не выходя из дома. Через неё можно даже переслать файл, используя специальный протокол



Скриншот терминальной программы.

Z-Modem. При этом протокол Z-Modem запускается на удалённом компьютере в режиме приёма, а терминальная программа передаёт содержимое файла байт за байтом так же, как и коды клавиш при наборе команд.

КАКИЕ БЫВАЮТ КОМПЬЮТЕРНЫЕ СЕТИ

Сети можно разделить на классы в соответствии с величиной. Небольшие сети, в пределах одного здания, объединяющие от 2 до 300 компьютеров, которые принадлежат обычно одной организации (или одной семье), называются *локальными вычислительными сетями*. Сети между учреждениями в пределах города, связывающие много локальных вычислительных сетей, называются *городскими* — MAN (Metropolitan Area Network). *Глобальные сети* — WAN (Worldwide Area Network) объединя-

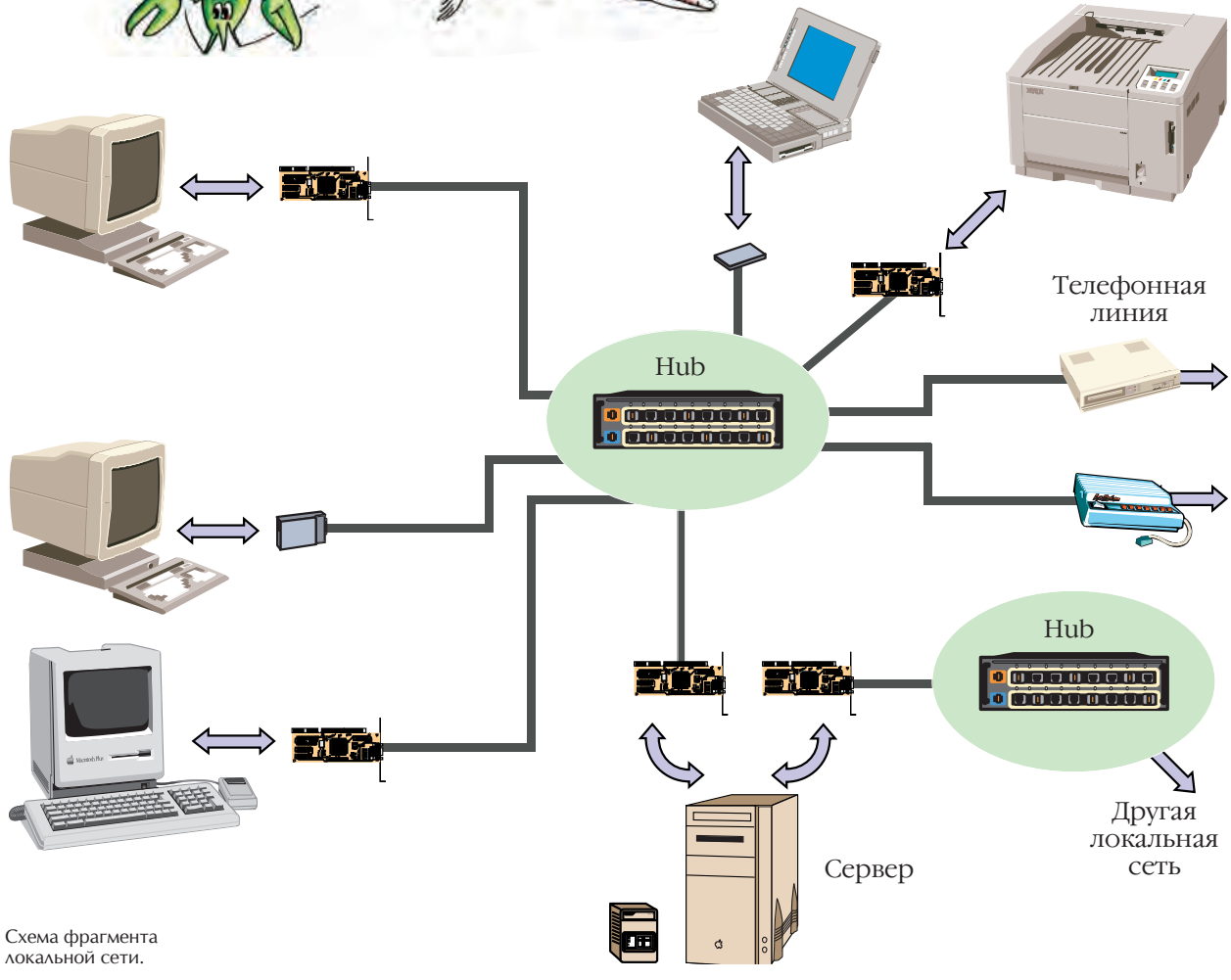


Схема фрагмента локальной сети.

ют сотни, тысячи узлов во многих странах мира.

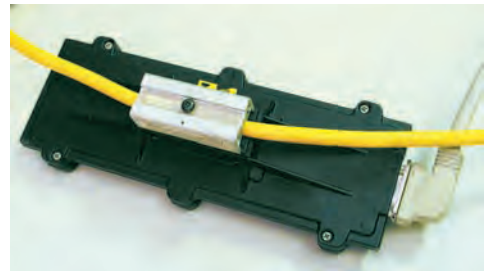
Локальная сеть создаётся для рационального использования компьютерного оборудования и эффективной работы сотрудников. В настоящее время трудно представить себе фирму или даже квартиру, где при наличии хотя бы двух компьютеров они не были бы соединены в сеть. Сеть позволяет пересылать файл с одной машины на другую, хранить совместный архив (как правило, компьютеры неравноценны, и у какого-то из них дисковое пространство больше) и делать распечатки.

В офисе обычно устанавливаются один сервер (для печати и хранения данных) и рабочие станции для сотрудников, один-два модема для выхода в Интернет (или прямое кабельное соединение), для получения и отправки электронной почты, факсов и электронных платежей, несколько сетевых принтеров, внутренняя АТС на десятки телефонных номеров. Организовывать переписку сотрудников внутри локальной сети вполне разумно — это, по крайней мере, упрощает документооборот. Дома ограничиваются настольным сервером с большими дисками, с принтером и сканером и одной-двумя машинами «послабее» (возможно даже, это ноутбуки, с которыми хозяева ходят на работу и частенько приносят работу на дом). Наличие пишущих CD ROMов при передаче достаточно большого объёма информации является малым подспорьем, так как современные ноутбуки «облегчаются» за счёт отсутствия всякой периферии, а гнездо для подключения витой пары есть практически везде. (В качестве среды передачи в локальных сетях преимущественно употребляется *неэкранированная витая пара* — похожая на обычные провода, свитые парами.)

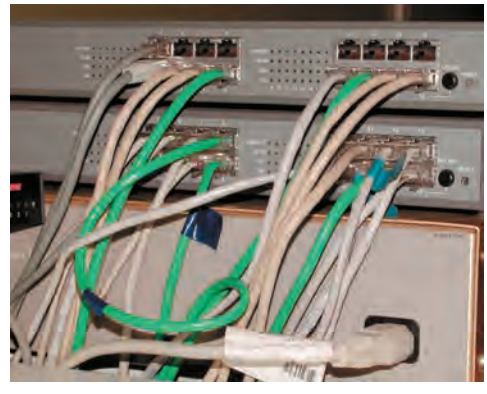
Применение наиболее популярного однотипного оборудования с однотипными протоколами для связывания компьютеров на основе технологии Ethernet упрощает интеграцию в сеть новых машин. Хотя Ethernet появилась в начале 80-х гг. XX столетия, однако широко применять данную технологию стали только в конце 80-х гг., когда Комитет IEEE 802.3 (Международная организация по стандартизации) разработал стандарт 10BaseT. Данный стандарт изменил саму природу локальных сетей: стала использоваться неэкранированная витая пара.

Раньше между всеми машинами, связанными в локальную сеть, прокладывался единый коаксиальный провод (очень похожий на антенный телевизионный). Если он обрывался, то из строя выходила вся сеть. А когда в результате аварии на провод попадало высокое напряжение (хотя бы 220 В), испорченными оказывались все компьютеры, находящиеся в сети. Такой провод называли *тонкий Ethernet*, в отличие от *толстого Ethernet*, который

Hub и тонкий Ethernet.



Толстый Ethernet.



Hub и витая пара.

Хотя наиболее широко распространена именно технология Ethernet, стоит упомянуть и другие протоколы (и оборудование), например Token ring.

может использоваться для выхода в городскую информационную сеть.

MAN и WAN — территориально распределённые сети, в них удалённые точки (локальные сети) объединяются с помощью скоростных оптоволоконных каналов связи. Волоконно-оптические кабели сегодня внедряются повсюду, уверенно вытесняя своих медных предшественников.

Стандарт 10BaseT определил использование технологии «звезда», когда любой компьютер подключался к прибору, называемому Hub, который позволял присоединить 8, 16 или 32 компьютера одновременно. Получалось, что Hub — центр звезды, а компьютеры — её лучи. Действия Hub напоминали работу телефонистки, он осуществлял коммутацию машин для передачи пакетов между ними. Более сложные Hub могли одновременно осуществлять передачу сразу между несколькими парами машин.

Если машины рассредоточены по зданию, то нерационально использовать один Hub (даже если машин в сети мало), так как прокладка проводов иногда весьма трудоёмка. Проблему легко решить, разместив по Hub в каждой комнате и связав стоящие в ней машины, а все Hub объединить затем с главным центральным Hub. Для дома, когда надо связать всего две машины, Hub не нужен, достаточно специального провода.

При описании сети используется термин «топология сети». Топология описывает физическое размещение компьютеров, кабелей, Hub и других компонентов локальной сети. Топология сети в значительной степени влияет на такие характеристики, как производительность, стандарт оборудования, стоимость монтажа, затраты на её поддержку, ремонтопригодность, надёжность и т. п.

Компьютерные сети различаются по типу: *одноранговые* и типа «*клиент—сервер*». *Одноранговая* сеть построена на равноправных компьютерах, каждый из них может использовать ресурсы другого. В сетях с большим количеством пользователей нежелательно, чтобы все пользователи получали доступ ко всем ком-

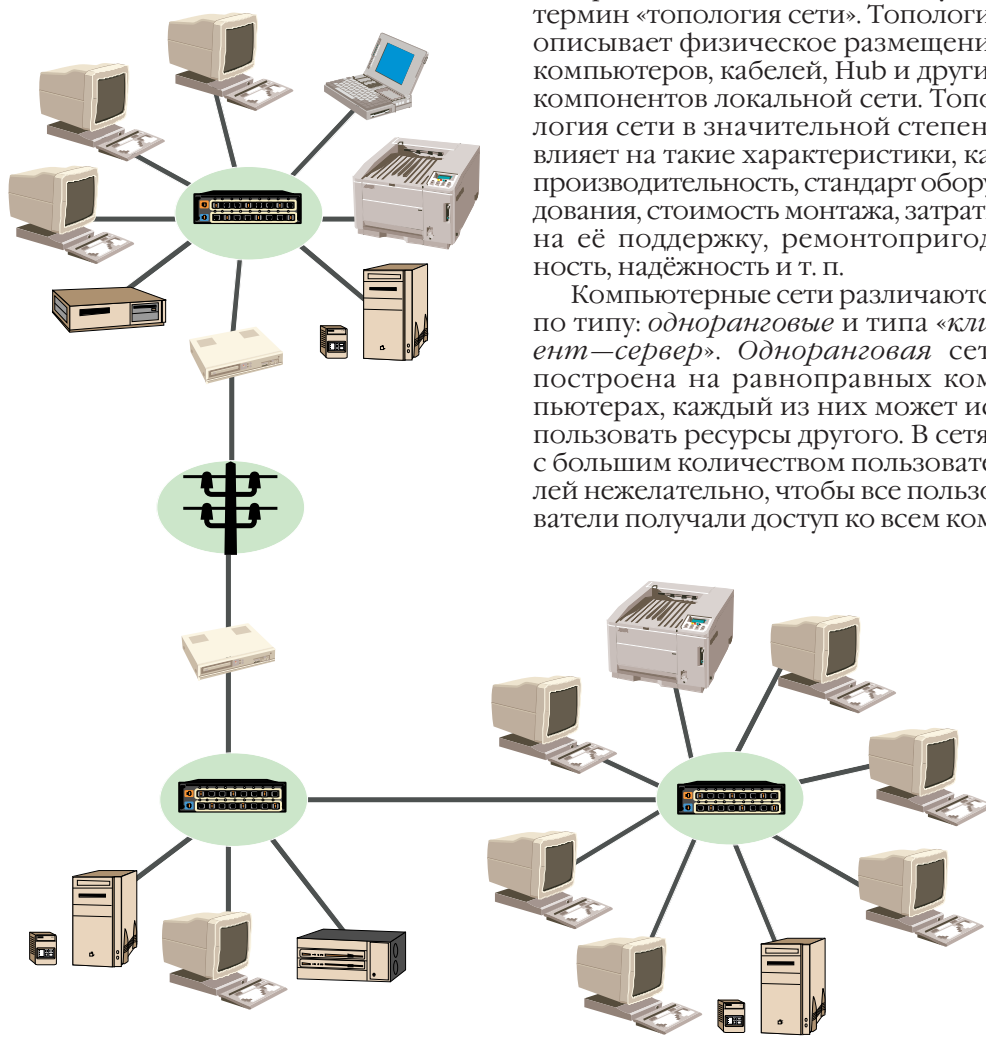


Схема фрагмента распределённой сети (WAN, MAN).

пьютерам сети. Именно поэтому одноранговые сети больше подходят для небольших групп, работающих над одним проектом.

Более популярны сети типа «клиент — сервер». При разделении программ на *клиентскую* и *серверную* части удаётся лучше использовать производительность настольных компьютеров, которые нерационально применять в качестве простого терминала. Такой подход даёт возможность перенести приложения с главных компьютеров — *мэйнфреймов* (*англ.* mainframe) в системы, основанные на локальных сетях. Программа, исполняемая на машине пользователя, меньше нагружает сеть передачей данных. Так, программы-навигаторы Интернета, например Internet Explorer, не беспокоят сеть, пока человек разглядывает страницы, загруженные из сети.

СЕТЕВАЯ ОС

Когда стоимость винчестеров была велика и устанавливать их на каждый компьютер было нерентабельно, использовались сетевые операционные системы. Подобные системы умели соединить несколько персональных компьютеров, обеспечив совместное использование небольшого количества данных и основных приложений. Загрузка операционной системы осуществлялась с центральной машины. Сетевые ОС напоминали центры коллективного пользования, когда терминалы подключены к одной машине. В качестве терминалов выступали персональные компьютеры, да и «интеллект» терминалов был существенно выше, потому что программы выполнялись не на центральной машине, а на компьютере пользователя. Наиболее успешно сетевые ОС применялись при обучении в университетах, так как, с одной стороны, студенту не требуется особенно больших компьютерных ресурсов, с другой — сетевой ОС удобно управлять. Сегодня сетевые операционные системы должны делать всё: работать со старыми файловыми системами и одновременно обрабатывать приложения «клиент — сервер». Самой распространённой сетевой ОС является Novell Netware.

ТЕХНОЛОГИЯ БЕСПРОВОДНОЙ СВЯЗИ

ПЕРЕДАЧА ИНФОРМАЦИИ БЕЗ ПРОВОДОВ. РАДИО

До 1920 г. в радиосвязи применялись волны длиной от сотен метров до десятков километров. В 1922 г. радиолюбители открыли свойство коротких (несколько десятков метров) волн, распространявшихся на любые расстояния благодаря отражению от верхних слоев атмосферы. Такие волны стали основным средством дальней радиосвязи.

В 30-х гг. XX в. были освоены метровые, а в 40-х гг. — дециметровые и сантиметровые волны, распространяющиеся в зоне прямой видимости; радиус действия связи на этих волнах ограничивался 40—50 км. Однако ширина частотного диапазона коротких волн (от 30 МГц до 30 ГГц) в 1 тыс. раз больше ширины частотных диапазонов волн длиннее 10 м. То есть на этих частотах можно передавать огромные потоки информации и организовывать многоканальную связь.

Диполь Герца (вибратор Герца), — простейшая антенна, которую использовал Генрих Герц в 1888 г. в своих опытах. Она состояла из медного стержня с металлическими шарами, в разрыв которого включалась так называемая катушка Румкорфа. Электрические колебания в диполе возбуждались с помощью единственного известного в то время источника электрических колебаний - искрового разряда. Поэтому передатчики, основанные на этом методе, получили название искровых.

НА ПУТИ К РАДИО

Ещё в 80-х гг. XIX в. американский изобретатель Томас Эдисон проводил опыты с радиосвязью и даже получил патент. В 1886—1889 гг. немецкий физик Генрих Герц экспериментально доказал существование электромагнитных волн и создал искровый излучатель электромагнитных волн. В 1895 г. русский физик и электротехник Александр Степанович Попов изобрёл электрическую связь без проводов (радиосвязь). 7 мая 1895 г. Попов продемонстрировал первый в мире радиоприёмник. В 1897 г. итальянский радиотехник Гульельмо Маркони сконструировал аналогичный аппарат. В том же году он зарегистрировал Компанию беспроводного телеграфирования и сигнализации в Англии, а в 1899 г. — в США. В декабре 1901 г. Маркони осуществил радиотелеграфную передачу через Атлантический океан.

Чтобы одна радиопередача не мешала остальным, требовались учёт и распределение частот радиовещания. По решению радиоконференции 1927 г. запрещалось применение искровых радиопередатчиков, которые вещали в широкой полосе частот (их оставили только для передачи сигналов SOS, чтобы увеличить вероятность приёма).

Малая дальность распространения и узкая направленность сигнала позволяют использовать одни и те же длины волн в близко расположенных пунктах без взаимных помех.

Передача на значительные расстояния осуществляется путём многократной ретрансляции в так называемых линиях радиорелейной связи или с помощью спутников, находящихся на высоте около 40 тыс. км над Землёй.

В 1953 г. была разработана аппаратура радиорелейной связи «Стрела-М», которая передавала 24 телефонных разговора одновременно, а несколькими годами позже её сменила радиорелейная станция Р-60/120, обеспечившая до 120 телефонных каналов.

Сотовые телефоны CDMA и GSM.

ЦИФРОВАЯ РАДИОСВЯЗЬ

На принципах радиосвязи основываются и методы передачи голоса (и любой другой информации) в сотовой телефонной связи.

Как разделить общий широкий радиоканал при передаче информации? Можно поступить очень просто: используемый диапазон разбить на равные части так, чтобы в выделенной полосе помещался весь спектр передаваемой информации. Если это голос телефонного абонента, то каждому абоненту достаточно 10 кГц, превышающие спектр человеческой речи. Чтобы передача разговоров одного абонента не мешала остальным, нужно оставить солидный запас.

В первых сетях радиосвязи использовался шаг 50 кГц, а затем его удалось уменьшить до 25 кГц. В аналоговых системах сотовой связи применён множественный доступ с разделением частот, или частотное разделение каналов — FDMA (англ. Frequency Division Multiple Access). Технически возможно уменьшить шаг до 6—8 кГц, правда с использованием цифровых методов сжатия звука. Уменьшать до бесконечности данную величину нереально, потому что из-за близости частот возникают сильные искажения. Для FDMA предел уже достигнут.

Все остальные методы относятся к цифровой радиосвязи, которая более устойчива к помехам, её проще кодировать, шифровать. Голос абонента даже проходит дискретизацию, преобразовываясь в цифровое представление, и в таком виде передаётся в эфир.

Наиболее популярный метод цифровой радиосвязи — *временное разделение каналов*, TDMA (*англ.* Time Division Multiple Access). В общей широкой полосе частот, выделенной для радиосети, применить этот метод в чистом виде не удаётся. Например, в сотовой сети GSM, самой популярной в Европе, используют комбинированное частотное и временное разделение каналов (FDMA+TDMA). Сначала методом FDMA общую полосу 25 МГц делят на каналы по 200 кГц, а уже затем групповой канал — на восемь пользовательских каналов методом

TDMA. Для каждого разговора (эквивалентного ширине 25 кГц) выделяется маленький квант времени, в течение которого по каналу передаётся его информация, в следующий момент — информация другого разговора и так далее, по кругу. Причем один и тот же разговор может не обязательно передаваться по одному каналу. На станции полученные куски «разбирают» по абонентам и «склеивают» в непрерывный разговор.

Можно ожидать, что в скором времени затраты на полосу на одного абонента снизятся до 4 кГц. В американском стандарте IS-54 с похожими методами кодирования затраты на полосу оказываются в три раза ниже, чем в GSM.

Существуют ещё системы цифровой радиосвязи с кодовым разделением каналов, или CDMA (*англ.* Code Division Multiple Access). Каждый канал занимает всю выделенную полосу частот и поэтому создаёт помеху для остальных. Казалось бы, это лишь усугубляет проблему. Однако представим себе ресторан, в котором за столиками сидят люди, одновременно говорящие на разных языках. Чужая речь при этом мало мешает отдельному разговору, так как беседующие из общего шума выделяют только понятные им слова родного языка. Примерно на тех же принципах основаны и сети CDMA.

В 1935 г. аспирант Ленинградского электротехнического института Дмитрий Васильевич Агеев написал статью «Основы теории линейной селекции». В ней он показал, что два радиосигнала, занимающие одну полосу радиочастот и совпадающие по времени, могут быть разделены и приняты приёмником, если в приёмнике заложена информация о форме передаваемого сигнала.

Идея очень понравилась военным, так как при этой технологии передача информации лучше всего защищена от прослушивания. Например, американцы использовали кодовое разделение сигналов для радиосвязи в военной авиации. Впервые такие возможности были продемонстрированы в 1989 г. американской компанией Qualcomm, которая и разработала стандарт CDMA для сотовой связи.

Хотя создаётся впечатление, что общая потенциальная пропускная способность радиосети CDMA снижается, на самом деле реальная эффективность систем CDMA оказывается даже выше, чем у TDMA. В CDMA нет потерь на разделение диапазона, когда надо сохранять некоторый запас между близко расположенными каналами. При специальных методах *помехоустойчивого кодирования* наряду с улучшением качества связи попутно осуществляется перераспределение общего ресурса полосы между разговорами в пользу более активно говорящего. Правда, приёмники CDMA значительно сложнее аналогичных приёмников TDMA.

СОТОВЫЙ ИНТЕРНЕТ

С помощью сотовой связи можно не только разговаривать, но и пользоваться Интернетом. К сотовому телефону можно подключить модем, но необходимо учитывать, что модем использует аналоговые методы передачи цифрового сигнала. А сотовая связь, как правило, цифровая. Кроме того, сотовая связь предназначена для передачи голоса или нескоростной цифровой информации: факса, SMS — коротких текстовых сообщений. Реальные скорости передачи данных по модему не превышают 10 кбит/с. Конечно, это не совсем удовлетворяет запросы человека, часто пользующегося Интернетом. На таких скоростях можно только получать электронную почту, да и то малого объёма.

В сети GSM был найден выход — система пакетной передачи данных по сети сотовой связи, или GPRS (*англ.* General Packet Radio Service). При этом

Интернет доступен с сотового телефона.

Д. В. Агеев известен также как основатель радиотехнического факультета Горьковского политехнического института, одной из самых сильных в мире школ радиоэлектроники.

Д.В.Агеев.

▶ Устройство безпроводной связи Bluetooth.

Стандарт CDMA-2000 предоставляет высокоскоростной метод передачи данных (в том числе доступ в Интернет) со скоростью порядка 150 кбит/с.

Синезубый — прозвище датского короля Харальда (около 890 — около 940), полученное им за тёмный передний зуб. Харальд вошёл в историю как великий завоеватель и человек, принесший Дании христианство.

методе информация собирается в пакеты и передаётся в эфир, заполняя «пустоты» между передачей голоса. Если при обычной передаче данных с помощью модема сначала устанавливается соединение и лишь после этого начинается передача данных, то при пакетной передаче соединение есть всегда, пока доступна сотовая сеть, ведь телефон включён.

При пакетной передаче данных радиоканал занят только в процессе передачи информации, пока по нему идёт пакет. Поэтому оплата взимается за так называемый *трафик*, т. е. за объём переданной и полученной информации, а не за время соединения.

GPRS может обеспечивать скорость передачи данных до 107 кбит/с и даже выше. Это вполне приемлемо и даже превышает аналогичные возможности обычной телефонной линии при использовании модема. GPRS так называемого первого поколения позволяет передавать информацию со скоростью до 13,4 кбит/с и принимать со скоростью до 40,2 кбит/сек.

БЕЗ ПРОВОДОВ

В начале 1998 г. пять компаний — Ericsson, Nokia , IBM, Intel и Toshiba — сформировали рабочую группу SIG

(Special Interest Group) для разработки нового стандарта беспроводной связи, названного Bluetooth, что в переводе с английского означает «синезубый».

Стандарт Bluetooth использует частоты 2,4 — 2,48 ГГц в так называемом диапазоне ISM (*англ.* Industry, Science and Medicine — «промышленный, научный и медицинский»), для этого не требуется специальной лицензии, в отличие от сотовой связи. Скорость передачи данных может достигать 720 кбит/с. Стандарт предполагал поддерживать связь на расстоянии не более 10 м, однако в первые годы XXI в. были выпущены микросхемы, позволившие увеличить расстояние до 100 м.

Излучение такой частоты способно обходить препятствия, поэтому устройства не обязательно должны находиться в зоне прямой видимости. Как только Bluetooth-устройства оказываются в пределах досягаемости, сразу происходит соединение, причём стандарт допускает широковещание.

То, что ISM не лицензируется, является не только достоинством, но и недостатком. В данном диапазоне работают также различные медицинские приборы, бытовая техника и прочее, что может привести к конфликту между ними. Во избежание этого Bluetooth действует по принципу скачкообразной перестройки частоты (1600 скачков в секунду). Новая частота выбирается генератором случайных чисел, что позволяет освободить нужные другим устройствам частоты.

Благодаря технологии Bluetooth абонент сотовой связи может использовать беспроводный наушник с микрофоном и разговаривать, пока телефон с таким же модулем лежит где-то в комнате. Предполагается, что Bluetooth будет интегрирован в мобильные телефоны, компьютеры и другие устройства, применяющие беспроводную связь.

ПО СЕКРЕТУ ВСЕМУ СВЕТУ...

Безпроводные технологии также имеют ахиллесову пяту. Проблема лежит на поверхности: раз информация передается без проводов, то соответственно она не защищена от прослушивания. Правда разработчики пытаются решить и эту проблему — методами шифрования передаваемой информации. Казалось бы существующая техника шифровки сообщений избавила потребителя от «лишних ушей» (см. статью «Современная криптография»). Однако допущенные ошибки при разработке безпроводных средств связи, обнаруженные хакерами, ставят под удар не только конкретное оборудование и конкретных пользователей (чей разговор подслушан), но и всю отрасль вцелом.

ЧТО С ЧЕМ СОЕДИНЯТЬ

Это традиционный вопрос покупателя персонального компьютера. Если компьютер состоит из двух больших коробок — монитора и системного блока, то, по крайней мере, есть шнур, соединяющий их друг с другом. Для мыши и клавиатуры, если они не используют радио- или инфракрасную связь для передачи, нужна ещё пара проводов. Кроме того, провода к усилителю или колонкам, чтобы машина воспроизводила музыку; провод к телевизору, чтобы смотреть DVD-диски не с монитора, а с экрана. Если нравятся компьютерные игры, то для подключения 3D-очков (или шлема) и руля с педалями также требуются провода. А ещё принтер, сканер, модем (когда он внешний), витая пара для локальной сети... Если же компьютер применяется для оцифровки и обработки звука или видео, то в проводах может запутаться и профессионал.

Конечно, разработчики вычислительной техники задумывались о foolproof («защита от дурака»), чтобы пользователь не испортил компьютер, неверно подключив провода. Поэтому разнотипные соединения компьютера имеют разные по форме разъёмы. Провода надо не только правильно подсоединить, но и аккуратно проложить, чтобы случайно не выдернуть из машины и т. п.

Той же «болезнью» страдают и дорогие музыкальные центры, и видеомагнитофоны. А если нужно подключить DVD-плеер, пару видеомагнитофонов, караоке-центр к одному телевизору? Простое присоединение всех выходов к одному входу не поможет, так как сигналы различных устройств будут мешать друг другу. В этом случае не обойтись без специального коммутационного устройства.

Более совершенно устроены стиральная машина и холодильник, так как у них всего один шнур — сетевой. Но неизвестно, что произойдёт, если стиральную машину подсоединят к домашней информационной сети, а холодильник — к Интернету.

Нельзя ли использовать единственный сетевой (по-русски «сеть» — это и система питания, и информационная связь, в отличие от английского языка, где power — «силовая сеть», а network — «информационная сеть») шнур и для передачи энергии, и как информационный канал, благо сеть уже проложена по дому?

Линия электропередачи — одна из наиболее сложных сред для передачи информации. Изначально электрическая сеть предназначена для передачи электрической энергии низкой частоты (50 Гц), на которую воздействуют сильные помехи от промышленных потребителей электрической энергии. Если использовать электрическую сеть для передачи информации в городских информационных сетях, то могут возникнуть проблемы из-за сильного затухания сигнала и высокого уровня помех. К тому же электропередача не должна ухудшать энергоснабжение, когда закодированная информация накладывается на частоту силовой сети.

Такие системы разработаны и уже существуют, но их повсеместное внедрение невыгодно экономически (по сравнению с беспроводной связью).

Так произошло и с Bluetooth, самой популярной технологией беспроводной связи XXI века.

Некий законопослушный гражданин Великобритании, Олли Уайтхауз в апреле 2004 года продемонстрировал методику взлома Bluetooth-устройств. При процедуре соединения двух таких устройств при помощи вспомогательного оборудования хакер получал возможность перехватить идентификационную информацию, которой обмениваются партнеры. При первом соединении на устройствах требуется набрать одинаковый четырехсимвольный PIN-код. Использую специальные Bluetooth-алгоритмы, перебрав примерно 10 тысяч вариантов хакер находил тот самый PIN-код. Далее, так PIN-код используется для генерации специального 128-битного защищенного ключа, фактически получал в распоряжения этот самый «защищенный» ключ. Таким образом, весь последующий зашифрованный «разговор» этих двух устройств для хакера становился открытой книгой.

Однако, это не принесло особого вреда индустрии Bluetooth связи, так как в случае, если первый «контакт» был установлен вне пределов досягаемости хакера, дальнейший обмен данными между устройствами риска не представлял.

Но рано успокоились разработчики Bluetooth. Весной 2005 года два гражданина Израиля Авишай Вул и Янив Шакед из университета Тель-Авива предложили новый, усовершенствованный и «красивый» метод взлома популярного беспроводного протокола. «Предложенный нами метод атаки позволяет взломать любой сеанс обмена данными между двумя Bluetooth-устройствами, даже если это не первая связь между ними», - пояснил Янив Шакед. Им удалось искусственно инициировать процедуру первого соединения Bluetooth-устройств. Подслушивающее устройство «вмешивалось» в контакт, «выдавая» себя за другое, посылалось «жертве» сообщение о том, что ключ якобы забыт. «Умное» устройство, в соответствии с протоколом аннулировало старый ключ и начинало процедуру первого соединения вновь. Дальше хакер действовал по вышеописанной схеме Олли Уйатхауза.

Для отправки ложного сообщения о «потере» ключа хакеру потребовался только чужой идентификатор, которые направо и налево автоматически рассылают все Bluetooth-устройства в пределах зоны действия.

Процедура вычисления защищенного 128-битного ключа после установления контакта на компьютере класса Pentium IV занимает всего 0,06 секунды.

«Это - катастрофа не только в теории, но и на практике», - поделился впечатлениями Брюс Шнайер, эксперт по системам безопасности из г. Маунтин-Вью, Калифорния.

Фактически абсолютно все Bluetooth-устройства стали уязвимы после этой демонстрации.

Даже малый (порядка десяти метров) радиус действия Bluetooth-устройств не спасет от хакеров. С помощью узконаправленных антенн хакеры научились производить «дальний взлом», мировой рекорд такого взлома летом 2005 года превысил расстояние в 1 км.

КОГДА ЭТО КОНЧИТСЯ?

Одна из причин слияния старых технологий связи — быстрый рост рынка беспроводной связи и коммуникаций (сотовые телефоны, спутниковая связь).

В скором времени количество приборов, которым требуется беспровод-

Телефон-зеркальце.

ная связь, настолько возрастёт, что снова встанет вопрос, как в заданном диапазоне частот разместить конечное число каналов. То есть, если используемые частотные ресурсы окажутся исчерпанными, придётся забираться всё выше и выше по частотной лестнице...

Но тут наряду с элементарной проблемой радиосвязи (электромагнитная энергия быстрее затухает, например во время дождя или при повышенной влажности) может возникнуть и другая: насколько вредным окажется влияние сверхвысоких частот на организм человека.

Тогда, пожалуй, человечество вспомнит о добрых старых проводах.

МОЛЧАНИЕ — ЗОЛОТО (НОВЫЕ ГОРИЗОНТЫ ПЕРЕДАЧИ ДАННЫХ)

О чем было нельзя мечтать вчера - стало реальностью сегодня. Компания со звучным названием «Сонет», предоставлявшая услуги мобильной телефонии стандарта CDMA, фактически перестала существовать в первых годах XXI века. Вероятно, перспективнейшая технология была вытеснена с рынка GSM-операторами. Но на ее месте, подобно волшебной птице Феникс, появилась более совершенная технология - CDMA-2000, развитие предыдущего стандарт CDMA, основанного на тех же принципах функционирования. Однако новая технология не только предоставляла возможность обмена данными на скоростях порядка 150 Кбит/с, но и не менее радужные перспективы по многократному увеличению этой скорости.

В системах CDMA для передачи сигналов используются радиоканалы с шириной полосы частот в 1,25 МГц, а повышение скоростных характеристик может быть достигнуто путем параллельного использования нескольких (до трех и более) радиоканалов. Поэтому в названии технологии словосочетание “Multi-Carrier” — многочастотная. Но, как всегда, оказалось, что можно использовать «внутренние резервы», лишь за счет усовершенствования системы кодирования и модуляции сигналов были достигнуты желаемые скорости передачи данных, до 153,6 Кбит/с. Этот факт получил свое отражение в названии: 1x — один радиоканал.

А при использовании 3-х радиоканалов можно добиться максимальной скорости передачи данных до 2,4 Мбит/с (IMT-MC-1x-EV-DO) и до 5,2 Мбит/с в стандарте IMT-MC-1x-EV-DV (*англ.* Evolution - DataVoice, эволюция данные и речь).

Но при реальном тестировании первых терминалов (так теперь уважительно называют телефоны с высокоскоростным доступом в Интернет) реальная скорость передачи данных оказалась в районе 25 Кбайт/с, что значительно ниже заявленного пятнадцатикратного увеличения и примерно соответствует EDGE — аналогичному эволюционному развитию GPRS. Вероятнее всего мешают разговоры абонентов, нещадно эксплуатирующие радиоканалы сотовой связи. Сам собой напрашивается лозунг текущего момента: «Молчание — золото!».

Технология ближайшего будущего носит имя EV-DO (*англ.* Evolution — Data Only) (Эволюция, только данных), а формально звучит так: IMT-MC-1x-EV-DO.

Аббревиатура IMT-MC обозначает группу стандартов сотовой связи (*англ.* International Mobile Telecom-munications-Multi-Carrier), разработанных Международным Институтом Электросвязи (ITU).

ИНТЕРНЕТ

ЗАРОЖДЕНИЕ, РАЗВИТИЕ И УСТРОЙСТВО ИНТЕРНЕТА

20 октября 1969 г. профессор Леонард Клейнрок из Калифорнийского университета в Лос-Анджелесе попытался со своего компьютера связаться с компьютером исследовательского центра в Станфорде и передать первое в истории электронное послание. Это было даже не письмо — Клейнрок собирался передать на компьютер, находящийся в Станфорде, команду LOG. Удалось переслать только две буквы, на букве G система дала сбой. В дальнейшем из этой попытки связать между собой две машины выросла глобальная компьютерная сеть, известная сейчас как Интернет. Этому предшествовала длинная цепочка событий, а 20 октября можно считать днём рождения Интернета.

Леонард Клейнрок.

ИСТОРИЯ СОЗДАНИЯ ИНТЕРНЕТА

История знает немало случаев, когда открытия и изобретения впоследствии использовались не в тех целях, для которых предназначались. Так, например, появился коньяк. Голландские купцы, чтобы снизить таможенную пошлину, перегоняли вино перед отправкой из Франции, уменьшая его объём, а у себя на родине снова разбавляли его. И лишь много лет спу-

стя французы обнаружили, как может быть хорош винный спирт, выстоявшийся в дубовых бочках. С Интернетом произошла похожая история.

Сразу после окончания Второй мировой войны США вступили в противостояние со своим недавним союзником — Советским Союзом. К середине 50-х гг. обе стороны обладали ядерным и даже термоядерным оружием и вели разработки ракетных носителей для него. В русле этого противостояния по приказу президента Эйзенхауэра при Министерстве обороны США было создано Агентство передовых исследований (Advanced Research Projects Agency — ARPA).

Агентство сосредоточило свои исследования на задачах военного применения компьютерных технологий. Главной задачей исследований стало создание компьютерной сети, которая была бы устойчива к повреждению её отдельных участков, например при ракетно-ядерной атаке противника. В 1967 г. был готов проект построения такой сети, получившей название ARPANET (Advanced Research Projects Agency NETwork), а в 1969 г. первыми узлами сети стали компьютеры калифорнийских университетов в Лос-Анджелесе и Санта-Барбаре, Станфордского исследовательского центра и Университета штата Юта. В 1971 г. к сети были подключены уже 23 компьютера, в 1973 г. — осуществлены первые международные подключения.

По мере роста сети ARPANET встала проблема, как связать отдельные сети, использующие различные методы передачи информации, между собой. Для этого необходимо было разработать специальный протокол, т. е. набор правил, которые определяли бы порядок обмена данными между различными программами, причём работающими на различных компьютерных платформах. В 1974 г. такой протокол был создан. Он включал правила налаживания и поддержания связи в сети, указания, как обрабатывать и передавать данные по сети. Этот протокол назвали TCP/IP.

К середине 80-х гг. к сети уже были подключены около тысячи компьютеров и TCP/IP был принят в качестве стандарта. В то же время появились настольные рабочие станции с операционной системой UNIX, оснащённой встроенными средствами для соединения с сетью, которая стала называться Интернет.

К 1986 г. Национальным фондом науки США (National Science Foundation — NSF) была создана опорная сеть для соединения его шести суперкомпьютерных центров. Сеть основывалась на протоколе TCP/IP. Скорость передачи данных по каналам составила 56 кбит/с. В дальнейшем к этой сети начали подключаться университеты.

В 1989 г. число хостов, т. е. компьютеров, подключённых к Интернету, превысило 100 тыс. Интернет стал использоваться не только в государственных и учебно-научных целях — к нему подключилась коммерческая сеть MCI Mail. К этому времени Сеть уже перешагнула границы Соединённых Штатов, в неё вошли Австралия, Великобритания, Германия, Израиль, Италия, Нидерланды, Новая Зеландия и Япония. В начале 90-х гг. Россия и страны Восточной Европы также подключились к Сети, которую теперь можно было назвать всемирной.

В 1991 г. Тим Бернерс-Ли, сотрудник лаборатории физики элементарных частиц Европейского центра ядерных исследований в Женеве (CERN), разработал систему World Wide Web (WWW), и началась новая страница в истории Интернета (см. статью «World Wide Web»).

Тим Бернерс-Ли.

КТО УПРАВЛЯЕТ ИНТЕРНЕТОМ

Интернет составляют соединённые между собой магистральные сети, иначе называемые опорными (backbone). Сети среднего уровня, региональные, подсоединяются к высокоскоростной опорной сети. Каждая из сетей отвечает за трафик (объём информации), который циркулирует внутри неё, и маршрутизирует (направляет) его, как считает нужным.

Каждая сеть несёт ответственность за соединение с сетью более высокого уровня, сама отвечает за своё финансирование, устанавливает собственные административные процедуры. Таким образом, не существует единого поставщика сетевых услуг, но все сети считаются полноправными частями Интернета.

Интернет в целом не контролируется какими-либо корпорациями, финансово-промышленными группами или государственными организациями. Он функционирует как единая система и развивается благодаря согласованным усилиям международного сообщества. Однако для поддержания единых стандартов, координации действий региональных и национальных сетей, являющихся частями Интернета, для объединения усилий при разработке новых технологий и решения других важных для жизни сетевого сообщества задач создан ряд организаций.

Координирует развитие мировой сети Собщество Интернета (Internet Society — ISOC). ISOC — это международная общественная организация, ставящая своей целью содействие глобальному информационному обмену через Интернет. ISOC назначает Комиссию по архитектуре Интернета (Internet Architecture Board — IAB), ответственную за техническое руководство. IAB координирует исследование и развитие протоколов Интернета, устанавливает правила присвоения сетевых адресов. В рамках IAB существует несколько рабочих групп специалистов. Одна — инженеры (Internet Engineering Task Force — IETF) — отвечает за разработку стандартов для протоколов и архитектуры Интернета. Другая — исследователи (Internet Research Task Force — IRTF) — концентрируется на развитии технологий будущего. Подразделение по реагированию на чрезвычайные ситуации (Computer Emergency Response Team — CERT) специализируется на вопросах безопасности Сети, оно координирует действия, направленные на защиту информации. За систему распределения адресного пространства Интернета отвечает международная организация ICANN (The Internet Corporation for Assigned Names and Numbers).

КАК УСТРОЕН ИНТЕРНЕТ

Если попытаться представить, как работает Интернет, то сначала напрашивается аналогия с телефонной сетью. И то и другое — средства электронной связи. И в том и в другом случае происходит соединение и передаётся информация. Многие подключаются к Интернету именно по телефону. Существенное различие состоит в принципе использования сети.

Телефонная сеть — это сеть с коммутацией каналов. Канал выделяется на время соединения двух абонентов и не может быть использован больше никем, даже если никакой информации не передаётся, например оба собеседника молчат. Интернет пользуется каналами связи более эффективно, это сеть с коммутацией пакетов. Информация разбивается на порции — пакеты, которые пересылают, как письма по почте. При этом канал используется для множества соединений одновременно. Различные участки Интернета связываются с помощью системы «маршрутизаторов» — компьютеров, которые анализируют заголовки пакетов и решают, какому узлу, находящемуся на пути к месту назначения, переправить тот или иной пакет. Они также выбирают альтернативные маршруты, если указанное сетевое соединение прервётся.

Это происходит примерно так же, как с обычными письмами. Отправитель вкладывает письмо в конверт, пишет на нём адрес получателя и отправителя, наклеивает марку... Почтовая служба при доставке корреспонденции следует определённым правилам, и точно так же существуют правила (протоколы), регламентирующие порядок работы Интернета.

За доставку пакетов и адресацию отвечает протокол Интернета (Internet Protocol — IP). Размер пакета данных, согласно IP, может быть от 1 до 1500 байт. К пакету прикладывается необходимая информация о пункте назначения. Каждому компьютеру, подключённому к Интернету, присваивается уникальный адрес — IP-адрес. Он состоит из четырёх чисел (байтов) от 0 до 255, отделённых друг от друга точками, например 62.3.250.189 или 193.20.1.147.

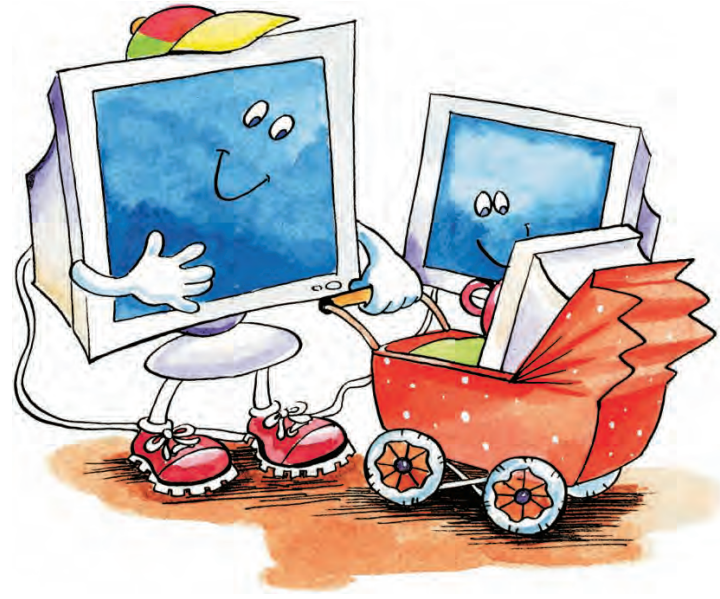
Между тем протокол IP не даёт гарантии, что из-за каких-либо неполадок пакет не может быть потерян. Также он не гарантирует, что будет соблюдаться последовательность доставки пакетов. В большинстве случаев объём пересылаемой по сети информации превышает 1500 байт. Решает эти проблемы протокол управления передачей — TCP (Transmission Control Protocol). Программное обеспечение протокола TCP на стороне отправителя разбивает информацию на пакеты и присваивает каждому из них порядковый номер. В пункте назначения TCP собирает пакеты и располагает их в соответствии с нумерацией. Если каких-то частей недостаёт, протокол требует повторной передачи. После размещения всей информации в правильном порядке эти данные передаются прикладной программе. Помимо этого TCP проверяет контрольную сумму пакета — сумму всех его байтов, вычисленную по специальной формуле. Если контрольная сумма не совпадает, значит, пакет доставлен с искажениями, и будет запрошена повторная передача.

Надёжность работы на любой платформе, простота и доступность протокола TCP/IP определили его успех. И если раньше TCP/IP считался протоколом исключительно для работы

с Интернетом и операционной системой UNIX, то в дальнейшем он нашёл более широкое применение. Он стал основным протоколом и в локальных сетях.

ДОМЕНЫ — ИМЕНА В ИНТЕРНЕТЕ

Интернет не стал бы такой удобной информационной средой, если бы приходилось пользоваться одними только цифровыми адресами. Они хороши при непосредственном взаимодействии компьютеров, но для людей предпочтительнее имена. Эта проблема обозначилась ещё в годы становления Интернета. Тогда всем входящим в сеть компьютерам были присвоены собственные имена. В качестве имён выбирали простые слова, каждое из которых обязательно



Названия доменов в именах отделяют друг от друга точками, например gmk.pp.ru, burleson.rtp.epa.gov, www.infomir.ru, abcd.abdn.ac.uk. В имени может быть любое число наименований доменов, но больше пяти, как правило, не встречается.

должно было быть уникальным. Подключаемый к сети компьютер регистрировался в Сетевом центре информации (Network Information Center — NIC). Данные обо всех компьютерах заносили в специальный файл, который регулярно рассылался по сети. Тогда при обращении к какому-либо компьютеру по имени система просматривала этот список и вычисляла IP-адрес. Вскоре Интернет сильно разросся, и файл со списком имён достиг таких размеров, что его стало невозможно оперативно рассылать; регистрация уникальных имён усложнилась, поиск замедлился. Для решения этой проблемы была введена доменная система имён (Domain Name System — DNS), используемая и поныне.

Сущность DNS состоит в том, что ответственность за присвоение имён возлагается на различные группы пользователей, контролирующих ту или иную область Интернета — домен (domain — «область»).

Доменное имя, в отличие от цифрового, разбирается слева направо. Самым первым обычно стоит имя компьютера, имеющего IP-адрес, далее следует имя домена, в который входит этот компьютер, потом идёт имя более крупного домена и т. д. Если владелец домена решит включить в сеть новый компьютер, он не обязан ни у кого спрашивать разрешения. Всё, что от него требуется, — это правильно внести соответствующую запись в базу данных на сервере имён своего домена.

Домены самого верхнего уровня существуют с момента введения DNS в начале 80-х гг. Изначально в США были созданы домены gov (правительственные учреждения), mil (военные учреждения), com (коммерческие организации), edu (учебные заведения), net (сетевые организации), org (прочие организации). Эта кодировка перешла и в XXI в. Когда к Интернету подключились сети других стран, в их ведение были отданы собственные домены. В соответствии с международным стандартом принята двухбуквенная кодировка государств. Позже появился домен int (международные организации).

В 2000 г. ICANN приняла решение о введении семи новых доменов верхнего уровня. Домен aero закреплён за авиалиниями, аэропортами и системами бронирования билетов; name предназначен для личного использования; biz — для сайтов коммерческих организаций; coop — для деловых союзов; info — для информационных проектов; museum — для музеев. За Североатлантическим блоком НАТО закреплён домен nato. Евросоюз для своих сетевых проектов планирует использовать домен eu.

Все программы, работающие на компьютерах, подключённых к Интернету, используют доменную систему имён. Каждый раз при обращении к тому или иному компьютеру по имени происходит преобразование запроса в IP-адрес. Компьютер может знать адрес, если он есть в его базе данных или если этот адрес ранее уже запрашивался и хранится в кэш-памяти компьютера. Если компьютер не знает адреса, тогда он запрашивает у сервера имён своего домена адрес сервера имён соответствующего домена, а если тот неизвестен, то запрашивает адрес сервера имён более высокого уровня и т. д.

При работе с доменной системой имён следует помнить, что части доменного имени сообщают только

В адресе www.InfoMir.ru www — это имя компьютера (вероятно, на нём установлен www-сервер), входящего в сеть компании «ИнфоМир» (имя домена группы — InfoMir), ru — наименование национального домена Российской Федерации.В адресе alpha.math.msu.su первым стоит имя компьютера в сети механико-математического факультета (math) Московского государственного университета (msu — Moscow State University), входящего в домен su — наследство Советского Союза.

о том, в чьём ведении это имя находится. При этом далеко не всегда легко определить, где физически находится компьютер. Более того, компьютеры одного домена порой находятся в разных сетях. Тем не менее использовать доменные имена предпочтительнее, чем IP-адреса. Во-первых, большинству людей их проще запоминать, во-вторых, IP-адрес компьютера может измениться по ряду причин (например, из-за переконфигурации локальной сети), имя же не меняется почти никогда. Наконец, у компьютера может быть несколько имён-синонимов помимо его канонического имени — обычно они образуются от имени соответствующего сервиса, предоставляемого этим компьютером.

ДОСТУП В ИНТЕРНЕТ

Интернет настолько вошёл в быт и стал доступен почти любому человеку, что уже воспринимается неискушёнными пользователями как часть компьютера. Нередко торговцы сталкиваются с такими вопросами покупателей: «А этот компьютер с Интернетом?», «Где у него Интернет?». Но, увы, Интернет находится не внутри компьютера. Компьютер лишь средство для доступа в Интернет.

Практически все современные операционные системы позволяют работать с Интернетом, иначе говоря, в них встроена поддержка протокола TCP/IP. Компьютер с такой системой готов к тому, чтобы подключиться к Интернету, стать его частью. Кроме того, он должен быть оснащён модемом или сетевой картой. Эти устройства, как правило, уже находятся в компьютере, но иногда их необходимо приобретать отдельно.

ПОДКЛЮЧЕНИЕ К ИНТЕРНЕТУ. ВИДЫ ДОСТУПА

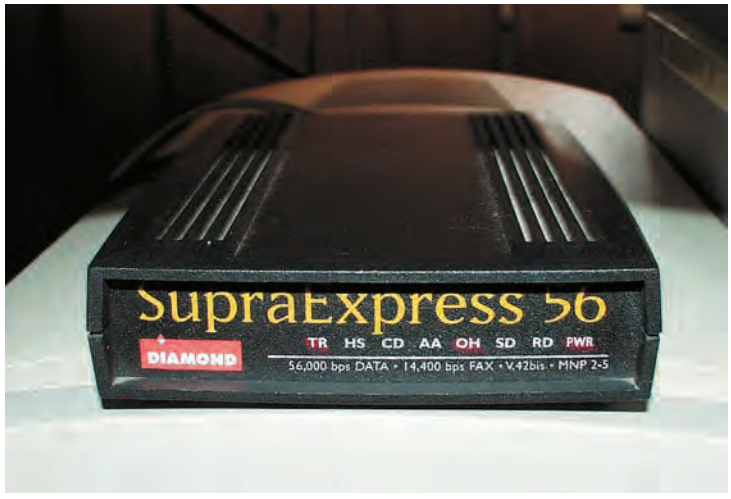
Доступ к Интернету предоставляют специальные организации, которые называются провайдерами (Internet service provider — «поставщик услуг Интернета»). Компьютерная сеть провайдера постоянно подключена к Интернету. Для этого арендуют выделенные скоростные каналы связи или прокладывают собственные. Сети провайдеров соединяются в опорные, магистральные сети, которые в конечном счёте и образуют Интернет.

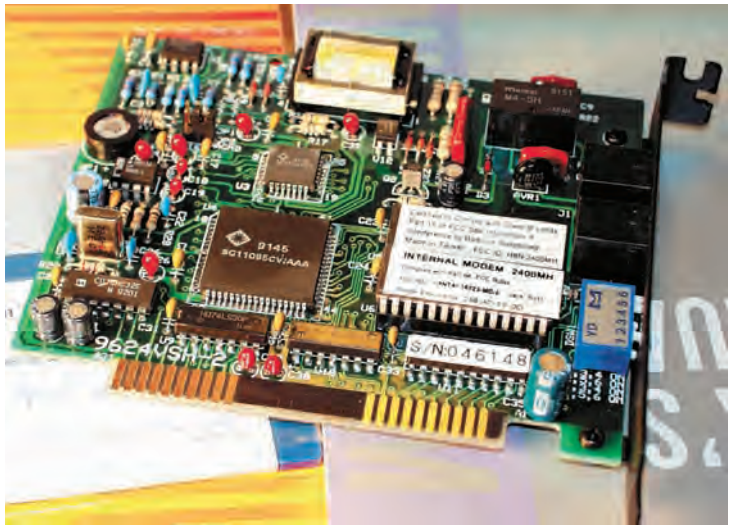
Способы доступа к Интернету отличаются схемами подключения, линиями связи и протоколами, которые определяют различные услуги, предоставляемые сетью. Чем больше возможностей предоставляет способ доступа, чем он быстрее, тем и дороже.

Если модем поддерживает протокол V34, скорость передачи данных не выше 28,8 кбит/с. Если модем поддерживает протокол V34+, скорость передачи не выше 33,6 кбит/с. Если модем поддерживает протокол V90, то скорость передачи может достигать 57,6 кбит/с.

DIAL UP

В начале XXI в. наиболее широко распространён коммутируемый доступ в Интернет по телефонным линиям (dial up). На момент соединения компьютер становится частью





Внешний (*вверху*) и внутренний модемы.

Интернета и получает доступ ко всем его ресурсам. Коммутируемый доступ использует специальный протокол PPP (Point-to-Point Protocol — «протокол связи компьютер — компьютер»). Для этого вида доступа необходим модем — специальное устройство преобразования цифровых данных в аналоговый сигнал телефонной линии и обратно.

Провайдеры, предоставляющие модемный доступ dial up, обычно предлагают различные схемы оплаты своих услуг (так называемые тарифные планы). Они назначают цену по времени связи с Интернетом. Фактически же оплата рассчитывается по времени телефонной связи с провайдером. А это не всегда одно и то же. Дело в том, что в силу различных причин связь провайдера с компьютерной сетью иногда прерывается или вовсе отсутствует. Как правило, работа ночью и в выходные дни дешевле. Иногда провайдеры назначают абонентскую плату и предоставляют неограниченный по времени доступ. Хорошие провайдеры предлагают бесплатный тестовый доступ, чтобы клиенты могли ознакомиться с качеством связи.

Модемный доступ имеет ряд недостатков. Телефонные линии изначально не предназначались для передачи цифровых данных, поэтому их использование в этих целях связано с некоторыми ограничениями и неудобствами. Максимальная скорость передачи данных по телефонным линиям составляет 56 кбит/с. В реальности она значительно ниже, так как зависит от качества телефонных проводов и возможностей АТС (автоматической телефонной станции). К тому же для соединения с Интернетом необходимо предварительно дозвониться до провайдера по телефонной линии (а это порой длится часами, если количество клиентов провайдера значительно превышает число модемных входов). Низкое качество телефонных линий не позволяет модемам долго удерживать надёжную связь. В результате прерванного соединения теряются данные, которые приходится скачивать заново.

ДОСТУП ПО ВЫДЕЛЕННОЙ ЛИНИИ

Многих активных пользователей Интернета уже не устраивает скорость передачи данных через модем, особенно если за это нужно платить дополнительные деньги (пользователю приходится оплачивать услуги не только провайдера, но и телефонной сети, часто тоже повремённо). Кроме того, телефон может понадобиться и для обычных разговоров. Поэтому на смену доступу dial up приходят другие, более совершенные способы подключения. Например, арендуется выделенный канал связи с большой пропускной способностью (один из вариантов — волоконно-оптическая линия). Пропускная способность выделенной линии, как правило, не ниже 64 кбит/с. В этом случае можно подключить к Интернету даже несколько компьютеров — локальную сеть. Каждый из компьютеров такой сети становится полноправным участником Интернета.

Подключение к Интернету по выделенному каналу имеет ряд преимуществ по сравнению с модемным доступом по телефонной линии. Выделенный канал для передачи данных рассчитан на круглосуточную бесперебойную работу (не нужно дозваниваться до провайдера) на высоких скоростях (например, 100 Мбит/с). При этом телефонная линия остаётся свободной. За счёт высокой скорости передачи данных стоимость полученной информации в среднем оказывается ниже, чем при доступе через модем.

Абонентская плата за доступ по выделенной линии зависит от пропускной способности канала. Кроме этого, деньги взимаются за объём информации, загруженной через Интернет, — так называемый трафик. Сумма за определённый трафик может быть включена в абонентскую плату, а превышение оплачивается дополнительно. Время соединения, как правило, не оплачивается.

Кроме доступа с помощью модема и по выделенным волоконно-оптическим линиям провайдеры предлагают подключение по беспроводным каналам связи — с помощью радиомостов, через спутники связи. При «спутниковом» доступе к компьютеру подключают параболическую спутниковую антенну-тарелку и преобразователь данных подобный тому, какой применяют для спутникового телевидения. Искусственный спутник Земли, связанный с Интернет-провайдером, может передавать пользователю данные на довольно высокой скорости (до 45 Мбит/с). Запросы и команды при этом способе доступа пользователь передаёт через обычный модем.

Ещё один из наиболее перспективных и активно развивающихся способов подключения к Интернету — ADSL (Asymmetric Digital Subscriber Line). Эта аббревиатура расшифровывается как

КАНАЛЫ СВЯЗИ

Компьютеры, подключённые к Интернету, соединены между собой с помощью каналов связи. Каналы характеризуются *пропускной способностью,* т. е. количеством информации, которую можно передать по каналу в единицу времени. Единица измерения пропускной способности — бит/с.

Пропускная способность различных каналов связи может составлять от нескольких килобитов в секунду до сотни мегабит в секунду. Пропускная способность канала бывает *средней* и *гарантированной.* Средняя пропускная способность измеряется в среднем за определённый промежуток времени. Гарантированной пропускной способностью называют минимальную пропускную способность, которую обеспечивает канал в наименее благоприятных условиях.

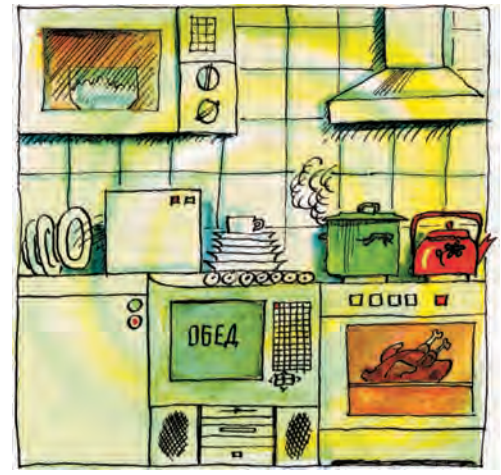
Пропускная способность — основная характеристика любого канала, по ней судят о его приемлемости для использования.

В зависимости от типа передаваемых сигналов различают *цифровые* и *аналоговые* каналы. Цифровой канал имеет цифровой (импульсный) сигнал на входе и выходе. Напротив, на вход аналогового канала поступает непрерывный сигнал, и с его выхода также снимается непрерывный сигнал. Параметры сигналов тоже могут быть непрерывными или принимать только дискретные значения. Сигналы могут содержать информацию либо в каждый момент времени (непрерывные во времени, аналоговые сигналы), либо только в определённые, дискретные моменты времени (цифровые, дискретные, импульсные сигналы). Цифровыми являются каналы системы ISDN, каналы типа T1/E1 и многие другие. В начале XXI в. цифровые каналы стали активно вытеснять аналоговые, перед которыми они обладают рядом преимуществ.

Аналоговые каналы (например, телефонные линии) просты в реализации, но подвержены влиянию помех и обладают малой пропускной способностью (несколько десятков килобайт в секунду). При передаче данных на входе и выходе аналогового канала должно находиться устройство, которое преобразовывало бы цифровые данные в аналоговые сигналы и наоборот.

Каналы также делятся на *коммутируемые* и *выделенные*. Коммутируемые каналы предоставляются потребителям на время соединения по их требованию. Обычные телефонные аппараты используют коммутируемые каналы. Кроме того, коммутируемые каналы предоставляет цифровая сеть с интеграцией служб (ISDN — Integrated Services Digital Network). Выделенный канал работает по-другому: соединение в этом случае постоянное, в любой момент позволяет передать данные от одного компьютера к другому. Выделенные каналы отличаются от коммутируемых высокой скоростью передачи (до десятков мегабит в секунду). Их качество выше качества коммутируемых каналов, потому что не зависит от состояния коммутационной аппаратуры на телефонных станциях.

По физическому устройству каналы бывают *электрические проводные, оптические* и *беспроводные*. Проводные каналы представляют собой электрический кабель, часто сложной конструкции. Во всех таких каналах передача данных происходит при помощи электрических импульсов. В оптических каналах связи применяют световоды, сигнал же передаётся при помощи лазеров. Беспроводные — спутниковые каналы, радиоканалы действуют по тому же принципу, что и радио и телевидение.



«асимметричная цифровая абонентская линия». Название подчёркивает изначально заложенное в этой технологии различие скоростей обмена данными в направлениях к пользователю и обратно. Асимметричность скорости передачи вводится специально: пользователь обычно загружает большой объём информации из Интернета на свой компьютер, а в обратном направлении идут или только запросы, или поток существенно меньшего объёма. В начале XXI в. ADSL обеспечивает входящую скорость до 8 Мбит/с и исходящую — 1 Мбит/с. Подключение осуществляется по телефонной линии, на концах которой установлено специальное оборудование. По линии может одновременно осуществляться и обычная телефонная связь, и работа в Интернете.

Это возможно потому, что для передачи данных используется диапазон частот, находящийся выше уровня различимости человеческого голоса (4 кГц). Экономически это выгодно: не нужно прокладывать отдельный кабель, а используются существующие телефонные линии. К большинству вновь строящихся домов выделенные каналы для связи с Интернетом подведены так же, как и электрические провода, водопровод и канализация. Это особенно актуально в связи с расширяющимися возможностями подключения к Интернету не только персональных компьютеров, но и других бытовых устройств, например кондиционера или микроволновой печи.

КАКОЙ КАНАЛ ВЫБРАТЬ?

Когда стоимость винчестеров была велика и устанавливать их на каждый компьютер было нерентабельно, использовались сетевые операционные системы. Подобные системы умели соединить несколько персональных компьютеров, обеспечив совместное использование небольшого количества данных и основных приложений. Загрузка операционной системы осуществлялась с центральной машины. Сетевые ОС напоминали центры коллективного пользования, когда терминалы подключены к одной машине. В качестве терминалов выступали персональные компьютеры, да и «интеллект» терминалов был существенно выше, потому что программы выполнялись не на центральной машине, а на компьютере пользователя.

Наиболее успешно сетевые ОС применялись при обучении в университетах, так как, с одной стороны, студенту не требуется особенно больших компьютерных ресурсов, с другой — сетевой ОС удобно управлять. Сегодня сетевые операционные системы должны делать всё: работать со старыми файловыми системами и одновременно обрабатывать приложения клиент — сервер.

Самой распространённой сетевой ОС является Novell Netware.

КЛАССИЧЕСКИЕ СЕРВИСЫ ИНТЕРНЕТА

Образ Интернета в массовом сознании прочно ассоциируется со страничками Всемирной паутины — World Wide Web. Но когда-то ситуация была совершенно иной.

Интернет первоначально был создан как компонент информационно-командной системы Вооружённых сил и разведывательной службы США. Затем стал инструментом в руках гражданских учёных, специалистов в области компьютерных сетей. Надо сказать, инструментом не таким удобным, как сейчас.

Кстати, и набор сервисов, которые теперь принято называть классическими (удалённый доступ, пересылка файлов и электронная почта), был разработан ещё в середине 70-х гг. XX в.

УДАЛЁННЫЙ ДОСТУП

Когда-то компьютеры имели большие размеры и стояли в специальных машинных залах. Терминалы — дисплеи с клавиатурой, позволяющие работать на компьютере, обычно располагались в другом помещении. Дисплеи были алфавитно-цифровые, поэтому диалог с компьютером заключался во вводе символьных команд, реагируя на которые машина печатала на экране

ответ — соответствующие данные. При создании системы удалённого доступа был сохранён этот способ диалога с компьютером.

Программа удалённого доступа называется TELNET (teletype network). Она хорошо опробована и широко распространена. TELNET позволяет пользователю входить в любой доступный компьютер сети, как бы далеко он ни находился, и работать на нём как с удалённого терминала, сменяя каталоги, просматривая файлы, запуская программы.

TELNET состоит из двух взаимодействующих между собой компонентов: клиента и сервера.

Программа-клиент выполняется на компьютере, посылающем запрос на соединение. Она должна установить сетевое соединение с требуемым компьютером, принять от пользователя входные данные в любой удобной форме, привести эти данные к стандартному формату и послать их серверу; затем принять от сервера выходные данные в стандартном формате; переформатировать полученные выходные данные для отображения их на дисплее пользователя.

Программа-сервер выполняется на компьютере, предоставляющем запра-

РАБОТА ПРОГРАММЫ TELNET

Для того чтобы войти в удалённый компьютер, пользователь вводит команду

`telnet` *имя_компьютера*.

Когда соединение произойдёт, удалённая система запросит входное имя (Login:) и пароль (Password:). Для окончания сеанса работы и отключения от удалённой системы пользователь должен ввести команду `logout`, и соединение будет прервано.

Программа TELNET позволяет переходить из режима работы в удалённой системе в командный режим. Это происходит при вводе так называемой escape-последовательности (но не клавиши Esc!), которая сообщается при соединении. Обычно это совместное нажатие клавиш Ctrl и Esc. Кроме того, в командный режим можно войти, введя команду `telnet` без указания имени компьютера.

Основные команды TELNET (именно КОМАНДНОГО режима!):

`?` — вывести список команд;
`close` — закрыть текущее соединение;
`quit` — выйти из программы TELNET;
`open` *имя* — установить соединение с указанным компьютером;
`set` — установить различные параметры соединения; например, команда `set echo` включает и выключает локальное эхо, `set escape` *символ* объявляет указанный символ escape-символом. Смена escape-символа бывает полезна для соединения с несколькими удалёнными компьютерами. Например, при входе в одну удалённую систему через другую и нажатии используемого по умолчанию Ctrl-] будет разорвана связь с обеими системами, даже если этого не требовалось;
`display` — выдать рабочие параметры сеанса;
`status` — показать информацию о состоянии соединения;
`z` — перевести соединение с удалённой системой в фоновый режим.

При нажатии клавиши Enter происходит возврат из командного режима.

(В режиме работы в удалённой системе вы работаете непосредственно на удалённом компьютере, а командный режим просто изменяет различные параметры сеанса связи.)

шиваемую услугу, в фоновом режиме. Она ожидает запроса на свои услуги и по мере поступления обслуживает его, передавая результаты программе-клиенту в стандартном формате. Если программа-сервер не активизирована, услуга удалённого доступа будет невозможна.

Программа-сервер и программа-клиент обмениваются данными на основе набора правил (протокола). Точное следование протоколу позволяет устанавливать соединение и проводить сеанс связи между компьютерами независимо от их статуса; это может быть как персональный компьютер, так и рабочая станция или сервер. Операционные системы на связывающихся компьютерах также могут быть различными.

Удалённый доступ как один из базовых сервисов Интернета не мог не попасть под прицел злоумышленников. Оказалось, что данные пользователя передаются по протоколу TELNET «открытым текстом» и могут быть «подслушаны». Чтобы решить эту проблему, в конце XX в. для удалённого доступа были разработаны криптографически защищённый протокол и основанная на нём программа SSH (Secure Shell).

ПЕРЕДАЧА ФАЙЛОВ

Программа для работы в удалённой системе TELNET может действительно очень многое, но она не позволяет перемещать файлы с одного компьютера на другой. Для этой цели существует программа FTP, получившая своё название от одноимённого прикладного протокола (File Transfer Protocol — «протокол передачи файлов»).

FTP позволяет пересылать данные файлов либо как двоичную информацию, либо как текст; переходить из директории в директорию, просматривать содержимое этих директорий, файлов; пересылать и одиночные файлы, и группы файлов, а также целиком директории вместе со всеми вложенными на любую глубину поддиректориями. Программа FTP, так же как и TELNET, состоит из FTP-сервера, выполняющегося на удалённом

РАБОТА ПРОГРАММЫ FTP

Чтобы соединиться с удалённым компьютером по протоколу FTP, необходимо ввести команду

`ftp` *имя_компьютера*.

При этом происходит соединение с указанным компьютером. После этого на удалённом компьютере следует ввести входное имя пользователя и пароль.

Основные команды FTP:

`quit` — выйти из FTP;

`dir` *файл* — просмотреть содержимое каталога на удалённом компьютере, в качестве параметра можно указать имя файла или имя группы файлов с использованием универсальных символов * или ?;

`cd` *каталог* — сменить каталог в удалённой системе. Имена каталогов отделяются косой чертой /; если имя каталога начинается с косой черты, то позиционирование осуществляется относительно корневого каталога. Для перехода на один уровень выше следует ввести команду `cd ..` или `cdup`;

`binary` — включить двоичный режим передачи файлов. При этом все байты пересылаются без каких-либо преобразований или сокращений. Таким способом пересылаются архивы, исполняемые файлы, изображения и т. п.;

`get` *файл1 файл2* — переслать из удалённой системы файл1, создав на компьютере пользователя (локальном) его копию с именем файл2. Если второй параметр не указывать, то в локальной системе будет создан файл с оригинальным именем. По команде `get` *файл* указанный файл будет выведен на экран;

`put` *файл1 файл2* — производит действие, обратное действию команды get, т. е. пересылает файл в удалённую систему из локальной;

`delete` *файл* — стереть указанный файл в удалённой системе;

`hash` — выводить символ # после пересылки каждого блока данных. Это полезно при пересылке больших файлов, так как позволяет убедиться, что данные действительно перемещаются. Для отмены этого режима следует дать команду `hash` повторно.

`help` — показать полный список доступных команд FTP.

компьютере в фоновом режиме, и программы-клиента. При использовании протокола FTP между соединяющимися компьютерами открывается два канала: один для передачи команд, другой для передачи данных.

АНОНИМНЫЙ FTP

Одной из важнейших возможностей, предоставляемых протоколом FTP, является анонимный FTP, позволяющий сделать те или иные файлы доступными всему сетевому сообществу. Анонимный FTP не требует от пользователя регистрации на компьютере, где установлен этот сервис. Для анонимного входа используется специально зарезервированное имя Anonymous. На анонимных FTP-серверах можно найти самую разную информацию. Это всевозможные драйверы, общедоступное программное обеспечение для разнообразных операционных систем, архивы компьютерных компаний, которые помещают на свои серверы обновлённые и демонстрационные версии своих программных продуктов, а также дополнения к документации или сообщения об обнаруженных ошибках. Там можно найти разнообразную документацию, библиотеки компьютерных изображений, карты, схемы, репродукции картин, коллекции фотографий, различные тексты: Библию, Коран, Тору, романы, научную фантастику и чёрную магию, исторические документы, прогнозы погоды, гороскопы и кулинарные рецепты; наконец, файлы с музыкой и кинофильмами, видеоклипы.

ЭЛЕКТРОННАЯ ПОЧТА

Электронная почта (E-mail — Electronic mail, от *англ.* mail — «почта») — наиболее естественное и легко осваиваемое приложение, поскольку для неё допустима прямая аналогия с обычной почтой. С её помощью пользователь может посылать сообщения, получать их в собственный электронный почтовый ящик, отправлять ответы своим корреспондентам автоматически по тем адресам, откуда пришли письма, рассылать копии своего письма сразу нескольким получателям, переправлять полученное сообщение по другому адресу.

Каждому пользователю электронной почты присваивается уникальный почтовый адрес, который обычно образуют присоединением имени пользователя к имени его компьютера. Имя пользователя и имя компьютера разделяет специальный символ «@» (изначально выбранный как не встречающийся в именах и фамилиях), который называется «эт коммерческое» (at). Например, если пользователь имеет входное имя frederick на компьютере ickenham.com, то его электронный адрес будет иметь такой вид: frederick@ickenham.com.

Для использования электронной почты на компьютере устанавливают почтовый сервер и виртуальный «почтовый ящик», доступный только его владельцу. Все приходящие письма складываются туда и ждут момента, когда пользователь прочитает их при помощи специальной программы-клиента. В этой же программе пользователь может подготовить своё письмо и послать его. Тогда программа-клиент передаст это письмо программе-серверу, которая и отправит его по сети адресату.

Электронное письмо, так же как и обычное, содержит адрес получателя, адрес отправителя, на нём есть «штампы отделений связи» — имена компьютеров, через которые прошло письмо, прежде чем добраться до адресата. Также в письме есть заголовок (Subject) — строка текста, позволяющая облегчить получателю классификацию писем, определить их срочность и необходимость немедленного

ответа. Современные почтовые программы-клиенты имеют удобный интерфейс в форме меню и диалогов, а также всё необходимое для хранения полученных и отправленных писем. Кроме того, они позволяют прикрепить к письму, имеющему обычный текстовый вид, картинку или любой другой файл. В действительности такой файл будет закодирован в последовательность букв и включён в письмо, но программа получателя автоматически раскодирует его.

У каждого почтового клиента есть адресная книга — список наиболее часто используемых адресов.

ТЕЛЕКОНФЕРЕНЦИИ

Система телеконференций — распределённая виртуальная сеть Usenet — позволяет участвовать в открытых дискуссиях, даёт возможность читать и посылать сообщения в группы новостей.

Обычно провайдеры наряду с электронной почтой предоставляют доступ к своему серверу новостей. Подключившись к серверу, необходимо загрузить полный список групп, который он поддерживает, затем подписаться на интересующие группы. Подписка будет сохраняться при всех последующих сеансах связи. Размещённые в конференции сообщения принято называть статьями (articles).

Существует несколько тысяч групп новостей, поэтому конференции Usenet организованы по иерархическому принципу. Изначально было создано семь категорий, лежащих в основе этой иерархии:

- comp — вычислительная техника, информатика, программное обеспечение и смежные области;
- sci — научно-исследовательская деятельность;

ОСОБЕННОСТИ «ЭЛЕКТРОННОГО» СТИЛЯ ПИСЬМА

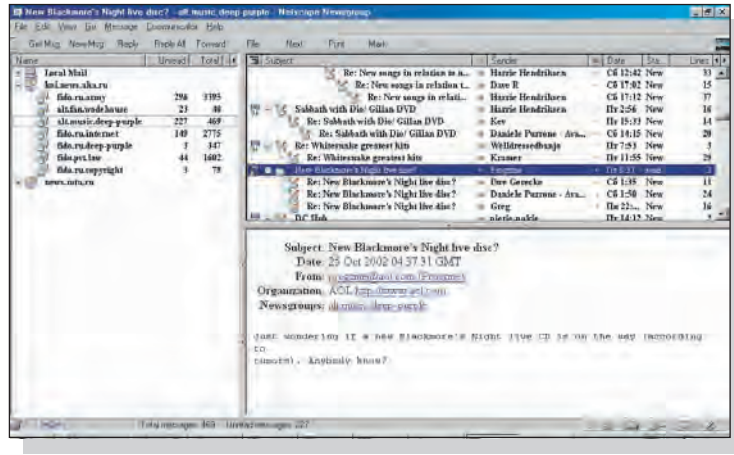
Большинство пользователей тяготеют к неформальному стилю общения и придерживаются принятых в электронном мире традиций. К ним относятся разнообразные сокращения. Вот некоторые из них:

IMHO — in my humble opinion — по моему скромному мнению;
FYI — for your information — для вашего сведения;
ASAP — as soon as posible — как можно скорее;
NC — no comment — без комментариев;
BTW — by the way — кстати;
THX — thanks — спасибо.

Придать эмоциональную окраску текстовым посланиям можно, используя стилизованные изображения мимики человеческого лица — смайлики (smile — «улыбка»). Идея состоит в том, что на обычные символы надо смотреть под углом 90°:

:-) — улыбка;
;-) — хитрая улыбка с подмигиванием;
:-(— хмурая физиономия;
:-> — саркастическая физиономия;
:-O — поражён.

Вам следует чаще проводить время в кругу семьи, если ваши домашние вызывают вас на обед по электронной почте.



Телеконференция, посвящённая группе «Дип Пёпл».

ИЗОБРЕТЕНИЕ ЭЛЕКТРОННОЙ ПОЧТЫ

Почти все изобретения последних столетий имеют автора. Так, в 1807 г. американец Роберт Фултон построил первый пароход, в 1903 г. американцы же братья Орвилл и Уилбер Райт подняли в воздух первый в мире самолёт. Изобретатель электронной почты также известен. Это профессор Рэй Томлинсон, тоже американец.

Когда в 1971 г. Томлинсон сделал своё гениальное изобретение, он работал в компании BBN, создававшей сеть ARPANET (сеть Управления перспективных исследований Министерства обороны США) — прототип современного Интернета. Электронную почту он создал без ведома руководства, просто ради собственного интереса. В её основу легла уже имевшаяся программа передачи текстовых сообщений, которую использовали программисты и исследователи, работавшие на компьютерах фирмы DeC. Но с помощью этой программы сообщениями могли обмениваться только пользователи одного компьютера. Томлинсону оставалось предусмотреть возможность пересылки сообщений на другие компьютеры, и, слегка изменив протокол передачи данных, он получил первую версию почтовой программы.

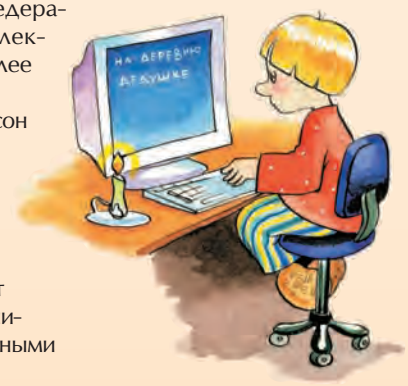
Для тестирования программы Томлинсон воспользовался двумя компьютерами, стоявшими в его комнате и соединёнными посредством сети. Первое сообщение по электронной почте он послал самому себе. По воспоминаниям Томлинсона, оно состояло из набора заглавных букв, расположенных в верхней части клавиатуры, — что-то вроде QWERTYUIOP.

Создав первую в мире почтовую программу, Томлинсон показал её другим сотрудникам и попросил их никому о ней не рассказывать, ибо опасался, что начальство будет упрекать его в пренебрежении своими служебными обязанностями. Но впоследствии электронной почтой стал пользоваться его шеф Ларри Робертс. Через некоторое время он перевёл основную часть своей деловой переписки в электронную форму, так что его сотрудникам волей-неволей пришлось последовать примеру босса.

Рэй Томлинсон.

Распространение электронной почты резко возросло в середине 80-х гг. прошлого века, когда появились первые персональные компьютеры: круг её пользователей пополнился за счёт частных лиц. К этому же периоду относится и создание первой массовой почтовой программы Eudora с удобным графическим интерфейсом.

В настоящее время практически все пользователи Интернета имеют как минимум один, а нередко и несколько почтовых адресов. По оценкам аналитиков, в 2004 г. электронной почтой пользовалось более половины жителей США, ежедневно через Интернет пересылалось около 16 млрд электронных писем. А в Российской Федерации пользуются электронной почтой более 15 млн человек.



Сам Рэй Томлинсон не получил от своего грандиозного изобретения никакой материальной выгоды. Он по-прежнему работает в компании BBN, занимается новыми научными разработками.

Преимущества электронной почты налицо. Она экономит время в отличие от обычной почты, которую американцы — активные пользователи E-mail назвали «улиточной» (snail mail).

• soc — социальные вопросы и проблемы культуры различных народов;

• news — сетевые новости, в частности о самой системе новостей;

• rec — всевозможные хобби и развлечения;

• talk — разговоры на любые темы, дискуссии по спорным вопросам;

• misc — всё остальное.

Впоследствии была организована категория alt — самая большая и популярная. В неё вошли конференции по альтернативным тематикам, т. е. всё, что не попало ни в одну из семи вышеперечисленных категорий.

Круг тем, обсуждаемых при помощи телеконференций, необычайно обширен и разнообразен и включает в себя практически все стороны жизни человека. Вот лишь несколько названий групп новостей:

• alt.music.deep-purple — поклонники британской рок-группы «Дип Пёпл»;

• alt.fan.wodehouse — место общения почитателей американского писателя Пэлема Грэнвилла Вудхауза;

• comp.lang.fortran — дискуссия по вопросам языка программирования FORTRAN;

• sci.bio.paleontology — конференция палеонтологов;

• rec.arts.tv.uk.comedy — для любителей английских телевизионных комедий;
• rec.pets.cats — любители кошек;
• soc.culture.armenian — проблемы армян по всему миру.

В связи с возросшей популярностью и повсеместным распространением World Wide Web были созданы средства для работы с системой конференций через web-страницы.

WORLD WIDE WEB

Интернет сделал возможным свободный обмен информацией, невзирая на границы и расстояния. К началу 90-х гг. XX в. в глобальной сети были накоплены огромные объёмы разнообразной информации. Однако эти ресурсы были доступны в основном узкому кругу специалистов, которым приходилось пользоваться программами, ориентированными лишь на пересылку файлов, работу в режиме удалённого терминала и обмен сообщениями. Ситуация изменилась с появлением и распространением World Wide Web (WWW).

World Wide Web — это глобальная распределённая информационная гипертекстовая мультимедийная система. Она позволила связать в единое целое информацию, хранящуюся в разных компьютерах по всему миру.

Начало WWW было положено в 1989 г., когда британский программист Тим Бернерс-Ли из Европейского центра ядерных исследований в Женеве (CERN) начал работу по созданию информационной системы, которая должна была объединить всё множество информационных ресурсов CERN (базы данных научных отчётов и результатов экспериментов, компьютерную документацию, списки почтовых адресов, наборы данных и т. п.). Гипертекстовая технология должна была позволить легко переходить от одного документа к другому.

Создавая World Wide Web, Тим Бернерс-Ли едва ли мог представить, что вскоре почти каждая компания, университет, клуб, магазин или журнал будет иметь свой сайт, а вокруг этой системы вырастет целая индустрия. Сейчас WWW позволяет людям бронировать места в отелях и билеты на поезд, знакомиться с научными исследованиями, читать газеты, напечатанные в разных странах, и узнавать погоду на другой стороне планеты.

> Идея гипертекста состоит в том, что его документы (страницы текста) можно просматривать в произвольном порядке, а не последовательно, как это принято при чтении книг. В гипертексте выделены особые места — ссылки. Они могут сразу привести читателя к нужным разделам, рисункам, описаниям. Благодаря этому процесс чтения становится принципиально иным: гипертекст можно просматривать многими различными путями, и читатель сам выбирает тот путь просмотра, который ему наиболее удобен. В качестве примера гипертекста можно привести энциклопедию.

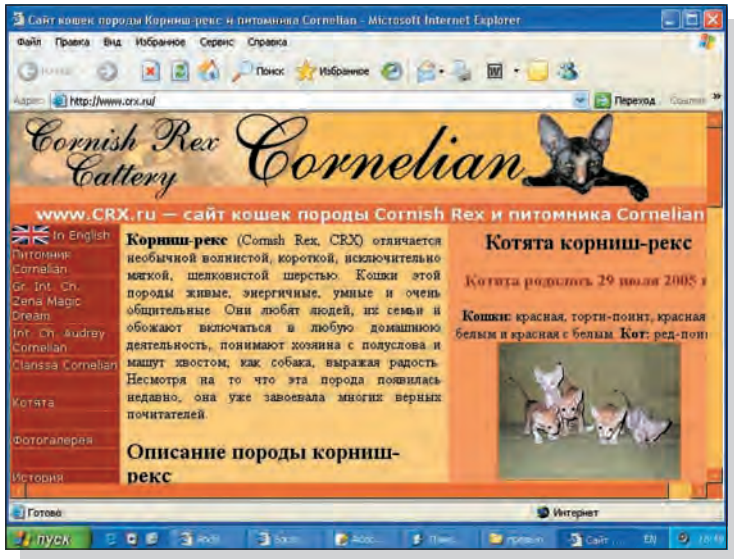
В основе технологии WWW лежат три составные части: язык гипертекстовой разметки HTML (HyperText Markup Language), универсальный способ адресации ресурсов в сети URL (Uniform Resource Locator) и протокол обмена гипертекстовой информацией HTTP (HyperText Transfer Protocol).

Тим Бернерс-Ли.

HTML

В 1986 г. Международная организация по стандартизации (ISO) приняла стандарт обобщённого метаязыка Standard Generalized Markup Language (SGML),

Разработкой и утверждением стандарта языка HTML занимается некоммерческая международная организация World Wide Web Consortium (W3C); в начале XXI в. широкое распространение получила четвёртая версия языка HTML.



Скриншот главной страницы сайта www.crx.ru.

позволяющего строить системы логической структурной разметки любых разновидностей текстов. Управляющие коды, вносимые в текст при такой разметке, лишь задавали его логическую структуру, т. е. создавали ссылки от документа к документу и не несли никакой информации о внешнем виде документа.

Разработчики WWW выбрали SGML в качестве основы для языка разметки гипертекстовых документов. Этот язык был назван HyperText Markup Language (HTML).

В отличие от прообраза, в котором гипертекстовые ссылки хранились в отдельных файлах специального формата, в HTML гипертекстовые ссылки, в том числе и управляющие внешним видом документа, внедрены в тело документа и хранятся в виде особых текстовых фраз, называемых *тегами*.

Гипертекстовый язык дополняет система единообразной адресации документов. Поскольку в WWW стали доступны ресурсы различного рода, для обращения к ним прежде приходилось применять специфическое программное обеспечение. Со временем эта задача упростилась, и сейчас требуется указать всего лишь адрес, а также вид сервера, т. е. протокол.

Наиболее часто URL используется для ссылок на страницы WWW: http://www.thehighwaystar.com, http://www.infomir.ru/KuMir/updates.html.

HTTP

Система WWW построена по традиционной схеме клиент — сервер. Взаимодействие между сервером WWW и клиентской программой осуществляется по протоколу передачи гипертекста HTTP.

В процессе взаимодействия клиент может запросить встроенную графику, принять и передать её параметры, получить новый адрес ресурса в сети и т. д.

Работа в WWW не ограничивается только получением уже готовых файлов с документами. Клиенту также предоставляется возможность запуска программ на сервере. При этом данные, выдаваемые этими программами, высылаются клиенту в виде документов. Так организована, например, работа через WWW с системами поиска информации и любыми интерактивными программами и системами, где требуется диалог с пользователем. Стандарт, по которому осуществляются запрос на исполнение программ и передача результатов клиенту, называется Common Gateway Interface (CGI).

ДОМАШНЯЯ СТРАНИЦА, САЙТ

Информация в системе WWW организована в виде страниц. Любой человек, имеющий доступ к Интернету, может разместить любую информацию в сети. К этой информации будет иметь доступ весь мир. Обычно поль-

зователи размещают информацию о себе на сервере своего провайдера или на серверах своих компаний. Такой способ предоставления информации называется домашней страницей (*англ.* home page). Также существуют серверы, бесплатно предоставляющие место под страницу. Адрес домашней страницы выглядит, например, так: www.infomir.ru/~agl.

Страницы системы WWW, находящиеся на одном сервере и объединённые по смыслу, образуют сайт (*англ.* site — «место», «участок»). Чаще всего сайты — это представительства компаний в Интернете. Также сайты имеют государственные структуры, общественные организации, те или иные информационные проекты. Частный пользователь Интернета тоже может организовать свой сайт. Адреса сайтов, в отличие от домашних страниц, выглядят, как правило, так: www.ds.ru.

ВОЙНА БРАУЗЕРОВ

Как и в системах удалённого доступа (TELNET) и передачи файлов (FTP), в World Wide Web существуют программы-клиенты, работающие на компьютере пользователя. Это так называемые программы просмотра гипертекстовых документов, или *браузеры* (от *англ.* to browse — «рассматривать»). Самые первые браузеры имели текстовый интерфейс, похожий на командную строку DOS. В начале 1993 г. в Национальном центре суперкомпьютерных программ (NCSA) Иллинойсского университета группой программистов под руководством Марка Андриссена был создан браузер Mosaic. Для того времени это была единственная профессионально написанная программа с удобным для пользователя графическим интерфейсом, работающая на различных компьютерных платформах. Mosaic сразу стала пользоваться огромным успехом и принесла своим создателям всемирную славу.

В 1994 г. разработчики Mosaic создали компанию Netscape Communications, и вскоре она выпустила первый коммерческий браузер —

ПЕРВЫЙ САЙТ

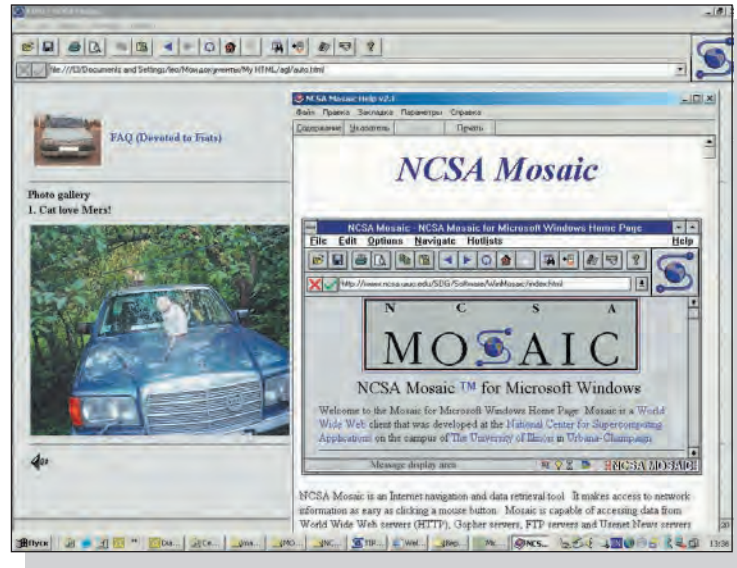
Пол Кунз.

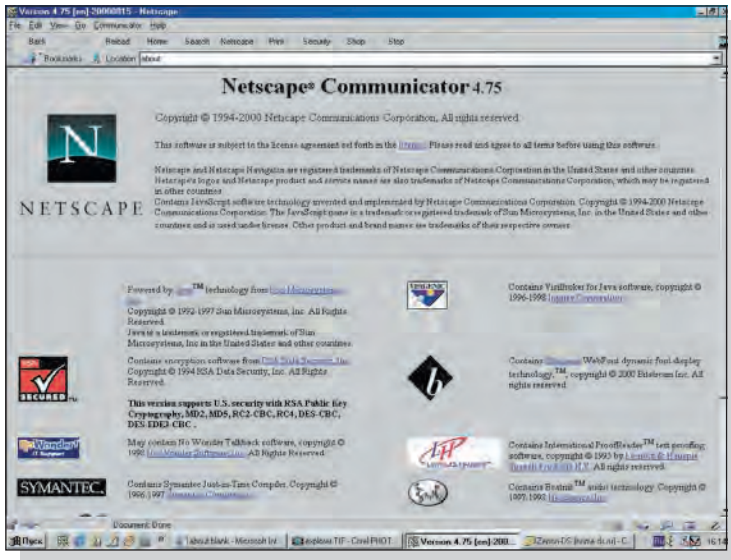
В декабре 1991 г. американский физик Пол Кунз из Станфордского университета создал первый сайт для системы World Wide Web, который открыл учёным из любой части света доступ к научным трудам университетского центра. В то время в базе данных университета содержалось 200 тыс. ссылок на статьи, и к ней обращались тысячи пользователей из 40 стран мира. Но нужна была технология, которая могла бы упростить и ускорить процесс поиска и получения необходимой информации. Именно для этого и пригодился сайт Кунза.

В январе 1992 г. Тим Бернерс-Ли продемонстрировал этот сайт на конференции во Франции, где присутствовало более 200 физиков из разных стран. Сначала он рассказал о системе WWW, а затем показал, как с её помощью можно связаться с заокеанским сайтом Кунза в США. Успех превзошёл все ожидания. Участники конференции были просто поражены тем, насколько просто и удобно можно получить нужную информацию.

Оформление сайта Кунза состояло из трёх строк текста и двух ссылок: одна на «записную книжку» с номерами телефонов и адресами электронной почты, другая на базу научных статей.

Окно браузера Mosaic.



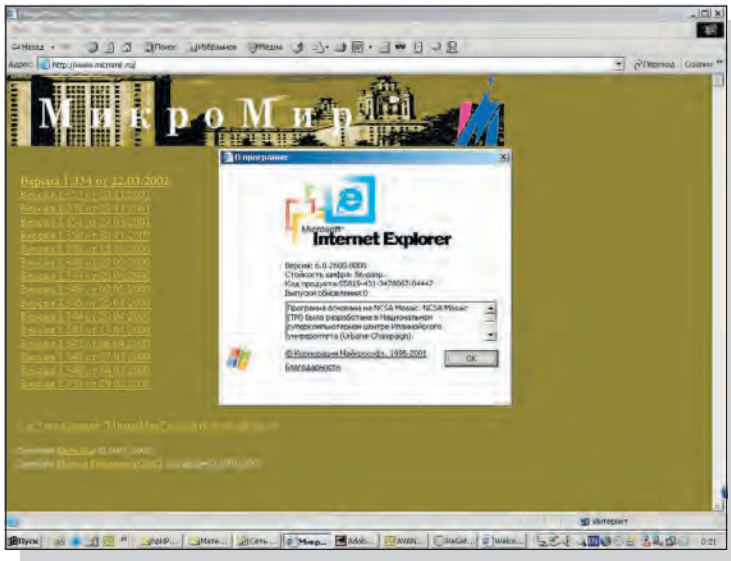


Окно браузера Netscape Communicator.

Netscape Navigator, который сразу же завоевал фантастическую популярность. Чтобы закрепить за собой лидирующее положение в этой сфере и привлечь новых пользователей, Netscape вводила в HTML новые усовершенствования, призванные улучшать внешний вид документа и расширять возможности форматирования. Новая версия языка поддерживалась только браузером Navigator. Такая политика принесла Netscape ещё больший успех — какое-то время доля Netscape Navigator в общем числе всех используемых браузеров составляла более 90 %. В третьей версии Netscape Navigator появились поддержка языка Java, встроенный язык сценариев JavaScript, возможность разбивки окна на кадры и другие новшества.

В 1995 г. на рынок браузеров, где господство Netscape, казалось, не оставляло шансов конкурентам, вступила корпорация Microsoft. Долгое время эта компания, привыкшая монопольно владеть своим сектором рынка, недооценивала перспективы Интернета и не собиралась как-либо участвовать в его развитии. Однако невероятный взлёт Netscape заставил руководство компании изменить своё мнение. Сначала трудно было поверить, что браузер Microsoft Internet Explorer, который тогда не представлял собой ничего выдающегося, сможет составить конкуренцию Netscape. Тем не менее выпущенная летом 1996 г. третья версия Internet Explorer, которая обладала оригинальным и привлекательным интерфейсом, понемногу стала теснить Netscape Navigator.

Решающим моментом в войне браузеров стал выпуск операционной системы Microsoft Windows 98 со встроенными средствами доступа к Интернету и очередной версией браузера Internet Explorer, являющегося ныне наиболее используемым в мире.



Сайт, открытый в браузере Internet Explorer.

ВЕБ-ДИЗАЙН

Рост WWW совпал с повсеместным распространением графических браузеров. Создатели сайтов с самого начала стали пользоваться всеми имеющимися возможностями по оформлению представляемой информации. Родилась новая отрасль деятельности — *веб-дизайн.*

В широком смысле веб-дизайн — это оформление веб-страниц. Он играет такую же роль для сайта, как вёрстка и полиграфический дизайн для книжного издания. Под веб-дизайном можно понимать не только создание графических элементов для сайта, но и проектирование всей его структуры, т. е. создание сайта целиком. Разработчику веб-дизайна необходимо знать стандарт HTML.

Гипертекстовый документ — это совокупность элементов, каждый из которых окружён тегами. По своему значению теги близки к скобкам begin – end («начало — конец») в языках программирования, которые задают области действия команд. Теги определяют область действия правил, по которым интерпретируются текстовые элементы документа. Например, тег <I> определяет область отображения курсива. Текст на языке HTML:

```
Этот текст <I>набран курсивом.</I>
```

Текст, отображаемый программой просмотра:

Этот текст *набран курсивом*.

В примере текст, который должен быть выделен курсивом, ограничен тегом начала стиля <I> и тегом конца стиля </I>. Конструкция перед содержанием текстового элемента называется тегом начала элемента и записывается так: <имя_элемента список_ параметров>. Конструкция, расположенная после содержания элемента, называется тегом конца элемента и записывается в таком виде: </имя_элемента>.

Для записи гипертекстовых ссылок в системе WWW используется URL. Например,

```
ссылка на <A HREF= http://www.
infomir.ru/press.html> пресс-релиз
фонда ИнфоМир.</A>
```

В этом примере элемент A (от *англ.* anchor — «якорь») использует параметр HREF, который обозначает гипертекстовую ссылку (Hypertext REFerence), для записи в форме URL. Данная ссылка указывает на файл press.html, хранящийся на сервере www.infomir.ru, доступ к которому осуществляется по протоколу передачи гипертекста HTTP. Как правило, ссылки выделяются подчёркиванием, а при наведении на них указатель меняет форму (например, указующий перст вместо стрелки).

Встраивание в документ рисунков осуществляется похожим тегом:

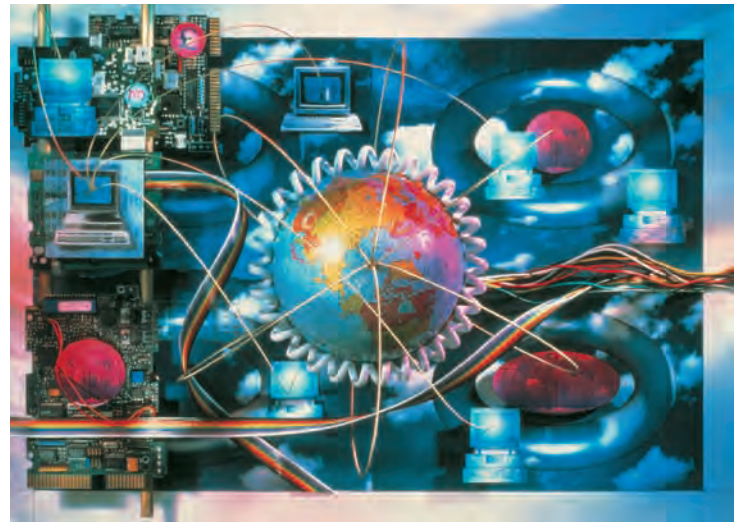
```
<IMG SRC=URL_рисунка>.
```

Допустимо (и желательно) применять такие параметры, как ширина (widht=), высота (height=) и описание (alt=) рисунка, а также его выравнивание (align). Автоматически рисунок выравнивается по нижнему краю сопровождающего его текста.

Сам гипертекстовый документ можно представить как один большой элемент с именем «HTML». Он состоит из двух частей: заголовка документа и тела документа, где содержится отображаемый на экране текст.

```
<HTML>
<HEAD>
Заголовок документа
</HEAD>
<BODY>
Тело документа
</BODY>
</HTML>
```

Форматирование происходит автоматически в процессе вывода документа на экран. Разбивка на строки зависит от ширины окна вывода, текущего размера шрифта, возможностей экрана и т. п., однако существует тег принудительного перевода строки
. Или, например, используя команду <HR>, можно разделить текст горизонтальной чертой. Чтобы отобразить заранее подготовленный текст, выключив форматирование программы просмотра, используют



Вам надо серьёзно задуматься об отдыхе, если вы представляете свою жену, как www.mylady.home.wife.

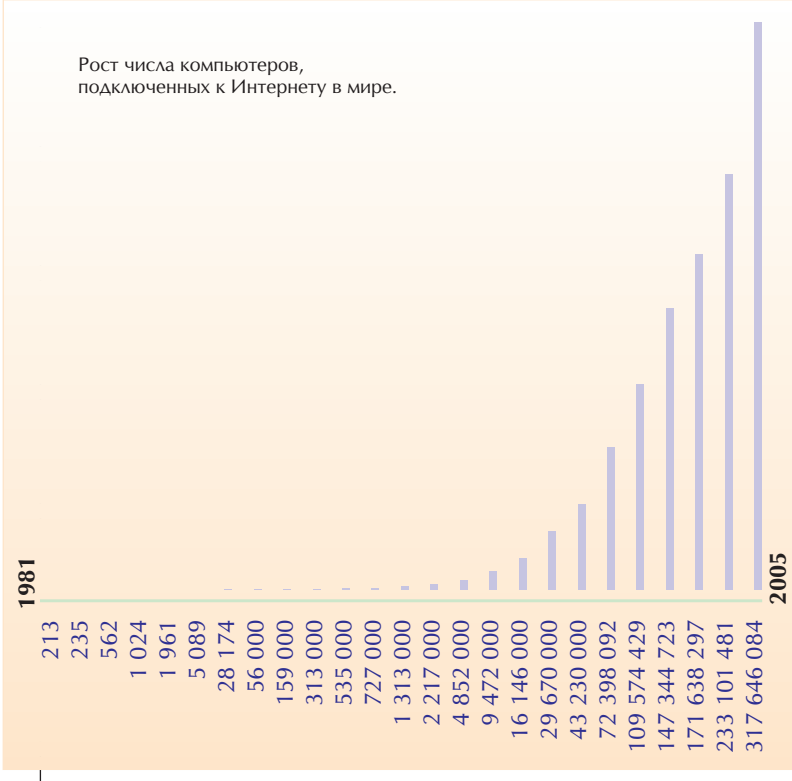
специальную пару тегов <pre> … </pre>. Есть теги для выделения текста жирным шрифтом, курсивом, подчёркиванием. Для более сложного форматирования применяются обычные и упорядоченные списки, таблицы, кадры. Существует возможность создать стиль для отдельных слов или предложений и документа в целом изменения этого стиля приведут к изменению сразу во всех документах, где он используется.

Полное описание языка HTML 4 (включая CSS-стили) — это многостраничный документ. Однако есть более лёгкий способ овладения языком гипертекстов — все программы просмотра обладают возможностью показа исходного кода WWW-страницы, поэтому можно учиться на примерах.

МАСШТАБЫ СЕТИ ИНТЕРНЕТ

Когда в 1969 г. были проведены первые эксперименты по объединению компьютеров в глобальную сеть, прообраз сети Интернет, в неё входило всего три компьютера. Однако Интернет непрерывно растёт. Его масштабы определяют, подсчитывая число компьютеров, подключённых к сети. На графике приведены данные организации Internet Software Consortium.

Если количество компьютеров подсчитать достаточно легко, то оценить число пользователей Интернета удаётся лишь приблизительно. Действительно, один человек может иметь доступ к Интернету с нескольких компьютеров, и, наоборот, один компьютер может предоставлять доступ к сети многим пользователям.



В 1996 г., по усреднённым данным, доступ к Сети имели около 50 млн человек. В 1999 г. их количество увеличилось до 200 млн, а к 2005 г. достигло миллиарда. Большинство из них (185 млн) — пользователи из США (60 % населения страны). В Великобритании 33 млн человек посещают Интернет, в Японии — 78 млн, в Китае — 99 млн, в Германии 42 млн, в Канаде — 20 млн.

В России в 2005 г., по данным фонда «Общественное мнение», насчитывалось более 17 млн пользователей Интернета, из них 10,3 млн посещают Интернет не реже одного раза в неделю. Резкий рост популярности Интернета был связан с появлением Всемирной паутины — WWW.

В 1993 г. насчитывалось около 130 WWW-сайтов, в 1994 г. 9 тыс. (тогда их ещё можно было сосчитать), в 1996 г. их число перешагнуло рубеж в 100 тыс., а в 1997 г. — в 200 тыс.

Подсчитать точное количество сайтов, появившихся к 2005 г., невозможно; по разным оценкам, оно составляет от нескольких десятков миллионов до почти миллиарда. Число их страниц — около 8 млрд. Есть оценка специалистов, что в 2005—2006 гг. число пользователей Интернета превысило 1 млрд. Их количество будет постепенно приближаться к числу людей, пользующихся телефоном, телевизором, водопроводом...

Число подключений к Интернету также будет расти. И не только за счёт компьютеров, но и других бытовых устройств, таких, как холодильник, или кондиционер, которыми можно будет управлять через Интернет.

КОММЕРЦИЯ В ИНТЕРНЕТЕ

На протяжении всей истории человечества бизнес почти всегда ставил себе на службу вновь изобретённые технологии вне зависимости от целей, для которых они были созданы. Так обстояло дело и с железными дорогами, и с телеграфом. Так стало и с Интернетом, который разрабатывался для решения научных и военных задач. Когда распространение Интернета достигло определённого уровня, на него обратили внимание коммерсанты.

Уже в середине 80-х гг. XX в. появились компании, предоставляющие платный доступ к электронной почте, а с развитием системы World Wide Web — полноценный доступ в Интернет. Таким образом, возникла самостоятельная отрасль экономики, которая поддерживает и обслуживает Интернет, обеспечивает доступ к нему пользователей.

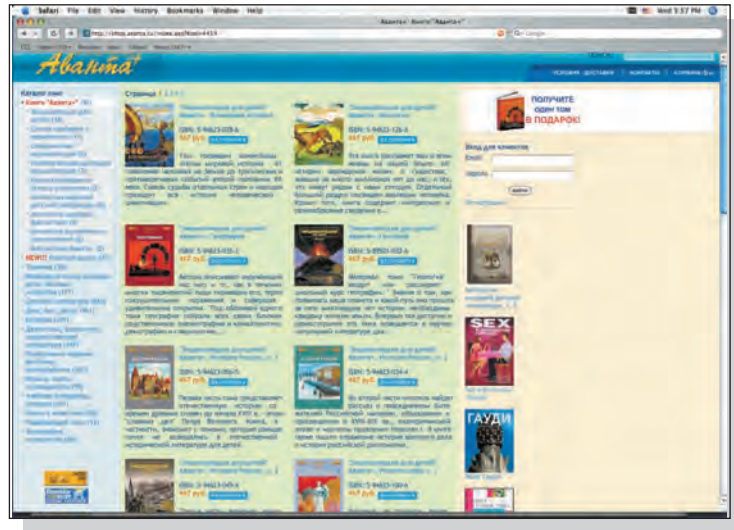
Помимо бизнеса, который осуществляет собственно поддержку Интернета и оказывает пользователям сопутствующие услуги, такие, как web-дизайн, Интернет взаимодействует и с традиционным бизнесом, являясь одним из его инструментов. В обиход вошло понятие *электронной коммерции* (e-commerce).

> Электронная коммерция — это набор технических и организационных форм ведения коммерческой деятельности и совершения сделок с использованием Интернета как средства взаимодействия с партнёрами, банками, поставщиками и потребителями товаров и услуг.

ВИДЫ ЭЛЕКТРОННОЙ КОММЕРЦИИ

Любые деловые операции, при которых участники взаимодействуют с помощью электронных систем и Интернета, а не путём прямых контактов, можно смело причислять к сфере электронной коммерции.

Как правило, в системах электронной коммерции присутствуют все этапы совершения сделки: поиск требуемой продукции или услуг, уточнение деталей сделки, оплата, получение (доставка) заказа.



Скриншот со страницы интернет-магазина издательства «Мир Энциклопедий Аванта+».

САЙТ eBay.com

Это огромная система аукционной торговли в Интернете, пользователи которой могут продавать и покупать различные вещи: антиквариат, игрушки, монеты, компьютеры, марки, памятные знаки, открытки и ювелирные изделия. Каждый день сайт посещают десятки тысяч человек. Система рассчитана в основном на индивидуальных пользователей, а не на фирмы. Ежедневно выставляется до полумиллиона лотов в нескольких тысячах категорий. Например, введя в окно поиска запрос, можно получить список более чем из тысячи наименований товаров, связанных с творчеством легендарной рок-группы Deep Purple: компакт-диски, винилове пластинки, плакаты, футболки, автографы самих музыкантов и т. п.

За счёт чего владельцы eBay получают доходы? Продавцы оплачивают размещение информации о своих товарах на сайте, а покупатели посещают его, чтобы приобрести интересующие их предметы. Каждый товар обычно выставляется на несколько дней. При совершении сделки владельцы eBay выставляют продавцу счёт на определённый процент от продажной цены. При этом не требуется складов для хранения товаров, не нужно арендовать торговые помещения, как на обычных аукционах, нет необходимости нанимать людей для их проведения. В сущности, eBay — это огромный посредник, основное достояние которого — репутация и популярность. Если нужно продать или купить что-то более или менее редкое, то стоит в первую очередь зайти на eBay.com. Рассказывают, что основатель Пьер Омидьяр создал этот сайт по просьбе жены, желавшей пополнить свою коллекцию редких игрушек, которые нельзя было купить в магазинах.

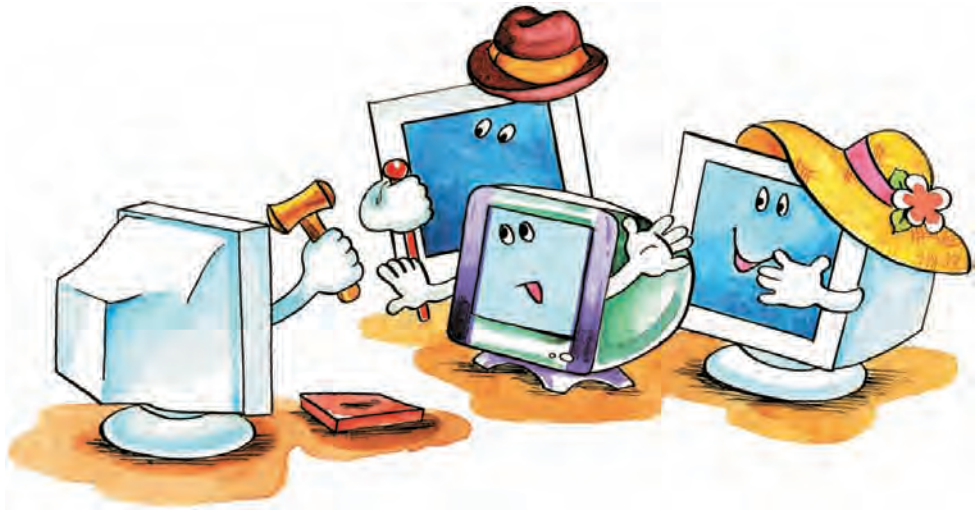
Электронную коммерцию принято разделять по типам отношений, устанавливаемых в процессе сделки.

В настоящее время наибольший объём торговых операций в Интернете осуществляется в сфере отношений *business-to-business* (B2B). Типичным примером B2B является взаимодействие крупных компаний с поставщиками. Некоторые компании в сфере высоких технологий не только закупают через Интернет почти все необходимые комплектующие, но и осуществляют большинство продаж. Технологии B2B развиваются весьма динамично. Многие крупнейшие мировые корпорации планируют в будущем полностью перевести в электронную форму все взаимоотношения с поставщиками.

Однако наиболее распространённая модель электронного бизнеса — *business-to-consumer* (B2C). Это хорошо известные Интернет-магазины, электронные каталоги, столы заказов и представительства компаний в Интернете. В первых Интернет-магазинах (наиболее известный из них — Amazon.com) продавались книги, программное обеспечение, компьютерные комплектующие, аудиозаписи (CD) и видеофильмы (VHS и DVD). Впоследствии посетители стали приобретать и более дорогие товары, например антиквариат, произведения искусства, ювелирные изделия, мебель, предметы роскоши. Покупатели отдают предпочтение тем товарам, характеристики которых обычно известны им заранее. Часто достаточно знать имя производителя, чтобы принять решение о покупке. А вот приобретение деликатесов или одежды и обуви без примерки требует от покупателя определённой смелости.

С другой стороны, посещение сайта Интернет-магазина не может воссоздать тех ощущений, которые возникают у человека при посещении реального магазина, во время прогулок по торговым залам, при разглядывании и выборе товаров. Интернет-магазин не может дать и чувства мгновенного удовлетворения от обладания только что приобретённым товаром.

Электронные коммерческие контакты между частными лицами называются *consumer-to-consumer* (C2C). Типичный пример — электронный аукцион, на котором одни частные лица выставляют на продажу различные предметы, предназначенные для по-



Интернет-магазин представляет собой коммерческую организацию, осуществляющую контакт с потребителем через сайт, на котором выставлены предложения товаров или услуг и предусмотрена возможность их заказа посетителями, находящимися непосредственно в Сети.

купки другими частными лицами. Один из наиболее известных и посещаемых электронных аукционов — eBay.com.

Оплата товаров и услуг, предложенных в Интернете, осуществляется самыми разными способами: наличными, наложенным платежом при получении товара по почте, банковским переводом, почтовым переводом, кредитной картой. Последний способ наиболее быстрый, однако покупатели часто справедливо опасаются передавать через Интернет реквизиты своих кредитных карт, боясь, что данные могут быть похищены из Сети злоумышленниками. Эти опасения являются одним из препятствий для осуществления электронной торговли. Чтобы преодолеть их, открывают специальные «электронные» счета, которые не сопряжены со счётом на кредитной карте. Важно также условие повышения безопасности электронных платежей.

КАКИЕ ТОВАРЫ ЛУЧШЕ ВСЕГО ПРОДАВАТЬ ЧЕРЕЗ СЕТЬ

Как показали исследования, с помощью Интернета чаще всего покупаются товары текущего потребления. Дорогие, давно планируемые покупки делаются здесь значительно реже. Ввиду особенностей доставки заказов очевидно, что в Сети легче продавать компактные, удобные для транспортировки товары, чем тяжёлые и громоздкие, осложняющие доставку.

Идеально соответствуют возможностям сетевой торговли товары, которые можно доставить покупателю через Интернет. Наиболее очевидным примером таких товаров является программное обеспечение. Книги и газеты также могут приходить к заказчикам в электронном виде. В общем, всё, что может быть оцифровано, в том

Вам немедленно следует пойти на вечеринку, если в Интернете у вас друзей больше, чем в реальной жизни.

числе музыка, графические изображения, видеоматериалы, оптимально для продажи в Сети. С конца 90-х гг. XX в. растёт популярность продажи через Интернет авиабилетов, туристических, биржевых, страховых и банковских услуг.

Тем не менее многие предприимчивые сетевые торговцы утверждают, что не существует такого товара, который нельзя было бы продать при помощи Интернета.

ПЕРСПЕКТИВЫ И ПРЕИМУЩЕСТВА ЭЛЕКТРОННОЙ КОММЕРЦИИ

В перспективе электронный бизнес может не только создать серьёзную конкуренцию традиционным каналам сбыта товаров, но даже полностью их вытеснить. Бурное развитие информационных технологий постоянно создаёт новые возможности. Наиболее существенными являются увеличение скорости передачи данных в Интернете и возможность передачи больших объёмов информации.

С появлением широкополосных каналов связи и подключением к ним всё большего числа индивидуальных пользователей в первое десятилетие XXI в. ожидается ускоренное развитие электронной коммерции.

Кроме того, значительное влияние на развитие электронной коммерции оказывает то обстоятельство, что мобильные телефоны и другие бытовые электронные устройства начинают оснащаться средствами подключения к Интернету, нарушая тем самым монополию персональных компьютеров.

По оценкам западных экспертов, реклама на телевидении и радио приносит пока более ощутимый результат, чем реклама в Интернете, однако Интернет по тем же показателям уже «победил» печатные средства информации — газеты и журналы.

Экономические возможности электронной коммерции гораздо шире, чем кажется на первый взгляд. Преимущества электронной коммерции заключаются не только в сокращении расходов на информирование покупателей о предоставляемых услугах и устранении ненужных посредников. Продажа книг через Интернет, например, позволяет избавиться от огромных книжных магазинов — на сервере Amazon.com заказы передаются непосредственно на оптовый склад.

Существенное достоинство электронной торговли в том, что продавец обычно вначале получает деньги от покупателя, а уже потом рассчитывается со своим поставщиком, в то время как в обычной торговле всё происходит наоборот. Таким образом, отсутствует необходимость привлечения большого оборотного капитала.

У электронной коммерции имеются и другие экономические преимущества. Например, покупатель выбирает наиболее подходящую цену. Цены гораздо проще сравнивать, нажимая на клавишу «мышки», чем обходя пешком бесконечную череду магазинов. Предположительно Интернет таким образом будет препятствовать искусственному завышению цен.

Интернет удобен и с точки зрения развития бизнеса. С помощью всего одного сайта можно осуществлять операции по всему миру: нужно только создать надёжно функционирующую систему, в каждый момент готовую к исполнению заказов посетителей сайта.

Можно расширять бизнес до любых масштабов, невзирая на территориальную удалённость клиентуры и разни-

цу часовых поясов. Интернет также позволяет создавать объединения пользователей и проводить для них специальные аукционы, которые невозможны вне Сети.

Весьма серьёзным является вторжение Интернета в сферу традиционных интересов туристических компаний. Отдыхающие начинают бронировать и выкупать авиабилеты через Интернет непосредственно у авиакомпаний, таким же образом они заказывают и оплачивают номера в гостиницах, минуя турагентства.

Активно наступает Интернет и на позиции торговцев книгами: многие оптовики занялись розничной торговлей, а всё расширяющаяся практика электронной публикации книжных новинок подрывает бизнес не только розничным торговцам, но и оптовикам и даже издателям. Такая же ситуация и с продажей аудиозаписей: меломаны могут получать через Сеть высококачественные копии последних музыкальных новинок.

Максимальное воздействие на общество Интернет, как и всякая технологическая революция, окажет тогда, когда с его помощью возникнут и широко распространятся абсолютно новые товары и услуги, а существующие обретут качественно новое воплощение. Борьба между традиционными и электронными формами коммерции только начинается.

ПОИСК В ИНТЕРНЕТЕ

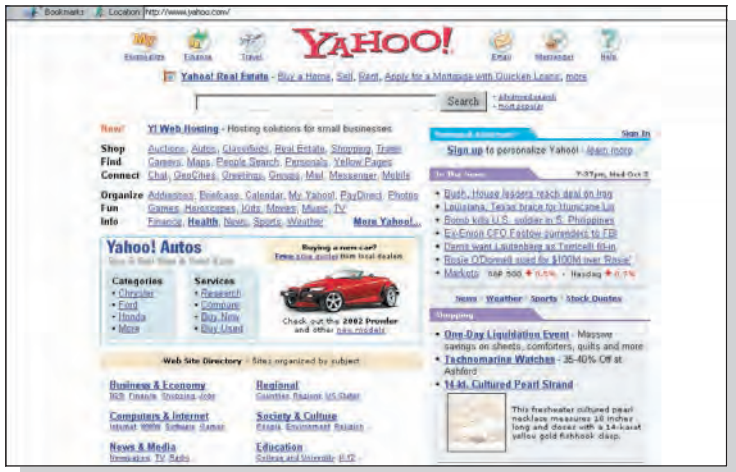
ЧТО МОЖНО НАЙТИ В ИНТЕРНЕТЕ?

Объём информации, содержащейся в Интернете и, в частности, в его олицетворении — системе World Wide Web, не поддаётся измерению. Можно лишь оценить его порядок. По всей вероятности, он составляет несколько десятков или даже сотен терабайтов (миллионов миллионов байтов). Это информация самого разного характера и направления.

В Интернете можно найти самые свежие новости — политические, культурные, экономические; научную, техническую, образовательную и справочную информацию абсолютно любого рода; рекламу разнообразных товаров и услуг. Это и ресурсы, посвящённые культуре и искусству: литературе, живописи, театру, кино, телевидению.

Кроме того, в Интернете находятся десятки и сотни тысяч развлекательных сайтов различной тематики: юмор, игры, путешествия, спорт. Каждый день в Интернете появляется порядка миллиона новых WWW-страниц, но при этом некоторое количество существующих страниц бесследно исчезает.

Таким образом, в Интернете есть всё. Задача пользователя состоит в том, чтобы найти желаемую информацию. Современный Интернет можно представить как библиотеку, книги в которой расположены без какого бы то ни было порядка; нет ни единой системы каталогов и классификаторов, ни библиотекарей. Посетители «библиотеки» время от времени добавляют новые тома или безвозвратно забирают их.



Окно Yahoo!

Для того чтобы извлечь полезную информацию из Интернета, нужно знать, где и как вести поиск. Этой проблемой люди озаботились почти одновременно с началом широкого распространения WWW. Тогда появились первые *каталоги ресурсов* и *поисковые системы.*

КАТАЛОГИ РЕСУРСОВ

В начале 90-х гг. XX в., когда серверы WWW ещё можно было сосчитать, существовали так называемые *отправные точки* (starting point). Это были страницы, на которых перечислялись ссылки на все WWW-серверы, сгруппированные по тематическому либо географическому признаку. Например, «Все WWW-серверы по биохимии» или «Все WWW-серверы Финляндии». (Сейчас вместо термина «WWW-сервер» чаще говорят «сайт».) Таким образом, открыв нужную отправную точку, можно было последовательно обойти все ссылки. Отправные точки стали прообразом современных каталогов ресурсов Интернета.

Каталоги ресурсов подобны справочникам, содержащим систематизированные ссылки на сайты. Ссылки объединяются в группы по определённым признакам, как правило по тематике. Каждая группа может иметь несколько уровней, т. е. каталоги имеют древовидную структуру. Кроме того, каталоги обеспечивают разнообразный дополнительный сервис: поиск по ключевым словам, списки последних поступлений, списки наиболее интересных из них. Ссылки в каталоги вносят администраторы, которые стараются сделать свои коллекции наиболее полными, включающими все доступные ресурсы на каждую тему. Также ссылки на свои ресурсы предлагают и владельцы. Администраторы каталога проверяют ссылку и вносят её в соответствующий раздел.

Старейший каталог ресурсов Yahoo! (www.yahoo.com) состоит из 14 основных разделов:

- Arts & Humanities (искусство и гуманитарные науки);
- Business & Economy (бизнес и экономика);
- Computers & Internet (компьютеры и Интернет);
- Education (образование);
- Entertainment (развлечения);
- Government (ресурсы правительства США);
- Health (здоровье);
- News & Media (новости и СМИ);
- Recreation & Sports (отдых и спорт);
- Reference (справочная информация);
- Regional (ресурсы по регионам);
- Science (естественные науки);
- Social Science (общественные науки);
- Society & Culture (общество и культура).

Поиск с помощью каталога ресурсов выглядит следующим образом: пользователь определяет, к какой теме относится разыскиваемая информация; передвигаясь в глубь по дереву, находит в каталоге соответствующий раздел; обходит все страницы, перечисленные в разделе.

Например, надо найти информацию о кошках породы корниш рекс.

Чаще всего адрес сайта соответствует названию компании или тематике. Если обратиться по ссылке www.имя_компании. com, то с довольно большой вероятностью откроется сайт именно этой компании. Например, www.adobe.com — сайт корпорации Adobe, а www.cnn.com — сайт медиакомпании CNN. Если искомая компания не американская, можно попробовать добавить обозначение национального домена — например, www.infomir.ru — адрес сайта фонда «ИнфоМир», а www.ferrari.it принадлежит итальянскому концерну Ferrari.

Тогда можно, постепенно углубляясь, перейти в раздел Science > Biology > Zoology > Animals, Insects, and Pets > Mammals > Cats > Breeds > Cornish Rex. Если же надо найти, кто продаёт котят этой породы, поможет раздел Business and Economy > Shopping and Services > Animals > Cats > Breeders > Cornish Rex.

К удобству применения каталогов ресурсов относится то, что, если пользователю известна тема искомого документа, он будет исследовать соответствующую ветвь каталога, не отвлекаясь на посторонние, не относящиеся к делу документы.

Однако объём каталога ограничен возможностями его администраторов и их субъективностью в выборе материала. Кроме того, тематику искомого документа не всегда можно сформулировать в пределах классификации каталога. В этом случае на помощь приходят поисковые системы.

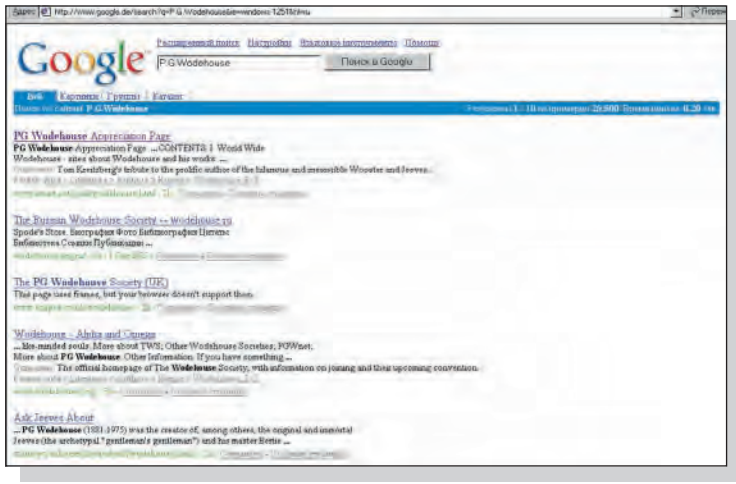
ПОИСКОВЫЕ СИСТЕМЫ

Действие поисковых систем заключается в постоянном, последовательном изучении всех сайтов Интернета, доступных данной системе поиска, а также всех тех страниц, на которые есть ссылки. В связи с постоянным обновлением информации поисковые системы регулярно возвращаются через определённый срок к уже изученным страницам, чтобы обнаружить и зарегистрировать изменения. Все прочитанные страницы индексируются, т. е. создаётся специализированная база данных, в которой записаны все исследованные системой сайты. Кроме того, владельцы сайтов могут самостоятельно добавить в поисковые системы ссылки на них.

При поступлении запроса от пользователя машина поиска просматривает свой индекс и выдаёт список страниц в Интернете, которые соответствуют критериям поиска. Найденные документы, как правило, упорядочиваются в зависимости от местоположения ключевых слов (в заголовке, в начале текста), частоты их появления в тексте и других характеристик.

Существует множество сайтов, предназначенных для поиска. Одни из них пользуются собственным индексом, другие, называемые метапоисковыми, их не имеют, а собирают и упорядочивают результаты из множества других поисковых узлов. Несмотря на схожий принцип работы, поисковые системы различаются по языкам запроса, зонам поиска, глубине поиска внутри документа, методам упорядочивания и приоритетам, поэтому применение разных поисковых систем даёт различные результаты. Старейшие поисковые системы — Lycos (www.lycos.com) и AltaVista (www.altavista.com).

Окно поисковой системы AltaVista.



Окно поисковой системы Google.

К достоинствам поисковых систем можно отнести исследование ими огромного объёма информации и её периодическую актуализацию. Однако при этом не учитываются документы, не содержащие ключевых слов, а, с другой стороны, в списке содержится много не относящейся к делу информации, отсеивание которой занимает немалое время.

Приступая к поиску, пользователь вводит одно или несколько ключевых слов и выбирает тип поиска. В большинстве поисковых систем есть три основных типа поиска: по любому из слов, по всем словам и точно по фразе. В зависимости от этого результаты могут сильно различаться.

Кроме того, поисковой системе можно попробовать задать вопрос на простом разговорном языке, например: «Каковы цены на услуги страхования недвижимости?».

БУДУЩЕЕ ПОИСКА

Объём накопленной информации увеличивается с каждым годом. Хотя долгое время его рост был экспоненциальным, тем не менее объём полезной информации в Интернете, которая может быть потреблена пользователями, и объём информации, которую там можно разместить, скорее всего, конечны. Поэтому от механизмов поиска требуется постоянное повышение эффективности, увеличе-

КАК ИСКАТЬ

Поиск по любому из слов
В результате поиска составляется список всех страниц, содержащих любое из ключевых слов. Нередко число совпадений при таком поиске огромно. Однако если поисковый узел хорошо сортирует результаты по тематике, то нужную страницу можно найти в верхней части списка. Поиск по любому слову может быть удобен в случаях, когда пользователь не уверен в ключевых словах.

Поиск по всем словам
В этом режиме поиска формируется список всех страниц, содержащих все ключевые слова в любом порядке. При этом сохраняется вероятность получения результатов, не соответствующих теме.

Поиск точно по фразе
В этом режиме поиска составляется список всех страниц, содержащих фразу, точно совпадающую с ключевой; знаки препинания игнорируются. В список не попадают те страницы, которые посвящены нужной теме, но в строке поиска описываются с использованием других фраз. Даже в этом режиме поиска возможны ложные результаты.

Сложный поиск
Многие сайты поисковых систем оснащены дополнительными функциями поиска. Вот некоторые из них.

Поиск с использованием языка запросов. Позволяет строить запрос с использованием логических операторов И, ИЛИ, НЕ; включать или исключать слова или фразы.

Поиск с использованием категорий. Индексированные страницы упорядоченно хранятся в многоуровневом каталоге категорий. Можно просто просмотреть каталог категорий в поисках нужных материалов или дойти до определённого уровня и затем провести поиск в рамках выбранной категории. Таким образом поисковая система объединяется с каталогом ресурсов.

Поиск среди обнаруженных страниц. Если найдено много страниц, то, добавив ещё одно ключевое слово повторим поиск. Во многих поисковых системах есть более быстрая функция поиска среди найденного.

Точная настройка параметров поиска. Можно указать язык, ограничиться сайтами в определённом домене, провести поиск только по заголовкам, просмотреть лишь сайты, информация на которых была обновлена в течение определённого периода, или показать страницы, на которых существуют ссылки на заданную страницу.

ние мощности вычислительных ресурсов.

Но совершенно необязательно, что та информация, которая была найдена вчера, сегодня останется на том же месте и не будет уничтожена. Уже сейчас многие материалы, когда-то расположенные в Сети, стали недоступны. Это открывает новые рубежи для поиска — археология (раскопки утраченной информации) в Интернете. Частные лица и организации пытались организовать хранение абсолютно всего в Интернете, что при

ГДЕ ИСКАТЬ

Google (www.google.com)
Самая быстрая и самая большая поисковая система. Проиндексировано более 8 млрд страниц (из них полностью — около половины, остальные представлены только в виде адреса и текста ссылки). Имеется возможность выбора языка интерфейса. Можно включать или исключать результаты с определённых сайтов или доменов. В отличие от большинства поисковых систем Google также оценивает популярность ресурса по количеству ссылок, ведущих к нему с других страниц. Кроме того, здесь содержится архив с возможностью поиска по всем телеконференциям системы USENET за последние 20 лет и поиск изображений.

Yandex (www.yandex.ru)
Наверное, лучшая из поисковых систем отечественного производства. Завоевала популярность умением проводить качественный поиск с учётом словоформ русского языка. Индексирует в основном русскоязычные ресурсы, при этом по возможностям не уступает зарубежным системам. Поиск можно осуществлять точно по слову или по любым его словоформам с ограничением по дате, с указанием сайта или его поддиректории. Можно вести поиск с учётом так называемого индекса цитируемости, задавать язык документа.

AltaVista (www.altavista.com)
Одна из старейших поисковых систем. Предоставляет расширенный круг критериев поиска: например, в разделе Advanced search есть выбор отрезка времени, к которому относится дата создания или изменения ресурса, поддержка 25 языков; имеется возможность выдачи одного результата на сайт (это сужает круг поиска без ущерба для качества).

Yahoo! (www.yahoo.com)
Один из первых каталогов ресурсов в Интернете. На Yahoo! составлен большой структурированный каталог категорий. Сначала поиск осуществляется в них, потом в собственном архиве, потом — с использованием поисковой системы. Поиск в категориях даёт хорошие результаты: их немного и соответствие критериям наиболее полное. Помимо стандартного набора функций позволяет отбирать ресурсы по дате, поддерживает возможность постановки знака «*» вместо любой последовательности символов в ключевых словах.

Рамблер (www.rambler.ru)
Одна из первых русских поисковых систем. Кроме стандартных возможностей поиска на сайте имеется рейтинг-каталог ресурсов.

существующем развитии техники теоретически возможно, но пока ещё дорого. Ожидается, что с появлением нанотехнологий задача будет решена. А пока всё, что может заинтересовать, надо сразу копировать себе на компьютер.

Ещё одна проблема состоит в том, что пользователям бывает сложно сформулировать, что же именно они хотят найти. Таким образом, можно предположить, что одной из особенностей будущих поисковых систем будет возможность предоставления информации по нечётко сформулированным запросам. Например, поиск изображений. Современные поисковые системы не умеют находить изображение по его описанию, а пытаются сделать это по контексту или даже просто по имени файла. Поисковой системе будущего достаточно будет сказать: «Найди фотографию девушки, похожей на рафаэлевскую Мадонну».

«ЭЛЕКТРОННОЕ ПРАВИТЕЛЬСТВО»

Идея создать «Электронное правительство» зародилась в 1999 г., когда был открыт всеевропейский проект Government Online. Он объединил правительства более 20 стран. Основной целью этого проекта является совместное сотрудничество по установке прямого электронного сервиса (*англ.* On-Line) между правительством, гражданами и бизнесом.

Технологии, привнесённые появлением Интернета, не обошли ни одну из сторон человеческой жизни и на рубеже XXI в. добрались и до государства.

К счастью, «Электронное правительство» (*англ.* E-Government) не имеет ничего общего с излюбленным будущим фантастов, в котором компьютерный супермозг захватывает Землю и устанавливает тиранию роботов, угнетая и планомерно истребляя несчастное человечество, как это происходило в фильме «Терминатор».

«Электронное правительство» — это не компьютерная система автоматизированного управления. «Электронное правительство» — набор Интернет-сайтов и открытых баз данных

государственных органов и правительства. Это правительственные программы, направленные на улучшение обслуживания населения, в частности социальные электронные карты и дистанционное образование.

«Электронное правительство» по замыслу создаётся прежде всего в помощь гражданам. Благодаря коммуникации, предоставляемой Интернетом, «Электронное правительство» стремится обеспечить политическую гласность, сократить до минимума чиновничью волокиту и бюрократизм, долгие и подчас бесполезные походы по государственным органам. Важнейшей частью всех предлагаемых услуг «Электронного правительства» является предоставление информации населению: ежедневные новости и юридические консультации, базы данных по законодательству.

На сайтах правительственных органов созданы группы телеконференций. Это способ вовлечения населения в политические дискуссии, участие в политической жизни страны, так как высказанные на конференциях мнения граждан внимательно изучаются и анализируются соответствующими правительственными службами. Можно сказать, что граждане участвуют в политической жизни, не выходя из дому.

Электронное заявление и налоговая декларация могут быть отправлены по Сети из любой точки планеты. Несомненно, это гораздо удобнее, чем лично посещать чиновников, налого-

вых инспекторов, посылать письма по почте.

Правда, традиционно письма отправляются в рукописном или печатном виде, а не электронном: «написанного пером не вырубишь топором». Гражданин доверяет бумаге, почте, уведомлению о вручении корреспонденции. Посланную бумагу не выбросишь! Но средства доставки электронной почты также могут гарантировать получение и регистрацию электронных писем, причём более надёжно и оперативно, чем почта или секретариат, регистрирующий письма граждан. А законодательство приравняет электронное письмо к обычному обращению граждан. Более того, гражданину не потребуется обзванивать инстанции, узнавая, как продвигается его дело. Это можно «увидеть» через Интернет с помощью своего компьютера.

Средства доставки электронной корреспонденции кажутся более уязвимыми — легко похитить конфиденциальную информацию. К Сети, как и к телефонной линии, может кто угодно подключиться и «подслушать» линию, украсть информацию, узнать, например, о доходах гражданина. Однако при передаче секретных данных применяются алгоритмы шифрования. При этом если преступник и сумеет «подслушать» линию передачи, он всё равно не сможет извлечь никакой полезной информации из шифрованного сообщения. Это будет просто абракадабра, и налоговая декларация не попадёт в руки злоумышленника.

Важным принципом электронных услуг будущего является One-Stop Service, который позволит обсудить просьбу гражданина на разных бюрократических уровнях порой без участия самого гражданина. Не потребуется сбора подписей на прошении, хождения по инстанциям. Всего лишь достаточно подать электронное заявление по E-mail, и «бумага» сама пройдёт по всем уровням управления. В идеале гражданин сможет проконтролировать прохождение бумаги прямо из дому. Каждое обращение получает уникальный номер. Этот номер-код пересылается гражданину с сообщением, что его письмо принято к рассмотрению и зарегистрировано. Используя этот код, нетрудно проследить всю историю прошения.

Очередей можно избежать!

Часто, получив «визу» чиновника, приходится отправляться в другой государственный орган за следующей. One-Stop Service и решает эту проблему. Письма граждан пересылаются между чиновниками и даже различными департаментами в электронном виде. Гражданин получит официальный ответ по E-mail, который имеет статус такого же документа, как и бумага с печатями.

Последний год XX века ознаменовался рождением «Электронного правительства». До полного успеха ещё далеко. Требуется принять соответствующие юридические акты, тем более что в разных странах уровень развития средств коммуникаций (Интернет) и обеспечение департаментов правительств соответствующими программными системами (системы учёта и делопроизводства, базы данных) и компьютерами различены. Одной из ведущих стран в этой сфере является Австрия.

ФЕДЕРАЛЬНАЯ ЦЕЛЕВАЯ ПРОГРАММА (ФЦП) «ЭЛЕКТРОННАЯ РОССИЯ 2002—2010»

К сожалению, Россия в области использования Интернет-технологий занимает далеко не лидирующее положение. По данным исследований международных организаций на 1999 г., она является единственной страной в мире, где количество компьютеров в школах сокращается. Менее 2 % школ имеют локальные сети, и лишь 1,5 % образовательных учреждений обладают доступом в глобальную сеть Интернет.

ФЦП — своеобразная попытка прорваться в «цифровое будущее», которое для развитых стран уже сделалось реальностью.

Успех программы будет означать в первую очередь, что органы власти всех уровней станут более доступными для граждан, а их деятельность — более прозрачной и эффективной. Прозрачности планируется достичь за счёт обязательной электронной публикации и создания баз данных по всем документам, не имеющим статуса секретных, эффективности — за счёт перевода в электронную форму значительных объёмов документооборота.

Для того чтобы при помощи информационных технологий приблизить российскую систему образования к стандартам развитых стран Запада, разрабатывается программа «Развитие единой образовательной информационной среды на 2002—2006 гг.». Эта программа предусматривает, в частности, компьютеризацию школ и распространение Интернет-

ВИРТУАЛЬНОЕ ОБРАЗОВАНИЕ В ЕВРОПЕ

С 2001/02 учебного года началась усиленная компьютеризация австрийских школ всех уровней (начальная и средняя школы, гимназия, школа с профессиональным уклоном). Муниципалитет Вены выделил около 17 млн долларов для 400 школ на приобретение персональных компьютеров и программного обеспечения и использование Интернета.

Министерство образования Австрии планирует широкое внедрение в школах переносных компьютеров, которые должны в ближайшем будущем заменить учебники, тетради, ручки и другие вспомогательные средства обучения.

В целом на компьютерную модернизацию австрийских школ, приобретение различного оборудования и обеспечение свободного доступа в Интернет, а также на повышение квалификации учителей будет потрачено около 70 млн долларов. В университетах предполагается, что все студенты будут иметь личные компьютеры для работы на лекциях и семинарах. Осенью 2001 г. ведущие провайдеры Интернета предложили специальные расценки для студентов, которые хотят установить дома свободный доступ в Интернет.

Уже ставший нормой виртуальный мир позволит упростить бюрократические процедуры образования и избавиться от всевозможных проволочек: безбумажного администрирования, записи на семинары, контрольных работ, экзаменов, видеоконференций, on-line-консультаций у преподавателей и профессоров.

Первопроходцами в области новой информации в Австрии являются Экономический университет Вены, Университет Кеплера в городе Линце. В Университете Вены более 30 % студентов записались на обучение и семинары через Интернет. (В Австрии приём в высшие учебные заведения проводится по окончании гимназии без вступительных экзаменов.)

Открытый университет Каталонии (Испания) увеличил количество своих web-студентов с 400 человек в 1995 г. до 20 тыс. в 2000 г. Университет в городе Хагене (Германия) в течение 26 лет держит первенство в виртуальном образовании и насчитывает в настоящее время более 60 тыс. студентов, из которых 10 тыс. учатся с помощью Интернета.

Хорошим примером дистанционного обучения является Открытый университет Милтона Кейнса в Великобритании, который насчитывает 200 тыс. web-студентов во всём мире; с 1992 г. открыт его филиал в Вене. Студенты Открытого университета общаются с преподавателями через Интернет, а также изучают учебный материал, представленный на CD ROM. Каждые шесть недель учащиеся встречаются в аудиториях для совместных занятий, а раз в год проходят совместную практику.

В Университете Милтона Кейнса разработан и введён учебный курс «Т171», призванный заменить традиционные формы преподавания. Обучение в его рамках происходит через Интернет; каждый студент поддерживает постоянный контакт со своим преподавателем по E-mail.

технологий. На её реализацию предполагается выделить в 2002—2006 гг. 56 млрд рублей.

Каждый второй компьютер подключат к Интернету, будет создана единая база данных, ежегодно планируется готовить 25 тыс. специалистов с высшим образованием и 60 тыс. — со средним. Все высшие учебные заведения намечается подключить к Сети к концу 2005 г., все школы — к 2010 г. В России один компьютер приходится на 500 школьников, в Европе — на 15. К концу 2005 г. предполагается увеличить в школе количество компьютеров до одного на каждых 80 школьников.

Кроме того, остро стоит проблема пропускной способности каналов. Её необходимо повысить до 256 кбит/с, иначе повсеместное подключение к Интернету будет лишено смысла.

«Электронная Россия» — это 3 млрд долларов на десять лет. «Единая образовательная информационная среда», параллельная с ФЦП программа, — это 2 млрд долларов на пять лет. В результате реализации программы, как показывают расчёты, до 2006 г. темпы роста отечественного производства в сфере информационно-коммуникационных технологий достигнут 15—25 % в год. Объёмы рынка информационных услуг и программного обеспечения увеличатся к 2010 г. в пять-шесть раз.

Компьютеризация федеральных ведомств также позволит получить некоторый эффект. До 65 % межведомственного оборота будет осуществляться безбумажным способом.

«E-AUSTRIA В ЕВРОПЕ»

14 апреля 2000 г. правительство Австрии опубликовало программу «E-Austria in E-Europe», реализация которой была расчитана на пять лет. Основой проекта являлись электронное образование и «Электронное правительство». Для того чтобы пользоваться услугами «Электронного правительства», нужно уметь не только писать и читать, но и обращаться с компьютером. Поэтому правительство Австрии уделяет столь пристальное внимание вопросам образования старшего поколения.

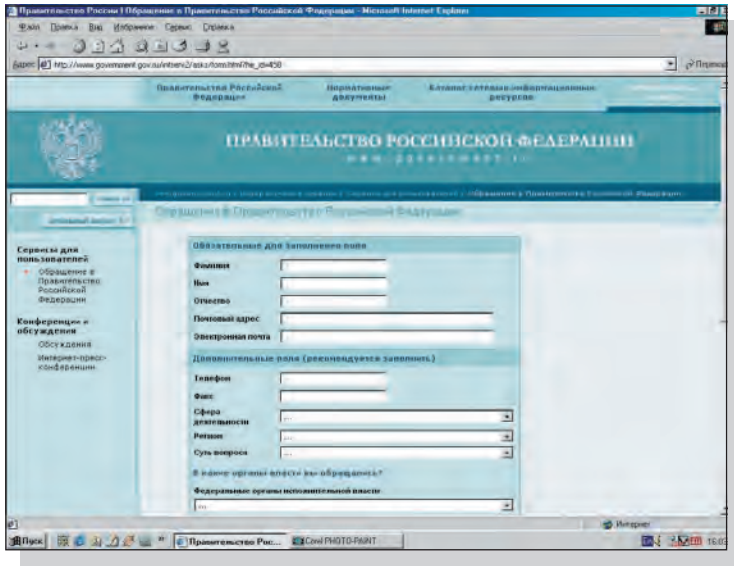
26 сентября 2000 г. правительство внесло проект «Виртуальный рынок Австрии». В Интернете предоставлен свободный доступ получения информации в таких областях, как политика, управление, право, экономика, техника, окружающая среда, образование.

В 2003 г. каждый австриец получил пластиковую карточку, которая заменяет собой полисы социальной и медицинской страховок (электронная карточка гражданина; *англ.* chipcard); разработка этих карт являлась одним из центральных пунктов проекта. Австрия здесь не одинока. В большинстве стран Европы также создаются электронные карточки граждан.

Полноценный ввод услуг «Электронного правительства» подразумевает достаточно высокие требования к работникам соответствующих государственных служб. Надо быть не только специалистом в собственной области, но и в совершенстве владеть компьютером и программным обеспечением, которое используется в определённом департаменте. Все работники государственных служб должны получить ECDL (*англ.* European Computer Driving Licence) — сертификат, подтверждающий умение работать на компьютере. Для этого проводят экзамены; не сдавшие их не будут допущены к занимаемой ими должности. Каждый должен обладать компьютерной грамотностью. Поэтому любой гражданин может при желании получить сертификат ECDL.

В конце 2001 г. был начат пилот-проект по представлению налоговых деклараций и налогообложению в режиме On-Line. Министерство финансов предоставило свободный доступ к системе всем австрийским гражданам. Большинство различных документов (паспорт, водительское удостоверение и т. д.) оформляются через Сеть уже с 2004 г. В 2005 г. должно быть обеспечено предоставление всех мыслимых услуг населению через Сеть.

Для пользования услугами One Stop Service необходимо обеспечить широкий свободный доступ к Интернету всем гражданам страны (дома и на работе). Для тех же, кто не имеет доступа к Интернету, будут созданы открытые компьютерные терминалы. Параллельно развиваются программы для инвалидов. Так, например, в настоящее время разрабатывается проект по ECDL для слепых.



* * *

Революционные изменения в «Электронном правительстве» не достигли своего возможного потенциала. Правительственные должностные лица обязаны внедрять современные технологии и сделать доступным пользование Интернетом для каждого гражданина. Однако повсеместное использование современных информационных технологий подразумевает высокий уровень компьютерной грамотности не только чиновников и студентов, но и всех слоёв общества независимо от социального положения, образования и возраста.

Смогут ли наши пожилые люди так же легко управляться с электронной почтой и пластиковыми карточками, как и молодые? Наверняка нет. Возникнет опасность дальнейшего расслоения общества. И, возможно, появятся люди, которые по объективным причинам не смогут овладеть всей современной премудростью Интернета.

ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ

Алан Кей.

Идея замены настоящих печатных изданий на электронные витала в воздухе давно. Её многократно описывали писатели-фантасты. В первый приезд в Москву в 90-х гг. XX века президента фирмы Apple Джона Скалли на его лекции в МГУ даже показывался футуристический рекламный ролик, где наряду с тогда ещё не существующей трёхмерной графикой демонстрировались электронные говорящие книжки, предназначенные для обучения неграмотных.

Но для внедрения идеи электронных библиотек надо знать, где взять и как получить электронные книги? Как стать абонентом электронной библиотеки? Сколько это стоит? Как читать? Проще ответить на последний вопрос — с монитора компьютера. Однако многим компьютер надоедает за рабочий день. Кроме того, такое чтение неудобно; во-первых, устают глаза, во-вторых, нельзя читать в метро или в автобусе.

Поэтому альтернативой является использование электронных записных книжек — PocketPC, PDA (Personal Digital Assistant — «персональный цифровой помощник»), карманного размера и весящих менее 200 г. На такой малютке без труда умещается несколько сотен книг. В них также можно хранить телефонную книгу,

В 1968 г. Алан Кай, разработал DynaBook — первый портативный интерактивный компьютер, выполняющий роль книги. В 1981 г. была создана рабочая группа Xanadu, которая впервые планировала заниматься так называемой подключённой литературой, т. е. продажей электронных книг.

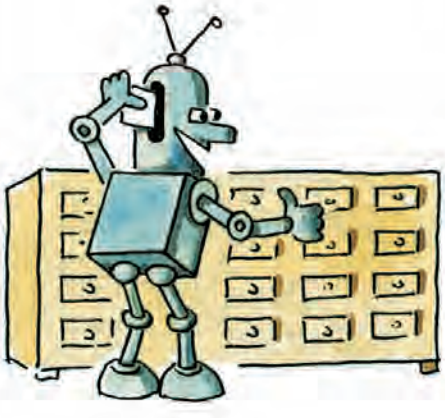
электронную почту, небольшие базы данных, электронные таблицы, игры и т. п. Естественно, что практически любой из этих «ручных» компьютеров можно подключить к Интернету.

В конце XX в. выпускались устройства, которые уже можно назвать электронными книгами. В США разработка специального оборудования для чтения электронных книг совмещена с платным подключением к электронному книжному магазину. SoftBook Reader (фирмы SoftBook Press) — нечто похожее на ноутбук (вес — 1,3 кг, размер экрана — 20×15 см) с обложкой из натуральной кожи.

Стоимость электронной книги с двухгодичной подпиской на услуги книжного магазина SoftBookstore (20 долларов в месяц) обходилась в несколько сотен долларов. Книжки хватает на хранение более 5 тыс. страниц текста, а батарей — на пять часов чтения. Каждый покупатель бесплатно получал 100 классических литературных произведений, за остальные же книги он платил (средняя цена составляла 10—25 долларов).

Rocket eBook (фирмы NuvoMedia; вес около 600 г, дисплей 11,5 x 7,5 см) стоит меньше, чем SoftBook Reader. Батарей хватает более чем на сутки, а объём памяти чуть более 3 тыс. страниц текста. В комплекте Rocket eBook покупатель в подарок получал «Алису в Стране Чудес» и Random House Webster's eDictionary объёмом 75 тыс. слов.

В XXI в. электронные книжки (устройства для чтения) совершенствуются почти ежедневно, и, что естественно , стоимость загрузки электронной книги из Интернет-магазина обходится дешевле, чем покупка обычной. Приобретённые по Сети книги, как правило, имеют защиту от копирования, которая порой столь неудобна, что затрудняет или вообще делает невозможным чтение книги.



В 1965 г. Теодор Нельсон, выпускник Гарварда, придумал термин *«гипертекст»*, основу хранилища мировой литературы в проекте Xanadu. Цель проекта состояла в организации доступа к электронным книжным хранилищам. При этом не ущемлялись авторские права и действовала система начисления авторского гонорара. Фактически это был прообраз электронного книжного магазина. В 1988 г. Xanadu получил финансирование у Джона Уокера, основателя Autodesk. Но в 1992 г. с появлением Всемирной паутины проект фактически был предан забвению.

Теодор Нельсон.

В 1971 г. профессор Иллинойсского университета (США) Майкл Харт создал «Проект Гутенберг».

У компьютера Xerox Sigma V (лаборатории Material Research) час машинного времени стоил более 70 долларов США. Однако такая дорогая машина отнюдь не всегда была загружена работой. И тогда Майкл Харт предложил способ хранения, передачи и поиска того, что находится в её «библиотеках», в то время, когда на ней не проводятся вычисления.

Основной тезис, на котором базируется «Проект Гутенберг», — всё, введённое в компьютер, можно скопировать неограниченное число раз. Автор назвал эту способность *репликационной технологией* (*англ.* replicator technology). Любой человек, где бы он ни находился, может скопировать себе электронную книгу (включая картинки, звук — всё, что сохранено в машине).

Поэтому формат электронного текста был выбран самым простым, (кодировка ASCII), чтобы на любой компьютерной платформе его удавалось прочитать.

Проект явно ждало большое будущее, но процесс шёл медленно, к началу 90-х гг. накопилось всего несколько сотен текстов. Развитие Интернета принесло «Проекту» не только читателей, но и армию наборщиков-добровольцев. Существенно подешевела стоимость магнитных носителей информации. Количество текстов стало стремительно расти.

Однако проект, базирующийся на бесплатной рабочей силе, предусматривал своей задачей популяризацию лишь классических произведений. Поэтому особенно актуально было точное соответствие текстов оригиналу.

«Мы ставим своей задачей выпускать электронные тексты, которые были бы на 99,9 % точны в глазах большинства читателей», — писал профессор Харт.

Тем не менее «Проект Гутенберг» позволил множеству государственных школ в разных странах мира пополнить их библиотеки несколькими тысячами классических произведений.

Одна из проблем, стоящая перед всеми компьютерными библиотеками (подобные проекты существуют и в других странах мира), — *копи-*

Иоганн Гутенберг — немецкий изобретатель книгопечатания *(слева)*.
Майкл Харт *(справа)*.

райт, представляющий подчас отнюдь не авторское право, а узурпацию права на копирование авторской работы. Это и волнует сегодня Майкла Харта.

Увеличение сроков, в течение которых авторское произведение не является публичным достоянием (первоначально 14 лет, потом — 28 и, наконец, совсем недавно — 70 лет в США), не может не наводить на грустные мысли. Ужесточение копирайта вообще ставит под сомнение существование электронных публичных библиотек. Если так пойдёт и дальше, вскоре вся информация может попасть под копирайт на срок дольше человеческой жизни.

> В прямом смысле понятие «авторское право» (*англ.* copyright) означает «право на создание копий». Упрощённо говоря, авторское право фактически сводится к праву копирования, которое принадлежит владельцу авторских прав, а не тому лицу, которое приобрело экземпляр охраняемого законом об авторских правах продукта.

ИНТЕРНЕТ И ОБЩЕСТВО

Значительные открытия и изобретения серьёзно влияют на общество и неотвратимо меняют его облик. Таковы, например, изобретение колеса, книгопечатания, железной дороги, электричества, телевидения, компьютера. Одни из них человечество оценило сразу, понимание того, что означают другие, пришло по прошествии века.

Интернет стал, пожалуй, одним из самых ярких изобретений XX столетия. Из технологического явления он быстро превратился в явление социальное. Создание глобальной компьютерной сети без преувеличения можно приравнять к мировой революции.

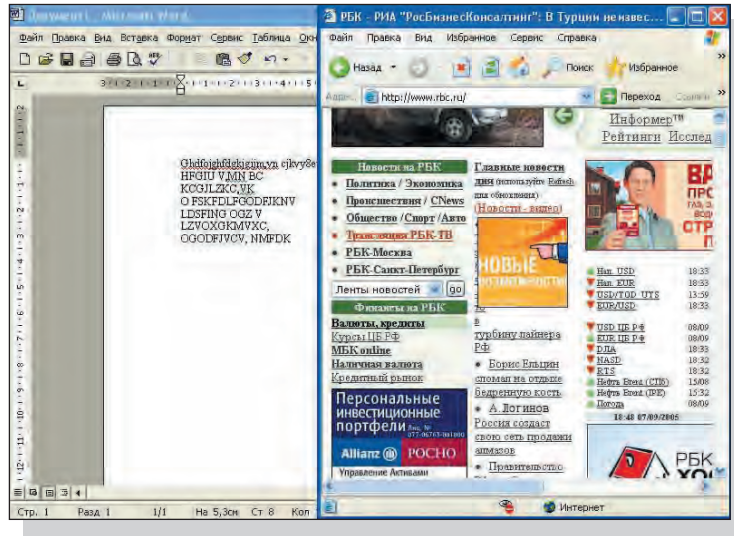
Интернет — планетарная структура, развивающаяся не по дням, а буквально по минутам. Это самое удобное и дешёвое средство распространения информации любого типа, будь то обыкновенный текст, изображения или же аудио- и видеоинформация.

Если ещё не так давно персональный компьютер использовался для вычислений и работы с документами, то сейчас, кроме того, он служит средством коммуникации между людьми, инструментом, с помощью которого можно получить доступ к любой информации со всего мира и даже вести бизнес. Вне всяких сомнений, рано или поздно Интернет придёт в каждый дом.

ИНТЕРНЕТ И ТЕЛЕВИДЕНИЕ

С каждым новым изобретением связывают определённые перемены. С появлением фотографии говорили, что она заменит живопись; кинематограф грозил вытеснить театр. В своё время полагали, что телевидение займёт место радио, но этого не произошло. Не такова ли ситуация и с Интернетом?

Результаты социологических исследований показали, что из тех опрошенных, кто имееет домашний доступ в Интернет, примерно одна треть предпочитает проводить время в Сети, а не перед экраном телевизора.



Скриншот демонстрирует открытие одного окна поверх другого.

УТРО 2010 ГОДА В ИНТЕРНЕТЕ

Что привнесли современные информационные технологии в нашу жизнь? Прежде всего, они объединили человечество в сообщество, имя которому — Интернет. И не беда, что не все ещё пользуются услугами глобальной компьютерной сети. Приход Интернета в каждый дом, в каждую семью не за горами. Это дело всего нескольких лет. Как же это будет? Давайте немного пофантазируем.

Обычный город России. Обычная семья. Папа, мама, двое детей. Отец работает ведущим инженером на заводе, мать — модельер. Дочь учится в университете. Сын ходит в школу и без ума от компьютеров.

Утро. Семья просыпается. Первым, как положено, встаёт папа. И сразу за газету. Но он не спускается к почтовому ящику и не выходит к киоску, а берёт плоский экран и читает утренние новости. Щёлк, папа касается пальцем экрана. Просматривает следующую страницу. Щёлк. Другая газета. Страницы мелькают одна за другой. Щёлк. Подбор новостей о его любимом заводе. Просмотрев электронную газету, папа просто сворачивает её трубочкой и кладёт в карман пижамы.

А вот и мама, ей к 9.00 на работу. Она торопится. У детей в комнате появляется её голографическое изображение. Она зовёт на кухню.

Все завтракают. И одновременно смотрят новости по WEB-TV — Интернет-телевизору. На экране размером во всю стену помимо основной передачи все важные новости сообщаются бегущей строкой, в углу — прогноз погоды и разнообразные рекламные вставки. Они уже не прерывают передачу, а тихо «висят» в верхнем углу телевизора. Новости не радуют. Растёт киберпреступность. Воруют программы и базы данных. С экрана диктор опять рассказывает про банду, которая совершает разбойные нападения на домашние компьютеры, уничтожает семейные базы данных. У соседа база стояла без охранной сигнализации, и ночью её «угнали». Общественность требует введения информационной смертной казни: отключения бандитов от Сети и лишения «логинов». Папа вспоминает, как было спокойно всего десять лет тому назад: ну грабили, ну угоняли автомобили, а теперь нападают в киберпространстве. Отнимают все деньги и стирают файлы. Автомобили и сами люди теперь никому не интересны. Прежних преступлений больше нет. Мама качает головой, дети обещают не ходить в Сети поздно поодиночке и никому не открывать «файрвол».

Мама заканчивает завтрак и уходит в комнату, где находится её рабочее место. Здесь она занимается проектированием новых моделей одежды. Мама включает компьютер. Что нового в мире моды? Читает почту с заказами. Отвечает на письма. Беседует с заказчиками. Разглядывает голографическую трёхмерную модель платья из новой французской коллекции. А вот выкройка её собственной модели на экране. Мама даёт команду машине: «Собираем». Вот так будет выглядеть модель в готовом виде. Чуть подправим форму рукава, изменим выкройку. Мама делает пассы руками, и модель на глазах меняется. Спереди, сбоку. Изменим материал, цвет. Всё готово. Можно шить. Это то, что она хотела. А вот и заказчик появляется на экране. Торопит. Мама его успокаивает.

Заказчик из Индии смотрит на трёхмерную модель платья у себя в кабинете. Он очень доволен. Каждый говорит на своём родном языке, автоматический перевод вполне сносный. Договорились об оплате. Вот и деньги поступили на мамин счёт, прямо по Сети. Выкройку она отправила в Индию. Там и сошьют. Ура, заказ выполнен. Мама свободна.

Сын с утра уже в Интернете. Делает задание. Ищет кино, по которому надо писать сочинение (за-

метьте: не книгу, а кинофильм). Нашёл. Бежит к папе. Говорит, что это недорого. Это задали в школе. Папа соглашается. Оплачивает фильм — прямо через Интернет. Сын радостно сгружает его по Сети. И садится смотреть.

Его 18-летняя сестра рассылает со своего компьютера звуковые письма через Интернет. Это удобно, когда посылаешь письма сразу трём поклонникам. И свои новые фотографии добавила к письмам... Изучать предмет римское право так скучно. Ни тебе картинок интересных, ни фильмов. Другое дело — разглядывание новых каталогов косметики. Отмечает, что её интересует, и заказывает через Сеть. Скоро пришлют по почте. Попутно через анонимный чат общается с такими же юнцами, только изображает из себя 60-летнюю матрону, утешающую несчастных в любви парней. Это её чрезвычайно забавляет. Но она даже не предполагает, что из пяти человек, участвующих в дискуссии о любви, только одному 18 лет. Двоим уже глубоко за 50, а ещё двоим не стукнуло и 14. Смешная вещь — чат. Здесь каждый может предстать тем, кем ему сегодня больше всего хочется быть.

Папа за компьютером. Обдумывает чертежи. Он в отпуске, но всё равно не может не работать. Исправил чертёж. Запустил расчёты модели. Вроде всё хорошо. Вдруг экстренный звонок с завода по Интернет-телефону. Звонит директор. Срочно требуется консультация. Папа присоединяется к телеконференции руководства завода и обсуждает со всеми участниками важную проблему. Прошло полчаса. Проблема разрешилась.

Папа идёт к сыну и застаёт того за игрой. Оказывается, к кинофильму прилагалась бесплатная игра. Это и было истинной целью ребёнка. Папа недоволен. Сын наказан. Ему грозит отправка на природу с отключением от Сети. На неделю.

Входит довольная мама. Она сообщает, что выполнила заказ и хочет с мужем и детьми съездить на неделю к морю. Семья у компьютера. Срочный заказ билетов, номеров в гостинице, такси. Семья начинает быстро собираться. Приезжает такси. Все на улице. Садятся в машину и сразу включают свои портативные компьютеры. Они снова в Сети.

Солнце в зените. День продолжается. День 2010 года в Интернете.

Стивен Кинг.

Фредерик Форсайт.

Информационное окно обновляется ежеминутно.

Причём среди молодёжи таких почти половина. Находятся в Интернете и одновременно смотрят телевизор 10 % опрошенных.

Можно ли говорить о том, что Интернет окончательно заменит телевидение? Пока ещё до этого далеко. Да и в дальнейшем это вряд ли произойдёт, так как надо учитывать, что получение информации из Сети связано с тратами времени, с определёнными усилиями и вниманием. В то же время одним из приятных достоинств телевидения является то, что человек, придя домой и опустившись в кресло, расслабляется и смотрит, что ему показывают по телевизору. Хотя уже сейчас некоторые телевизионные каналы предлагают свои программы в Сети...

ИНТЕРНЕТ И КНИГИ

Есть предположение, что Интернет вытеснит и заменит книги. Сейчас этому препятствует ряд факторов. Во-первых, отсутствие комфорта при чтении книг с монитора компьютера. И хотя портативные устройства для чтения электронных текстов уже существуют (нет никаких сомнений в том, что они должны быть именно портативными), разрешение их экранов явно недостаточно. Специализированные же мониторы с высоким разрешением, используемые в издательских системах, громоздки и довольно дороги. Во-вторых, не до конца разработаны авторские права на электронные публикации. В любом случае только автор может определять условия распространения собственного произведения. Поэтому должна быть надёжная защита от несанкционированного копирования.

В конце XX в. известные писатели американец Стивен Кинг и англичанин Фредерик Форсайт публиковали свои очередные бестселлеры исключительно в электронном виде, ведь чтобы загрузить книги через Интернет, нужно было заплатить чисто символическую сумму. Возможно, со временем на бумаге будут выпускать только какие-то особенные издания, подарочные или коллекционные.

Преимущества электронных книг очевидны: быстрота доставки (не надо идти в магазин или библиотеку), малый объём (многотомное собрание сочинений умещается на миниатюрном диске), удобный поиск по тексту. Немаловажно, что снижение доли бумажных изданий позволит сохранить леса от вырубки.

ИНТЕРНЕТ И СМИ

В СМИ технологии, предоставляемые Интернетом, получили максимально эффективное развитие и применение. Это объясняется несколькими причинами, одна из них — оперативность. Помимо автора над подготовкой телепередачи работают операторы, звукорежиссёры, осветители, редакторы, монтажёры и многие другие. Над газетной статьёй — автор, редактор, корректор, художник, её надо напечатать и доставить читателю. При этом проходит немало времени. В Интернете же информация размещается в считанные минуты и доставляется читателю прямо на рабочее место. В отличие от газет и журналов материалы в Интернете всегда актуальны и могут обновляться в реальном времени (минимальный срок обновления информации в газете — один день). Стоимость раз-

мещения электронной версии газеты в Интернете меньше стоимости её печатного аналога (происходит значительная экономия на бумаге, полиграфических и транспортных расходах).

Интернет в перспективе значительно потеснит традиционные средства массовой информации благодаря гибкости, оперативности и интерактивности (возможности постоянной и быстрой обратной связи с читателями).

ИНТЕРНЕТ И ЗАКОН

С развитием компьютерных и телекоммуникационных технологий, и Интернета в частности, появилась так называемая киберпреступность: несанкционированный доступ к компьютерной информации; умышленное нарушение работы web-сайтов; финансовые махинации в Интернете; нарушения авторских прав; распространение вирусов; пропаганда человеконенавистнических идеологий.

В современном уголовном законодательстве практически любого государства совершение преступления с использованием технических средств (а Сеть и компьютер — технические средства) является отягчающим обстоятельством, увеличивающим наказание.

ИНТЕРНЕТ-ЗАВИСИМОСТЬ

Интернет — это уникальное средство для расширения контактов, но оно не может заменить полноценное человеческое общение. Люди, сильно увлекающиеся Интернетом, иногда теряют грань реальности. Это приводит к Интернет-зависимости, или аддикции, которая представляет серьёзную угрозу для здоровья (физического и психического). Человек становится беспомощным перед своим пристрастием: волевые усилия ослабевают и не дают возможности противостоять зависимости.

Значительный ущерб наносится межличностным отношениям. Поведение зависимых выражается в стремлении уйти от реальности путём постоянной фиксации внимания на Интернете, это сопровождается развитием интенсивных эмоций. Данный процесс настолько захватывает человека, что начинает управлять его жизнью. Интернет заменяет дружбу, любовь; поглощает время, силы, энергию и эмоции.

ИНТЕРНЕТ И ПОЛИТИЧЕСКАЯ ЖИЗНЬ

Всего за несколько лет Интернет из узкоспециальный компьютерной среды развился в мощное информационное пространство. Сеть выступает как средство общения, обмена информацией, служит своеобразной трибуной для политической пропаганды. Первым

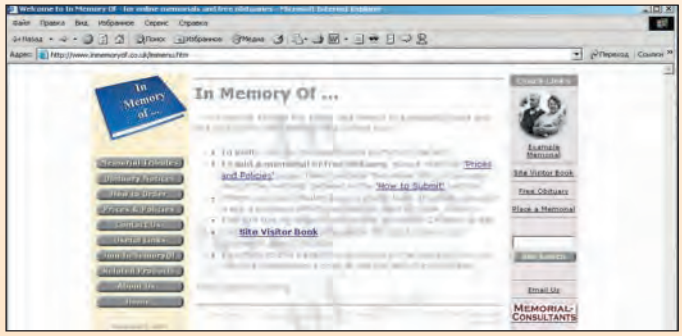
ПРОГРАММИСТ СМЕНИЛ ФАМИЛИЮ НА НАЗВАНИЕ СВОЕГО САЙТА

Израильского программиста Томера Крисси не устраивала собственная фамилия. Фамилии, по его мнению, являются архаизмом и имеют только историческую ценность. Томеру хотелось, чтобы в том, как его зовут, был какой-то смысл. Приложив некоторые усилия, он сумел получить новый паспорт на имя Томер.com.

Теперь, утверждает программист, ему достаточно представиться, чтобы собеседник понял, что у Томера есть собственный сайт, где можно найти о нём всю необходимую информацию. Кроме того, такая фамилия подходит ему как нельзя лучше: некоторое время назад он страдал от Интернет-зависимости и проводил в Сети около 12 часов в сутки.

КАК ИНТЕРНЕТ ЗАТРАГИВАЕТ САМЫЕ РАЗНЫЕ СТОРОНЫ ЖИЗНИ. ВИРТУАЛЬНОЕ КЛАДБИЩЕ

Свои услуги родственникам усопших предлагает первое виртуальное кладбище в Интернете — сайт www.pazeterna.com.ar, его открыл аргентинец Хорхе Луис Кальвильони. Заказчики могут увековечить память о близких, выбрать памятник, венки и цветы, разместить на виртуальных могилах фотографии, видео- или аудиозаписи, некрологи. По желанию родных доступ посторонних к могиле свободен либо закрыт. Однако эта услуга платная, стоимость её составляет 70 долларов в год.



сайтом российской политической партии в Интернете стал DS.ru, открытый в 1996 г. службой информации партии «Демократический союз».

Осенью 1999 г. парламентские выборы вызвали значительную активизацию политической деятельности в Интернете. Практически все партии и движения постарались открыть собственные страницы в Сети. Тогда же впервые появились специальные информационные ресурсы, освещающие ход предвыборных кампаний и результаты выборов. На некоторых из них размещались сенсационные материалы о той или иной политической силе. Политические скандалы, потрясавшие страну с экранов телевизоров и страниц печатных изданий, продолжались и в Интернете.

Наличие собственного сайта стало непременным атрибутом политического имиджа. Собственные сайты имеют, в частности, Правительство РФ — www.government.gov.ru., Государственная дума — www.duma.ru., Президент РФ — president.kremlin.ru.

СВОБОДА СЛОВА И ЦЕНЗУРА В ИНТЕРНЕТЕ

ИНТЕРНЕТ И СВОБОДА СЛОВА

Свобода слова — основа демократического общества, это основное право человека и фундамент всех свобод. С появлением Интернета методы получения и распространения информации радикально изменились. В отличие от прежних средств передачи информации Интернет позволяет каждому, у кого есть компьютер и доступ к Сети, мгновенно устанавливать связь с людьми по всему миру. Интернет преодолевает государственные границы и устраняет барьеры на пути потока информации. Эта технология не только глобальна — она отдаёт контроль над созданием и распространением информации в руки отдельных людей.

Каждый человек имеет право на свободу убеждений и на свободное выражение их; это право включает свободу беспрепятственно придерживаться своих убеждений и свободу искать, получать и распространять информацию и идеи любыми средствами и независимо от государственных границ.

Всеобщая декларация прав человека, статья 19

«Тот, кто готов поступиться вечной свободой во имя сиюминутной безопасности, не достоин ни свободы, ни безопасности».

Бенджамин Франклин

Под свободой слова в Интернете следует понимать, что никакие государственные органы не должны регулировать содержание информационных ресурсов.

Среди основных прав человека есть право на доступ ко всем проявлениям знаний, творческой мысли и интеллектуальной активности, а также право на публичное высказывание собственных взглядов. Любой человек имеет право разместить на личном компьютере произвольную информацию и предоставить право пользования ею любому желающему, если распространение такой информации не запрещено международным законодательством или законодательством его страны (законы охраняют государственную тайну, запрещают разжигание социальной, расовой, национальной и религиозной розни). Причём для этого не требуются разрешительные либо уведомительные документы или процедуры в отличие от радио- или телевизионного вещания. Любое посягательство на эти права со стороны других людей, организаций или государства — вмешательство в личную жизнь и свободу творчества граждан.

Всеобщий и беспрепятственный доступ к информации является благом, поскольку владение и умелое пользование информацией развивают экономическую, социальную и культурную интеграцию, способствуют установлению демократических ценностей.

ЦЕНЗУРА И ИНТЕРНЕТ

Цензура существует с давних времён, в Западной Европе она появилась в XV в. С начала формирования государства идёт борьба между властью и обществом, которое стремится быть независимым, высказывать свои оценки, критиковать политику властей.

В разные исторические периоды цензуру использовали сначала Церковь, затем монархи, государственная власть. В демократических государствах, где права человека самоценны, попытки их ограничения неизменно наталкиваются на протест граждан. Появление Интернета с его анархичной структурой усложнило проблемы цензуры. Благодаря децентрализованному характеру Интернета многие

Бенджамин Франклин.

1. Каждому гарантируется свобода мысли и слова.

2. Не допускается пропаганда или агитация, возбуждающая социальную, расовую, национальную или религиозную ненависть и вражду. Запрещается пропаганда социального, расового, национального, религиозного или языкового превосходства.

3. Никто не может быть принуждён к выражению своих мнений и убеждений или отказу от них.

4. Каждый имеет право свободно искать, получать, передавать, производить и распространять информацию любым законным способом. Перечень сведений, составляющих государственную тайну, определяется федеральным законом.

5. Гарантируется свобода массовой информации. Цензура запрещается.

Конституция Российской Федерации, статья 29

средства контроля со стороны государства оказываются неэффективными и даже бесполезными, так как в распоряжении пользователей имеются многочисленные способы их обойти.

Причинами возникновения цензуры в Интернете стали угроза национальной безопасности и безопасности крупных корпораций в связи с действиями хакеров; распространение информации непристойного характера; посягательство на интеллектуальную собственность; использование Сети экстремистскими политическими группами для публичного распространения своей идеологии, а также доступность или распространение опасной информации о наркотиках, способах самоубийства и т. п. С точки зрения общественной морали (а в большинстве случаев и законодателя) распространение подобных материалов в Интернете расценивается как зло.

Когда издателем способен стать каждый, право распространять информацию приобретает новый смысл. Уникальные возможности Интернета соответствуют понятию «независимо от государственных границ», фигурирующему в важнейших международных документах о правах человека.

Меры властей по контролю за Интернетом противоречат нормам международного права. Цензура, осуществляемая в одной стране, в другой может привести к прямому нарушению права личности распространять информацию.

Реакцией на угрозу появления цензуры в Интернете стало немедленное создание сетевых объединений. «Международное движение за свободу в Интернете» (Global Internet Liberty Campaign) — это самая крупная организация, которая объединяет 45 организаций, пропагандирующих свободу слова и доступа к информации по всему миру. Другой пример — международное сетевое объединение «Репортёры без границ» (Reporteurs sans frontiers), выступающее против любого ограничения доступа к ресурсам Интернета.

По мнению членов подобных организаций, всё, что можно осудить с моральной точки зрения, необязательно должно осуждаться и законом. Они стремятся донести до общественности такую мысль: бороться с преступностью следует, не вводя цензуру в Интернете, а искореняя непосредственные источники преступлений. Интернет — одно из важнейших средств коммуникации в современном мире и в принципе не должен подвергаться цензуре.

Вопрос свободы доступа к Интернету довольно сложен с этической точки зрения. Каждый человек имеет право оградить себя и своих детей

УПРАВЛЕНИЕ ИНТЕРНЕТОМ

Первые попытки регулировать содержание Интернета совпали по времени с его стремительным внедрением в жизнь общества. Уже в 1995 г. был обнародован Акт о пристойности коммуникаций США. Этот законодательный акт приравнивал неприличные материалы, посылаемые с помощью Интернета, к противозаконным и определял меру наказания от штрафа (до 10 тыс. долларов) до лишения свободы (на срок до двух лет). Особо запрещалось осознанное использование интерактивной компьютерной службы для пересылки или показа любых оскорбительных по своей сути материалов лицам моложе 18 лет. Акт был принят 1 февраля 1996 г. и подписан президентом Клинтоном 8 февраля. В тот же день группа организаций, возглавляемая Американским союзом в защиту гражданских прав, получившая название «Коалиция граждан за права в Интернете», обратилась в Верховный суд США с иском о неконституционности данного акта. В Интернете была развёрнута широкомасштабная кампания по защите свободы слова против цензуры. Суд признал акт неконституционным, как противоречащий первой поправке к Конституции США, которая запрещает издавать законы, ограничивающие свободу слова.

Цензура (от *лат.* censeo — «оцениваю») — система государственного надзора за печатью и средствами массовой информации.

от нежелательной, на его взгляд, информации, и он может жёстко «фильтровать» на личном компьютере всё, что ему заблагорассудится. Но только на собственном. Если кому-то какой-либо сайт и не нравится, пусть не смотрит его сам, не разрешает смотреть детям, но не требует отмены для всех.

В демократических странах законодатель не может ввести цензуру (в США это запрещается первой поправкой к конституции, в России — статьёй 29.5 конституции). Однако провайдеры, будучи частными коммерческими структурами, имеют право распоряжаться принадлежащим им дисковым пространством и каналами связи в соответствии со своими предпочтениями, т. е. осуществлять контроль над материалами, передаваемыми по их информационным каналам. Все недовольные клиенты могут подыскать себе других провайдеров. Но государство не должно каким-то образом фильтровать частную информацию, какими бы общественными интересами оно ни пыталось это оправдать, потому что свобода слова — наивысшая ценность цивилизованного общества.

ЗАПРЕТ НА ИНТЕРНЕТ

Международная организация «Репортёры без границ» распространила список стран — врагов Интернета, в которых значительно ограничивают или вовсе запрещают своим гражданам свободный доступ в Интернет под предлогом защиты от распространения разрушительных идей и ради гарантии безопасности национального единства. Среди них — Азербайджан, Белоруссия, Вьетнам, Ирак, Иран, Казахстан, Киргизия, Китай, Куба, Ливия, Саудовская Аравия, Северная Корея, Сирия, Судан, Таджикистан, Туркмения, Тунис и Узбекистан.

Руководитель Северной Кореи Ким Чен Ир, желая, чтобы его государство оставалось закрытым для всего остального мира, принял решение о запрете Интернета. Единственный сайт, на котором режим ведёт официальную пропаганду, зарегистрирован в Японии.

Кубинский лидер Фидель Кастро считает Интернет инструментом, который капитализм использует для манипулирования общественным мнением. На Кубе доступ в Интернет разрешён только иностранным туристам или функционерам компартии.

Как Древний Китай был отгорожен от мира Великой стеной, так компьютерные сети современного социалистического Китая отделены от Всемирной паутины, а подключения и трафик контролируются государством. Созданы специальные полицейские подразделения, специализирующиеся на преследовании авторов антикоммунистических и антиправительственных статей, опубликованных в Интернете. Инакомыслие в Сети карается длительными сроками заключения.

В Белоруссии требуется регистрация средств доступа к Интернету (модемов) в органах внутренних дел.

ЗАЩИТА ИНФОРМАЦИИ

АВТОРСКОЕ ПРАВО В ЦИФРОВОЙ ВЕК

В России действует Закон РФ от 23 сентября 1992 г. «О правовой охране программ для электронных вычислительных машин и баз данных».

В России авторское право регулируется Законом РФ от 9 июля 1993 г. «Об авторском праве и смежных правах».

АВТОРСКОЕ ПРАВО И КОМПЬЮТЕРИЗАЦИЯ

Компьютеры так стремительно ворвались в жизнь человека, а новые информационные технологии столь быстро развивались, что сначала законодатели оказались к этому неподготовленными. Некоторое время оставалось неясным, какую правовую охрану предоставить компьютерным программам. Однако было очевидно, что защита необходима, чтобы обеспечить должное развитие компьютерной индустрии.

В настоящее время компьютерные программы (редакторы, компиляторы, базы данных) приобрели значение товарной продукции. Например, база данных — это объективная форма представления и организации совокупности данных (статьи, расчёты и пр.), систематизированных таким образом, чтобы они могли быть найдены и обработаны с помощью ЭВМ. Независимо от формы своего выражения программы и базы данных защищаются так же, как и произведения литературы. Кроме того, в компьютерных программах используются другие объекты авторского права: произведения литературы, изобразительного искусства, кинематографии, музыкальные произведения.

В отличие от материальной собственности объекты интеллектуальной собственности, в том числе и компьютерные программы, покупатели могут использовать лишь на особых условиях, устанавливаемых автором. При этом в собственность покупателя переходит только носитель информации, например компакт-диск, на котором программа записана. Однако покупатель вправе без согласия правообладателя и без выплаты ему дополнительного вознаграждения изготовлять копии программы при условии, что они предназначены исключительно для архивных целей.

АВТОРСКОЕ ПРАВО И ИНТЕРНЕТ

Развитие новых информационных технологий положило начало процессу перевода всей информации в цифровой вид, а появление и повсеместное развитие Интернета, облегчившее копирование и распространение

Закон определяет программу как объективную форму представления совокупности данных и команд, предназначенных для функционирования электронно-вычислительных машин (ЭВМ) и других компьютерных устройств с целью получения определённого результата, включая подготовительные материалы, полученные в ходе её разработки, и порождаемые ею аудиовизуальные отображения.

ЧТО ТАКОЕ АВТОРСКОЕ ПРАВО

Авторское право — это часть гражданского права, которая регулирует порядок использования произведений литературы, науки и искусства. Объектами авторского права признаются также программы для компьютеров и электронные базы данных.

Охраняемое произведение должно быть результатом творческого труда, продуктом интеллектуальной деятельности человека, оригинальным, отмеченным индивидуальностью автора. Результаты интеллектуальной деятельности называют *интеллектуальной собственностью*.

Субъектами авторского права, т. е. лицами, обладающими исключительным правом на произведение, считаются прежде всего авторы произведений. Правообладателями также могут быть различные компании, приобретающие право на коммерческое использование произведения, или наследники автора.

Авторские права делятся на личные неимущественные и имущественные права. Из неимущественных прав автору принадлежат

- право авторства — право признаваться автором произведения;
- право на имя — право использовать или разрешать использовать произведение под именем автора или его псевдонимом;
- право на обнародование — право обнародовать или разрешать обнародовать произведение в любой форме;
- право на защиту репутации автора, право на защиту произведения, включая его название, от всякого искажения или иного посягательства, способного нанести ущерб чести и достоинству автора.

Личные неимущественные права принадлежат автору независимо от его имущественных прав и сохраняются за ним, например, даже в случае передачи исключительных прав на публикацию произведения.

К имущественным правам автора относятся

- право на воспроизведение;
- право на распространение;
- право на переработку;
- право на импорт;
- право на перевод;
- право па публичный показ;
- право на публичное исполнение;
- право на передачу в эфир;
- право на сообщение для всеобщего сведения по кабелю (т. е. передавать произведение, включая показ, исполнение или передачу в эфир, для всеобщего сведения по кабелю, проводам или с помощью других аналогичных средств);
- право на доведение до всеобщего сведения (т. е. сообщать произведение таким образом, при котором любое лицо может иметь доступ к нему в интерактивном режиме из любого места и в любое время по своему выбору).

Авторское право действует в течение всей жизни автора и 70 лет после его смерти. По истечении срока охраны произведение переходит в общественное достояние, его можно использовать, не спрашивая разрешения. Право авторства, право на имя и право на защиту репутации автора охраняются бессрочно.

ИЗ ИСТОРИИ АВТОРСКОГО ПРАВА

История авторского права началась с изобретения книгопечатания, благодаря которому стало возможным механическое размножение произведения вместо переписывания его от руки. Сначала охрана авторского права предоставлялась в форме привилегий, выдаваемых королями. В 1557 г. английская королева Мария Тюдор издала указ о том, что одна из британских типографских гильдий наделяется монополией на книгопечатание и книготорговлю в стране. Согласно данному указу, книготорговец, принадлежащий к этой гильдии, купивший книгу у автора и зарегистрировавший её в специальном реестре, становился владельцем права на печатание.

11 января 1709 г. в британской палате общин был заслушан проект закона «О поощрении образования путём закрепления за автором или приобретателями копий печатных книг прав на последние на время, устанавливаемое отныне». 10 апреля 1710 г. проект стал первым законодательным актом об авторском праве. В историю он вошёл под названием «Статут королевы Анны». Закон предоставил авторам правовую охрану на их произведения на 14 лет; при жизни автора она могла быть продлена ещё на 14 лет.

Королева Мария I Тюдор.

Королева Анна Стюарт.

Первым многосторонним международным соглашением, которое касалось авторских прав, была принятая в 1886 г. постоянная Бернская конвенция об охране литературных и художественных произведений.

В 1952 г. была принята вторая основная международная конвенция — «Всемирная конвенция об авторском праве».

Особенностью авторского права в России являлась тесная связь с цензурным законодательством. Первым авторским законом в России стал Цензурный устав (1828 г.). Он содержал специальную главу «О сочинителях и издателях книг», предоставлявшую автору исключительные права на собственные произведения на протяжении его жизни и наследникам — на 25 лет после смерти автора. В 1830 г. вступил в действие ещё один закон — «Положение о правах сочинителей, переводчиков и издателей».

ОТВЕТСТВЕННОСТЬ ЗА НАРУШЕНИЕ АВТОРСКОГО ПРАВА

Российское законодательство отвечает всем принципам международных соглашений об авторском праве. Оно включает разнообразные формы его защиты. За нарушение авторских прав наступает гражданская, уголовная и административная ответственность.

Экземпляры произведения или программы, изготовление или распространение которых влечёт за собой нарушение авторских прав, называются по закону контрафактными (от *фр*. contrefaçon — «подделка»). В обиходе же часто употребляется термин «пиратская копия» (от *англ*. piracy — «нарушение авторского права»). Обладатели авторских прав могут обратиться за защитой в суд и вправе требовать от нарушителя признания прав автора; восстановления положения, существовавшего до нарушения права, и прекращения действий, нарушающих право или создающих угрозу его нарушения; возмещения убытков, включая упущенную выгоду; взыскания дохода, полученного нарушителем вследствие нарушения авторских и смежных прав; выплаты компенсации в сумме от 10 до 50 тыс. минимальных размеров оплаты труда (МРОТ). Контрафактные экземпляры произведений подлежат конфискации и уничтожению.

В соответствии со статьёй 7.12 Кодекса Российской Федерации об административных правонарушениях незаконное использование контрафактных экземпляров влечёт наложение на граждан административного штрафа в размере от 15 до 20 МРОТ.

Согласно статье 146 Уголовного кодекса РФ, незаконное использование объектов авторского права, а равно присвоение авторства, если эти деяния причинили крупный ущерб, наказываются штрафом в размере от 200 до 400 МРОТ, либо в размере заработной платы или иного дохода осуждённого за период от двух до четырёх месяцев, либо обязательными работами на срок от 180 до 240 часов, либо лишением свободы на срок до двух лет.

Диски с контрафактной продукцией.

информации, только заострило проблемы охраны авторских прав. Современное законодательство отражает эти изменения и охраняет объекты авторского права в любой форме. Так, произведение, размещённое в Интернете, отвечает всем признакам объекта авторского права: оно, во-первых, является результатом творческой деятельности независимо от назначения и достоинства произведения, а также от способа его выражения; во-вторых, закреплено в объективной форме — в письменной форме (текстовые файлы и web-страницы) и в форме изображений. Вместе с тем широкая доступность Интернета породила ряд распространённых заблуждений относительно авторского права на материалы, размещённые в системе World Wide Web.

• Если не проставлен знак уведомления об авторском праве (буква «с» в окружности — ©), то произведение не охраняется. Это не так: авторское право на произведение или программу возникает в силу их создания. Проставлять знак © не обязательно.

• Авторское право не нарушается, если нарушитель не извлекает прибыли. Это опять неверно: использование произведения без согласия автора — нарушение в любом случае.

• Произведения, размещённые в Интернете, являются общественным достоянием. Однако произведение переходит в общественное достояние только по истечении срока охраны или если автор явно выразил такое желание.

• Размещение чужого произведения в Интернете считается личным использованием. Это не так: доведение произведения до всеобщего сведения относится к исключительным правам автора.

Лёгкость копирования объектов авторского права вынуждает законо-

Каждому гарантируется свобода литературного, художественного, научного, технического и других видов творчества, преподавания. Интеллектуальная собственность охраняется законом.

Конституция Российской Федерации, статья 44-1

THE DIGITAL MILLENNIUM COPYRIGHT ACT OF 1998
U.S. Copyright Office Summary

December 1998

INTRODUCTION

The Digital Millennium Copyright Act (DMCA)[1] was signed into law by President Clinton on October 28, 1998. The legislation implements two 1996 World Intellectual Property Organization (WIPO) treaties: the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty. The DMCA also addresses a number of other significant copyright-related issues.

дателей усиливать их защиту. В 1998 г. конгресс США принял закон «О защите авторских прав в цифровом тысячелетии» (*англ.* Digital Millennium Copyright Act — DMCA). DMCA предусматривает гражданское и уголовное преследование субъектов, нарушающих авторские права, в том числе и в Интернете. Согласно этому закону, люди, предоставляющие информацию о том, как нарушить авторское право, а также проектирующие и создающие инструменты для взлома защиты авторского права, могут быть задержаны и предстать перед судом точно так же, как и те, кто незаконно использует защищённые авторским правом материалы. Подобные DMCA законы, предусматривающие ответственность за обход технических средств защиты от копирования, приняты как в странах Евросоюза, так и в Российской Федерации.

COPYRIGHT И «COPYLEFT»

Линус Торнвальд.

Сегодня каждый человек, имеющий дело с компьютерами, слышал про свободно распространяемую операционную систему Linux. Многие знают и о её создателе (точнее, основателе) — Линусе Торнвальде. Но не всем известно о знаменитом предшественнике Линуса Торнвальда — Ричарде Столмене, заложившем много лет назад философскую, юридическую и солидную техническую базу, на которой и расцвёл Linux.

В 70-х гг. XX в., когда Ричард Столмен работал преподавателем и программистом в знаменитом Массачусетсском технологическом институте, он был в группе специалистов, которая использовала исключительно свободное программное обеспечение. Сам Столмен активно проповедовал следующее: интеллектуальные достижения, включая исходные тексты компьютерных программ, должны быть доступны всему человечеству для использования, копирования, исправления ошибок, развития, полной переделки и т. д.

В те годы, как и сейчас, большинство программ являлось собственностью отдельных компаний. Для законного использования таких программ нужно было купить лицензию. Обычная индивидуальная лицензия сильно ограничивает права покупателя:

- позволяет устанавливать данную программу только на одном компьютере;
- разрешает сделать лишь одну резервную копию программы;
- запрещает другому человеку работать с программой даже на том компьютере, где программа законно установлена;
- запрещает исправлять ошибки в программе или даже изучать коды программы с целью найти источник ошибки.

«Проект GNU» (рекурсивная аббревиатура GNU's Not UNIX. — *Прим. ред.*) зарождён в 1983 г. как способ возвращения духа сотрудничества, который преобладал в сообществе пользователей компьютеров в ранние годы, чтобы сделать сотрудничество снова возможным, удалив препятствия, создаваемые владельцами закрытого программного обеспечения», — пишет в своём манифесте Ричард Столмен. Под покровительством Массачусетсского технологического института он организовал некоммерческое объединение Free Software Foundation (Общество поддержки свободно распространяемого программного обеспечения).

В самом прямом смысле понятие «авторское право», или «copyright», означает «право создания копий».

Для реализации его идеи в первую очередь требовалась операционная система, под управлением которой это программное обеспечение начало бы функционировать. У Столмена не было никаких сомнений: ОС должна быть совместима с UNIX. Кроме того, операционная система и все инструменты (компьютерные программы) должны быть свободно распространяемыми. Первой программой проекта GNU оказался текстовый редактор GNU Emacs, что вполне естественно, ведь, чтобы написать программу, нужен редактор. Программа быстро стала популярной, и у проекта появились приверженцы.

К 1990 г. коллекция программ GNU включала все основные компоненты операционной системы, кроме ядра. Основным ядром, под управлением которого работает коллекция программ GNU, стало ядро Linux. Работы над ядром HURD ведутся до сих пор.

Большинство программистов сегодня согласны с тем, что создание Linux без использования результатов Free Software Foundation было бы невыполнимо. Какой из результатов основополагающий, мнения разделятся: часть скажет: «Разумеется, главное — это компилятор», а другая часть возразит: «Нет, это разработанная Столменом лицензия». Лицензия называется GNU Public License, сокращённо GPL, и суть её весьма проста. Предположим, что автор *X* создал программу *Y* и хочет подарить её всему человечеству, сделав свободно распространяемой. Если *X* говорит: «Все могут использовать мою программу и делать с ней что угодно», то возможно следующее: злоумышленник незначительно меняет программу *Y* и превращает её в программу *Z*, объявляя своей собственностью. GNU-лицензия запрещает такую ситуацию. Согласно этой лицензии, переделка GNU-программ разрешена, только если в результате переделки будет снова получена GNU-программа.

«Цель проекта GNU — дать всем пользователям свободу распространять и изменять программное обеспечение GNU. Если бы посредник мог убрать свободу, у нас, возможно, было бы много пользователей, но у них не было бы свободы. Поэтому, вместо того чтобы делать программное обеспечение GNU общественным достоянием, мы выпускаем его под “copyleft”... Copyleft — это такое правило, что при распространении программы вы не можете добавлять ограничения, отбирающие у людей основные свободы», — пишет Столмен.

HURD — рекурсивная аббревиатура: HIRD of UNIX-Replacing Daemons, где HIRD — аббревиатура от HURD of Interfaces Representing Depth. Перевести эту рекурсивную аббревиатуру (в соответствии с традициями GNU) на русский язык практически невозможно. Знакомые с английским могут попытаться сделать это сами. (Подсказка: словом daemons в мире UNIX принято называть программы-серверы.)

Проблема свободного программного обеспечения заключается не только в свободе самого программного обеспечения, но и в свободе документации к нему.

«Критерий свободного руководства во многом такой же, как и критерий свободной программы... Повторное распространение (в том числе коммерческое) должно быть разрешено, чтобы руководство могло сопутствовать каждой копии программы, в интерактивной форме или на бумаге. Разрешение на внесение изменений также играет решающую роль».

Ричард Столмен.

Как правило, даже свободно распространяемое программное обеспечение часто имеет закрытую документацию, которую запрещено копировать, изменять, распространять. Идеально, чтобы документация к свободно распространяемому программному обеспечению также распространялась под лицензией GNU Public License.

Неправильно думать, что свободно распространяемое программное обеспечение всегда копируется бесплатно. Свобода и означает, что можно получить копию как бесплатно, так и заплатив. Просто существуют тысячи способов получения свободно распространяемого программного обеспечения, среди них указан и способ бесплатного получения. То же относится и к документации.

Однако вот вопрос: зачем платить, если можно получить бесплатно?

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

С интенсивным развитием компьютерных и информационных технологий неизмеримо возросла важность защиты информации. Давно известно, что информация бывает настоящим сокровищем. Именно поэтому часто много усилий затрачивается как на её охрану, так и на её получение. Информацию нужно защищать в тех случаях, когда есть опасения, что она станет доступной для посторонних, которые могут обратить её во вред законному пользователю. В таких случаях говорят о конфиденциальной информации.

Офис современной компании представляет собой наполненные электроникой помещения, где конфиденциальная информация находится на бумаге, на компьютерах, передаётся по телефону, факсу, обрабатывается, записывается и воспроизводится на магнитофонах, диктофонах, видеомагнитофонах, наконец, просто содержится в разговорах. Существует множество угроз нанесения ущерба системам обработки информации. В первую очередь он может быть вызван стихийными, природными явлениями, не зависящими от человека: пожаром, наводнением, землетрясением и ударом молнии...

Более широк и опасен тот ущерб, который связан с человеческой деятельностью:

• неумышленный ущерб, вызываемый ошибками в проектировании, в действиях обслуживающего персонала, программном обеспечении, случайными сбоями в работе средств вычислительной техники и линий связи, энергоснабжения, ошибками пользователей, воздействием на аппаратуру электромагнитных полей;

• умышленный ущерб, обусловленный несанкционированными действиями обслуживающего персонала и доступом к информации посторонних лиц.

Последствиями могут быть уничтожение, разрушение информации, а также искажение, подделка, утечка, копирование и т. п.

Как и все серьёзные мероприятия, защита информации должна осуществляться комплексно, т. е. для получения наилучших результатов все виды защиты информации нужно объединить. Комплексная система защиты информации, включает:

• организационную защиту, т.е. специальные мероприятия — от собраний до разработки планов и организации отделов по защите информации от НСД (несанкционированный доступ);

• программно-аппаратную защиту, компьютерные системы и программы;

• инженерно-техническую защиту: камеры видеонаблюдения, интеллектуальные замки, ограничивающие доступ в секретные помещения (к конфиденциальной информации) посторонних;

• законодательную защиту.

В России действует Федеральный закон «Об информации, информатизации и защите информации» от 20 февраля 1995 г. Закон определяет информацию как сведения о лицах, предметах, фактах, событиях, явлениях и процессах независимо от формы их представления. Он регулирует отношения, возникающие при формировании и использовании информационных ресурсов на основе создания, сбора, обработки, накопления, хранения, поиска, распространения информации; создании и использовании информационных технологий и средств их обеспечения; защите информации, прав субъектов, участвующих в информационных процессах и информатизации. Закон определяет, что защите подлежит любая документированная информация, неправомерное обращение с которой может нанести ущерб её собственнику, владельцу, пользователю и иному лицу. Закон запрещает сбор, хранение, использование и распространение информации о частной жизни, информации, нарушающей личную тайну, семейную тайну, тайну переписки, телефонных переговоров, почтовых сообщений граждан без их согласия, кроме как на основании судебного решения. Также защищается право на доступ к открытой информации.

ОШИБКИ И АВАРИИ

Ошибки и сбои в программном обеспечении так же опасны, как и компьютерные преступления. Услышав про очередной скандал, всегда хочется спросить, это ошибка или диверсия?

В столице Венесуэлы Каракасе Верховный суд отложил проведение выборов в мае 2000 г. из-за ошибки в компьютерной системе.

17 июня 2000 г. в Великобритании произошёл крупный сбой в системе авиаперелётов. В течение дня поломка была ликвидирована, но тысячам людей пришлось ночевать в аэропортах, менять время вылета. Отрадно, что все они остались в живых и сбой не привёл к авариям в воздухе.

Из-за неполадок, случившихся в операционной системе компьютера австралийской телефонной компании Telstra, 76 тыс. абонентов не могли звонить по своим аппаратам.

В millennium после установки недостаточно отлаженной компьютерной системы в Northeastern University (Северо-Восточный университет), США, было принято на 600 студентов (25%) больше, чем предполагалось. Руководство университета не знало, что делать с лишними зачисленными абитуриентами.

Беркли Банк заявил, что в результате установки более совершенного программного обеспечения нарушилась работа банка через Интернет, так как некоторые клиенты получили доступ к чужим счетам.

Резервный банк Австралии отправил 64 клиентам информацию о 0,5%-ном плановом увеличении процентной ставки на 6 мин раньше срока, и за это время биржевые брокеры ухитрились продать долговых обязательств и векселей на сумму 3 млрд австралийских долларов, составив целые состояния на этой ошибке.

Были и просто курьёзные случаи. Две женщины носили одно и то же имя — Белинда Ли Перри и родились в один день, 7 января 1969 г. Такое совпадение гарантировало им постоянную путаницу. Регулярно одна Белинда обнаруживала, что её данные в страховом полисе, банковских карточках заменены другими. Единственный положительный момент во всей этой истории состоит в том, что женщины подружились. Как видно, имя и дата рождения не подходят на роль уникального идентификатора.

Уголовным кодексом РФ предусмотрена ответственность за преступления в сфере компьютерной информации. К ним относятся:

• неправомерный доступ к охраняемой законом компьютерной информации, если это повлекло уничтожение, блокирование, модификацию либо копирование информации, нарушение работы ЭВМ или сети;

• создание, использование и распространение вредоносных программ или внесение изменений в существующие программы, заведомо приводящих к несанкционированному уничтожению, блокированию, модификации либо копированию информации, нарушению работы ЭВМ или сети;

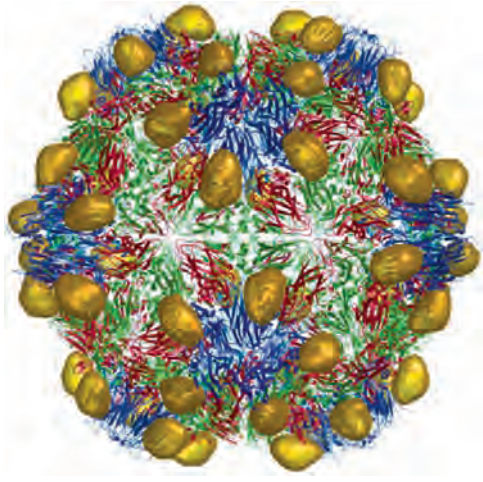
• нарушение правил эксплуатации ЭВМ или сети, повлекшее уничтожение, блокирование или модификацию охраняемой законом информации.

ВИРУСЫ

Цифровой век несёт не только освобождение людей от громоздких вычислений, автоматизацию всего, что можно автоматизировать, но и свои проблемы и заботы. Одной из таких проблем нашего времени являются *компьютерные вирусы*. Здесь всё как в жизни — есть свои болезни и свои доктора.

Компьтерный вирус — это программа, работающая на ЭВМ без ведома пользователя, наносящая обычно вред компьютеру, как обычный вирус — человеку, и способная создавать свои копии на доступных носителях информации (дисках, дискетах).

ВИРУСЫ, КОТОРЫЕ МЫ ВЫБИРАЕМ



Структура мозаичного вируса, вторгнувшегося в биологическую клетку.

Интересно отметить, что такое определение довольно точно описывает компьютерный вирус, но под него подпадает ещё множество других программ.

Например, операционные системы PC-DOS или MS-DOS и им подобные могут создавать свои копии. А в Windows работает множество частей системы, о которых пользователь даже не догадывается. Что касается вреда, то любая программа, имеющая в своём коде ошибки, способна нанести вашему компьютеру и хранящейся в ней информации большой урон.

Вирусы начали представлять серьёзную угрозу для компьютера где-то в 70—80-х гг. XX в. Авторы вирусов, пользуясь для создания новых «электронных приветов» различными недоработками в операционных системах компьютеров, наплодили огромное количество разной «заразы». Порой здравомыслящему человеку трудно судить о движущих ими мотивах. Кто-то просто проводил эксперименты, кто-то желал славы поджигателя-Герострата, кто-то всего лишь учился программированию.

Предыстория компьютерных вирусов начинается с появлением ЭВМ в учебных заведениях. Именно там талантливые студенты-программисты писали программы, которые «воевали» между собой за ресурсы компьютера — память, место на диске и даже за такой фактор, как *процессорное время* — миллисекунды, отпущенные процессору для решения одной

задачи перед переключением на следующую.

Первым нашумевшим вирусом стал так называемый вирус Морриса, который был написан в США и там же парализовал на длительный срок большое количество компьютеров в государственных учреждениях.

С появлением персональных компьютеров вирусы начали расти как грибы после дождя. Этому способствовали растущая популярность персональных ЭВМ и несовершенство программных средств для них.

Географию развития «вирусописательства» можно легко проследить во времени:

- США, Европа — начало 80-х гг.;
- Болгария, Индия — середина 80-х гг.;
- Россия (тогда ещё СССР) — конец 80-х гг.;
- Китай, Тайвань, другие страны Азии — начало и середина 90-х гг.

Итак, вирусы появлялись там, где вспыхивал очередной бум информатизации общества. Вначале Америка, потом Европа, затем Азия... Время действия — 80—90-е гг. XX в.

В Россию первые вирусы завозились вместе с бывшими в употреблении компьютерами из стран с развитой экономикой. Потом компьютеры стали покупаться в Болгарии и Индии, где они «заражались» порой уже на заводах (очевидно, по незнанию).

Вскоре в России появились свои «авторы» вирусов. Несмотря на то что множество людей просто «сдирали» код программ друг у друга, русский бум вирусописательства был одним из самых страшных для остального мира. Традиционно хорошая подготовка русских программистов, а также их своеобразное чувство юмора вводили западные компании, которые специализировались на «отлове» вирусов, в состояние лёгкого помешательства. Но в то же время в России появляются отечественные разработки антивирусных программ. Несомненно, они во многом превосходят зарубежные аналоги, так как именно российские программисты являются одними из самых талантливых в мире. Их разработки отмечены во многих странах и являются лидерами продаж. Позднее центр производства вирусов смещается ближе к Азии, а ещё позже центр как таковой исчезает. Причина тому — развитие Интернета, который фактически стирает границы.

Вирусы можно классифицировать по-разному:

- *по способу размножения*. Вирусы поражают исполняемые файлы, почтовые программы, даже файлы с документами. Современные текстовые редакторы имеют свои встроенные средства создания надстроек и дополнений. Этими средствами можно пользоваться не только для дополнительного удобства в работе, но и для создания вредоносных программ;
- *по способу действия*. Вирус может проявляться по-разному, спектр его действий — от демонстрации звуковых и визуальных эффектов до тихого стирания данных на компьютере-жертве;

Скорее всего, ваш компьютер поразил вирус, если Windows вывел надпись: «Ваши руки ввели неверную команду и будут ампутированы».

ИСТОРИЯ ОДНОГО ВИРУСА

2 ноября 1988 г. Роберт Моррис-младший, аспирант факультета информатики Корнеллского университета, с помощью написанного им вируса инфицировал большое количество компьютеров, подключённых к Интернету.

Роберт Моррис-младший.

Для своего распространения вирус использовал некоторые дефекты стандартной операционной системы UNIX, установленной на многих системах, а также механизм доступа к удалённым компьютерам в локальных сетях.

Вирус состоял из двух частей — главной программы и программы, обеспечивающей его распространение. Главная программа после запуска на очередной машине собирала информацию относительно других компьютеров в сети, с которыми та имеет связь. Затем производилась пересылка программы распространения на эти машины, где она запускалась и обеспечивала пересылку и компиляцию остальной части вируса. Затем весь процесс повторялся.

Наиболее заметным эффектом при распространении вируса была непрерывно возраставшая загруженность поражённых вирусом машин. По истечении определённого времени некоторые компьютеры оказались настолько загруженными распространением копий вируса, что не были способны выполнять никакой полезной работы; другие машины исчерпывали свою память, и их приходилось перезагружать.

• *по уровню наносимого вреда*. Вирусы могут быть как безобидными, так и весьма опасными.

Компьютерная программа — это набор инструкций для ЭВМ. Она хранится на постоянном носителе информации, а при запуске находится в оперативной памяти компьютера. Вирус, как уже было сказано, тоже программа. Её задача в первую очередь — изменить как можно больше программ на ЭВМ таким образом, чтобы они содержали в себе дополнение в виде самого вируса.

Заражение происходит следующим образом: пользователь запускает на своём компьютере программу, уже содержащую вирус. Это может быть программа, «скачанная» из Интернета или переписанная у знакомого. Перед загрузкой самой программы или после неё вирус активизируется и начинает работу. Сценарий его работы примерно таков:

1) найти все доступные для заражения программы на ЭВМ;

2) дописать себя в начало или в конец программы;

3) провести разрушительные действия, если настала «критическая» дата (тот день, в который вирус наносит «удар»);

4) если дата не настала, то, возможно, сделать какую-нибудь «мелкую» пакость, например зашифровать небольшой участок жёсткого диска ЭВМ.

А когда пользователь запустит «свою» программу, которая была у него на ЭВМ до заражения и теперь содержит в себе вирус, весь сценарий повторится.

Помимо этого, вирус может записать себя в специальную область на диске — загрузочный сектор. Это позволит ему активизироваться, когда пользователь включит компьютер и начнётся загрузка системы с жёсткого диска. Такой способ был популярен в то время, когда компьютер надо было загружать с дискет. Вообще, *загрузочные* вирусы не получили такого большого распространения, как *файловые*.

Вирусы, которые поражают файлы с программами, называются файловыми. Те же, которые «пишут себя» в загрузочный сектор, — загрузочными. А «монстры», поражающие и файлы, и загрузочный сектор, называются *файлово-загрузочными*.

Как правило, вирусы имеют очень маленький размер и пользуются не описанными в инструкциях к программам и специальной литературе возможностями операционной системы. Это позволяет им поражать файлы быстро и незаметно. Некоторые вирусы могут «маскироваться» — принимать вид запутанного и бессмысленного кода, который не понимает большинство антивирусных программ и который способен сам изменять себя в процессе работы. Такие «изделия» называются *полиморфами*.

Развитие Интернета сильно повлияло на скорость распространения вирусов. Кроме этого, вирусы приобрели совершенно новые качества — они получили возможность использовать Интернет и локальные компьютерные сети для переноса себя на другие машины и для связи со

своими «создателями». Очень часто распространители вирусов пользуются для заражения компьютеров-жертв оригинальными трюками. Эти приёмы затрагивают слабости человека — его интерес к «запретному плоду», т. е. бесплатным пиратским программам, «крэкам» и тому подобным соблазнам Всемирной сети. Схема действия этих программ такова: пользователь находит на просторах Интернета некую программу, сулящую ему бесплатный доступ к сайтам или, например, предлагающую взломать защиту популярного программного продукта. Он скачивает её на свой компьютер и запускает. Всё, дело сделано. Программа сообщает, что не может запуститься по той или иной причине, а сама «подсаживает» на компьютер пользователя вирус. Или, например, вам приходит письмо от незнакомого человека, который желает познакомиться и предлагает посмотреть фотографии, приложенные к письму. Только вот почему-то фотографии оказываются исполняемыми файлами... Такой приём позволит злоумышленнику или вообще всем желающим получить доступ к компьютеру жертвы и добыть из него любую информацию. Как правило, эти программы используются для получения паролей доступа в Интернет. Эта разновидность вирусов называется *«троянцем»* или *«троянским конём»*.

Помимо программ-«троянцев» существует такая разновидность вирусов, как *«черви»*. Эти программы, как и «троянцы», ориентированы на работу через Интернет. «Черви» переезжают с компьютера на компьютер по сети, и основной вред, который они причиняют, — это чрезмерное использование ресурсов компьютеров-жертв и каналов Интернета. Пример компьютерного «червя» — вирус Морриса.

Вирусы представляют реальную опасность для современных персональных компьютеров. Если 5—7 лет

САМЫЙ ИЗВЕСТНЫЙ «ТРОЯНЕЦ»

Самым нашумевшим «троянцем» стал, пожалуй, BackOrifice, имевший наибольший «успех» в 1997—1998 гг. Название его можно перевести как «Чёрный ход». Эта программа была задумана как система удалённого управления компьютером и позволяла делать буквально ВСЁ — начиная от выдвигания лотка дисковода CD ROM до изъятия информации из видеокамеры, подключённой к компьютеру-жертве, или записи звуков через микрофон. Небольшие изменения в коде BackOrifice позволили пользоваться им в неблаговидных целях, незаметно подсаживая вирус на компьютеры-жертвы.

Авторы программы даже не предполагали, что их детище произведёт столько шума. Они создали лишь инструмент для удалённого управления компьютером, но вскоре их программа стала настоящей напастью для пользователей Интернета. Пользователи пересылали друг другу файлы, к которым был «приклеен» BackOrifice, и при запуске вложенной в письмо программы вместе с ней запускался и BackOrifice. Он маскировался под безобидный процесс в системе, и пользователь не мог обнаружить его без помощи специальных антивирусных программ.

Вы платите только за операционную систему! Ошибки, недоработки и «дыры» в защите — совершенно бесплатно, т. е. даром!

Для удалённого управления чужим компьютером совершенно необязательно подсаживать на него «троянского коня», рискуя быть замеченным. Современные операционные системы громоздки и уязвимы, и, к сожалению, количество ошибок и «дыр» в защите от версии к версии не снижается. Последние версии популярного браузера Internet Explorer, входящего в состав ОС Windows, любезно предоставляют доступ к компьютеру пользователя без его ведома через просматриваемые им web-страницы. То есть во время просмотра web-сайта пользователю на компьютер может быть передан код, при исполнении которого будет создана возможность запустить ЛЮБУЮ программу на его компьютере.

назад основной целью автора вируса был просто вывод из строя компьютера жертвы, то в начале XXI в. вирусы нацелены главным образом на то, чтобы незаметно украсть у вас какую-то информацию и предоставить посторонним доступ к вашей ЭВМ. Конечно, не последнюю роль в истории компьютерных вирусов сыграло развитие Интернета. Число каналов распространения вирусов значительно возросло — это и электронная почта, и пиратские сайты, и даже просто документы. Вирус, который похищает информацию, может серьёзно «подмочить» репутацию коммерческой

фирмы, выставив на обозрение документы, не предназначенные для посторонних глаз. А если, например, представить, что такой вирус попадёт в компьютер, связанный с деятельностью военных или государственных организаций?

Убытки, наносимые компьютерными вирусами, на Западе оцениваются в сотни миллионов долларов.

Как же защититься от компьютерного вируса? Для этого существуют программы, называемые *антивирусами*.

Работает такая программа следующим образом. Она «знает» о том, как «выглядит» вирус. То есть имеет в своей базе данных фрагменты программ-вирусов. Базу данных можно регулярно пополнять, добавляя в неё «образцы» новых вирусов. При запуске антивирус ищет в файлах на вашем компьютере похожие фрагменты и, если находит, пытается провести процедуру «лечения».

«Лечение» происходит так: антивирус пытается вычистить ваши файлы, просто удаляя «критические» куски вируса, а затем устраняет последствия работы вируса, например восстанавливает испорченную информацию.

Существуют также программы-мониторы, которые постоянно находятся в памяти ЭВМ и отслеживают все критические моменты и подозрительные действия — запуск программ, попытки записи в критически важные системные файлы и т. д.

Но уповать на совершенство программ-антивирусов нельзя. Для того чтобы не подцепить какую-нибудь «инфекцию», нужно придерживаться простых правил:

- не давайте работать на вашем компьютере посторонним;
- пользуйтесь легальным программным обеспечением. Порой в пиратских копиях обнаруживаются целые коллекции «троянцев»;
- получив почту от незнакомых людей, убедитесь, что с письмом не «приехал» какой-нибудь «довесок» в виде исполняемого файла;
- не скачивайте из Интернета программы без разбору, даже если они обещают вам 300%-ное увеличение работоспособности вашего компью-

тера. Лучше поищите информацию о них на сайтах с обзорами программ.

Кто же является автором вирусов? Злодеи, мечтающие завоевать мир? Специальные отделы фирм, которые занимаются борьбой с конкурентами? Диверсанты? Обычно нет. Авторами часто становятся студенты вузов, порой даже школьники старших классов. Что движет этими людьми? В первую очередь научный интерес. Авторам любопытны все стороны компьютерной жизни, но они не всегда понимают, какой вред могут нанести своими действиями...

Порой они сами не подозревают, что их детище вырвалось на свободу и зажило собственной жизнью.

В Уголовном кодексе РФ существует статья, под которую подпадают действия авторов вирусов — создание и распространение вредоносных программ для ЭВМ. Так что, если автор «попался», ему грозит тюремное заключение.

Какие же перспективы на ближайшее время у компьютерных вирусов и антивирусных программ?

С антивирусами ситуация вполне ясна: они будут совершенствоваться

и дальше, алгоритмы тоже будут совершенствоваться, поиск вирусов и «лечение» станут эффективнее.

Что касается вирусов, то здесь ситуация сложнее, хотя некоторые прогнозы сделать всё-таки можно. Во-первых, мир переживёт ещё не одну волну «вирусописательства». Во-вторых, для вирусов открывается большое поле деятельности, и целью их уже стали не только домашние и офисные ЭВМ, но и карманные компьютеры, а также сотовые телефоны и прочие сложные электронные приборы, в которых есть процессор.

ТЕСТОВОЕ ПРОНИКНОВЕНИЕ

«Я положил перед сидевшим молча президентом компании технологию изготовления самого важного из продуктов. Президент продолжал молчать, когда на его стол лёг развёрнутый план всех разработок компании. Президент откинулся на спинку кресла, когда к стопке добавились документы, описывающие переговоры о многомиллионном судебном процессе. Наконец, президент промолвил: "Мы должны благодарить Бога, что вы не работаете на наших конкурентов".

Я украл всё это и ещё кое-что, будучи временным работником. Думаете, речь идёт о компании с плохой службой безопасности? Едва ли. Внешняя защита организована превосходно, однако в руководстве подозревали, что компания может быть уязвима изнутри. Ко мне обратились, чтобы проверить, что может украсть целенаправленно действующий информационный вор.

На работу в компании мне отвели три дня. Я украл всё...»

Так начинается рассказ Айрэ Винклера, автора книги «Корпоративный шпионаж», одного из самых известных экспертов в области информационной безопасности, основателя и президента Internet Security Advisors Group (ISAG, примерный перевод «группа экспертов по безопасности в Интернете»), непревзойдённого специалиста в тестовом проникновении. Он описывает методы шпионажа, которые обнаруживают брешь в системе безопасности.

Информация любой корпорации тщательно охраняется. При этом

Айрэ Винклер.

средства защиты подчас сопоставимы со средствами, применяемыми в Форт-Ноксе. Прямой взлом, подобный тем, что показывают в кинофильмах, практически исключён. Никакой серьёзной службе безопасности не придёт в голову использовать последнюю версию Windows на компьютерах корпорации, имеющих выход в Интернет. Оставленные открытыми «калитки» в подобных несерьёзных операционных системах не требуют никаких специальных средств для взлома. Надо просто войти. А список «дыр» постоянно обновляется, и любой желающий может получить его в Интернете.

Другое дело — операционные системы, сертифицированные специальными комиссиями по определённым классам безопасности. В США, Европе и России они обозначаются по-разному, однако требования, предъявляемые к ним, одинаково высоки.

Но есть ещё один метод, самый изощрённый — взлом изнутри. Для этого нужно провести определённую работу.

При шпионаже используются абсолютно все доступные методы, или *атаки,* — как технические, так и нетехнические.

Усиленно охраняемое хранилище золотого запаса США Форт-Нокс находится почти в центре военного городка с тем же названием.

Итак, в шпионаже против корпорации Винклер применил пять видов атак:

- поиск по открытым источникам;
- маскарад;
- превышение прав доступа;
- хакерство штатного сотрудника;
- внутренняя координация действий внешних сообщников.

ПОИСК ПО ОТКРЫТЫМ ИСТОЧНИКАМ

Прежде всего необходимо досконально ознакомиться с объектом атак. Удивительно, какое огромное количество сведений можно просто почерпнуть из Интернета: узнать о разработках компании, прочитать научные статьи об этих исследованиях. На сайте компании, как правило, указаны имена руководителей, её финансовое положение, множество общей информации о компании и её корпоративной философии. Иногда удаётся выяснить имена ведущих специалистов. Обычный поиск по названию компании в группах новостей Интернета может принести удивительные сюрпризы. Айрэ Винклер узнал не только всё вышеперечисленное, но и нашёл имена многих сотрудников компании. Письма сотрудников в компьютерные группы новостей дали информацию об аппаратной и программной среде компании, а письма в прочие группы помогли узнать о личных интересах авторов.

МАСКАРАД

Шпион должен быть артистом, причём хороший шпион — вдвойне артистом, чтобы не провести десяток лет в тюрьме в случае разоблачения. Человеку свойственно с подозрением относиться к чужакам и принимать с распростёртыми объятиями тех, кто работает с ним в одной корпорации.

Маскарад начался с изготовления визитных карточек (в качестве образца использовалась подлинная визитная карточка сотрудника компании), правда, должность была изменена: из

мелкого временного работника шпион превратился в сотрудника информационной безопасности.

Только поступление на временную работу обошлось без маскарада, но тем не менее бланки временного служащего заполнялись ложной информацией: поддельные номер социального страхования, адрес и телефон.

У каждого шпиона свой стиль. Но удачно выбранная должность позволяет творить чудеса. Винклер придумал легенду о защите информации компании. Исследователю, работавшему над важнейшим проектом, он сказал, что недавно принят в штат и что ему поручено выяснить, какая часть проекта наиболее нуждается в защите и где хранится соответствующая информация. Тот посоветовал обратиться к руководителю проекта. Айрэ Винклер так и сделал. Встретившись с руководителем проекта и вручив ему визитную карточку, он попросил объяснить, какая информация является критичной и кто имеет к ней доступ. Ему ответили, что важнейшая информация — сведения об изготовлении продуктов компании, и показали основной источник этих сведений — копии протоколов заседаний проектной группы и перечень сотрудников, получающих эти протоколы. Руководитель проекта не только сделал для шпиона копии всех документов, но и добавил его в список рассылки протоколов.

Подобные беседы с другими ведущими сотрудниками компании дополнили информацию. Маскарад поистине приносит богатый урожай, однако не стоит забывать и про случайность. Среди полученной конфиденциальной информации значился секретный документ с указанием директории и пароля для доступа. Рядом с секретным документом стоимостью больше миллиарда долларов в том же каталоге находились и другие, описывающие ещё две важные разработки компании. Естественно, документы были тут же скопированы.

ПРЕВЫШЕНИЕ ПРАВ ДОСТУПА

Как правило, в офисе компании используется более слабая защита компьютеров, чем от взлома извне. Например, дистанционно нельзя проникнуть в конфиденциальную сеть компании, так как она просто может быть не подсоединена к внешним каналам связи. Однако внутри компании шпиону удастся взломать компьютер сотрудника, если он подберёт пароль. Часто работники компании не утруждают себя придумыванием более или менее сложных паролей, в основном они

Кадр из фильма «Крепкий орешек».

используют имена собственные: Barbie, Ken, Starwars и т. п.

Имея пароль, легко получить доступ к конфиденциальной и даже секретной информации, хранящейся в машине высокопоставленного сотрудника.

При шпионаже важна любая мелочь . Отсутствие замка на двери одного из ведущих сотрудников или разбросанные по столу дискеты с надписью: «Отчёты компании» — дают почву для создания примитивного плана хищения.

Шпион работает как уборщик, ищет незапертые кабинеты, шкафы и ящики. Его интересуют мусорные корзины (там могут быть записки с системными паролями), компьютеры, на которых не выполнена процедура завершения сеанса (иногда просто выключают монитор, оставляя включённым системный блок).

Винклеру повезло. Оставленный компьютер с сохранёнными электронными письмами позволил скопировать основной график разработок — один из самых секретных документов компании.

ХАКЕРСТВО ШТАТНОГО СОТРУДНИКА

При современном уровне развития вычислительной техники переносные компьютеры стали действительно миниатюрны. Это позволяет шпиону использовать собственный ноутбук, заранее подготовленный для хакерства. В Интернете чрезвычайно много хакерских инструментов, помогающих «взламывать» системы, используя бреши, выявленные сканером (хакерская программа). Шпиону достаточно в своём кабинете отключить офисный компьютер от сети Ethernet и подсоединить ноутбук.

Сканер напускают на компьютеры компании, и если обновление средств защиты хоть немного отстаёт от выявленных «дыр» в системе, о которых открыто сообщается в Интернете, то судьба таких машин предрешена. Иногда шпиону даже удаётся получить права суперпользователя. В любом случае брешь позволит скопировать на хакерский ноутбук секретную информацию со взломанных машин. Затем наступает пора подбора паролей.

ВНУТРЕННЯЯ КООРДИНАЦИЯ ДЕЙСТВИЙ ВНЕШНИХ СООБЩНИКОВ

Если у шпиона есть внешняя поддержка, то соответствующая команда способна взломать внешнюю сеть компании. Для этого требуются входное имя, пароль и номер телефона для модемного доступа. Обычно сотрудники компании имеют такой доступ. Эту информацию можно передать, например, по телефону или даже послать с курьером.

Если шпион вдобавок обладает списком входных имён и паролей сотрудников компании, то сообщники могут проникнуть практически во все компьютерные системы компании. Средства удалённого доступа, спроектированные для защиты от несанкционированного доступа, не смогут остановить авторизованных пользователей.

Для того чтобы не выдать себя, шпион не покидает компанию. Поэтому действия хакеров-сообщников бросают тень на честных сотрудников компании.

Айрэ Винклер за три дня добыл более 300 Мбайт секретных данных.

У него в руках оказалась детальная информация о технологии производства пяти наиболее важных продуктов компании. Кроме того, он получил информацию практически обо всех новых разработках. Попади всё это к конкурентам, компания могла быть разорена.

Пока шпион действовал, никто не заметил ничего подозрительного, несмотря на грубые методы.

Арсенал профессиональных промышленных шпионов значительно шире. Можно, например, установить «жучки» и прослушать телефонные линии или завербовать сотрудников. Шпионы тратят больше времени на подготовку и планирование, чем на осуществление реальной атаки.

Компания с прекрасно организованным внешним рубежом обороны беззащитна перед шпионажем штатного работника, когда внешняя защита перестаёт действовать, а секретная корпоративная информация становится доступна.

ХАКЕРЫ

«Теперь это наш мир! Мир электроники и коммуникаций, мир красоты скорости передачи данных. Для нас не существует национальности, цвета кожи и религиозных предрассудков. Вы развязываете войны, убиваете, лжёте и пытаетесь запудрить нам мозги, говоря, что всё это для нашего блага, а в итоге преступники — это мы! Да, я преступник! Меня толкает на преступление моя любознательность. Я — ХАКЕР, и это мой манифест! Вы можете поймать меня, но всех НАС вам не переловить» (из «Манифеста хакера»).

Кто же такие хакеры? Это слово прочно вошло в обиход всего за десяток лет.

Словарь даёт следующее толкование английского глагола to hack:

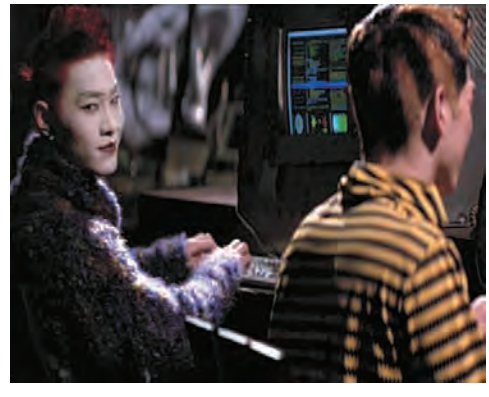
- рубить, разрубать, кромсать, разбивать на куски;
- тесать, обтёсывать, делать зарубку, зазубривать;
- использовать на нудной, тяжелой работе;
- использовать в качестве литературного подёнщика.

Хакерами называют людей, виртуозно владеющих компьютерными технологиями, программистов-профессионалов, специалистов по взлому компьютерных сетей и операционных систем. Вообще, в этом слове заложена идеология целого поколения, его стиль жизни и философия.

Вот типичный портрет хакера. Это молодой человек от 15 до 35 лет, всерьёз, до сумасшествия, увлечённый компьютерами, нестандартно мыслящий, ведущий замкнутый образ жизни, малообщительный (практически нет друзей), окружающие считают его тихоней. Он хорошо знает математику и физику, редко занимается спортом, и вся жизнь его — это Компьютер, в школе, в университете, на работе и дома.

Кинематограф подарил нам, по крайней мере, два замечательных фильма о хакерах — «Военные игры» (War Games) режиссёра Джона Бэдэма (John Badham) и «Хакеры» (Hackers) режиссёра Айэйна Софтли (Iain Softley).

В первом фильме американский школьник, играющий на персональном компьютере в видеоигру «Всеобщая ядерная война», неожиданно подключается к компьютерной системе Пентагона (военное ведомство США). Война на экране чуть не обратилась в реальность...



Кадр из фильма «Хакеры».

Британское Министерство гражданской авиации указало на ряд случаев, когда радиохакеры внедрялись в систему контроля за полётами и давали неправильные инструкции. Количество таких происшествий растёт:
в 1988 г. — 3,
в 1999 г. — 18,
в 2000 г. — 20.

В «Хакерах» действие происходит в Нью-Йорке. Родители двенадцатилетнего компьютерного гения оштрафованы на крупную сумму, а сам он осуждён на семь лет условно, т. е. без заключения в тюрьму, за заброску вирусов в полторы тысячи локальных компьютерных сетей и панику на бирже Уолл-стрит. Ему запрещено пользоваться компьютером и телефоном до 18 лет. В новой школе он знакомится с компанией таких же хакеров. Один из них случайно проникает в сеть крупнейшей корпорации и переписывает файлы, разоблачающие главу службы компьютерной безопасности, который использовал вирус-«червь» для кражи миллионов долларов. Вор, узнав об этом, решил обвинить в грабеже хакеров, но все хакеры мира объединились с помощью Интернета и начали войну с циничным злодеем...

В реальной жизни хакеры часто работают программистами или специалистами по компьютерной технике в больших и маленьких фирмах. Нет такой профессии — хакер. Скорее это состояние души и образ мысли. В детстве многие ломают игрушки для того, чтобы понять, что находится у них внутри. Но дети вырастают, и вместо игрушек у некоторых из них в руках оказываются компьютер и новейшая операционная система. Любопытство толкает их разобрать «по винтикам» программу и сам компьютер, изучать и что-то улучшать. И порой из этого выходит что-нибудь по-настоящему полезное.

Современные операционные системы весьма сложны и громоздки, а чем сложнее система, тем более она уязвима с точки зрения безопасности и стабильности работы. Первыми такие ошибки и недоработки в программах обычно находят хакеры и, как правило, сообщают о найденных неполадках производителям.

То же самое относится к web-сайтам. Сотни сайтов в Интернете ежедневно исследуются хакерами с целью выявления брешей в защите.

В шутку говорят, что хакер — это человек, умеющий нажимать нужные клавиши в определённом порядке. Поэтому иногда хакеров путают с обычными студентами, которые, воспользовавшись программами для взлома (их полным-полно в Интернете), проникают на сайты университетов, благотворительных организаций, школ и портят их себе на потеху. Такие сайты наиболее уязвимы в плане безопасности, поскольку не содержат секретных материалов и поддерживаются бесплатно. К сожалению, такие хулиганские взломы чрезвычайно часты в Интернете. Но подобный вандализм имеет мало общего с хакерством, хотя компьютерные хулиганы гордо именуют себя таковыми. Если взять ружьё и начать для интереса стрелять по прохожим, то всё равно не станешь суперменом.

Вообще, хакерская деятельность с точки зрения закона трактуется не всегда однозначно. С одной стороны, в международном законодательстве существуют статьи, которые предусматривают не самый маленький срок тюремного заключения за незаконное проникновение в компьютерные системы и базы данных. Конечно, хакер, получивший доступ к конфиденциальной информации без ведома

«НЕЗАПЕРТЫЕ ДВЕРИ»

Хакерами пугают многих, но порой воровство возможно не благодаря суперинтеллекту молодых программистов, а просто из-за «незапертых дверей».

В конце 1999 г. 19-летний русский хакер проник на web-сайт CD Universe, Интернет-магазина, специализирующегося на продаже музыкальных дисков, и завладел сведениями о 300 тыс. кредитных карточек покупателей этой компании. Он потребовал 100 тыс. долларов, но, получив отказ, начал публиковать на своём сайте информацию о 25 тыс. кредиток. Правда, эту электронную страницу закрыли, и всё же некоторыми данными успели воспользоваться посторонние люди.

В 2000 г. ФБР арестовало 18-летнего Рафаэля Грея, который взломал девять Интернет-магазинов и похитил информацию о 26 тыс. кредитных карт в США, Канаде, Таиланде, Японии и Британии. (Кстати, одна из кредиток принадлежала Биллу Гейтсу, «королю» Microsoft.) Своим поступком он хотел показать, насколько не защищены сайты.

В России хакеры ухитрились попасть в компьютерные системы, принадлежащие Газпрому — российской газовой монополии. Согласно данным Министерства внутренних дел России, им удалось взять под контроль систему потока газа в трубопроводах.

властей или владельцев, является преступником. Но, с другой стороны, если эта информация просто лежит «на поверхности» и каждый, кто хоть немного знаком с устройством операционной системы и web-сервера, может с лёгкостью получить к ней доступ, то можно ли судить человека за это? Или такой пример: после проведения взлома хакер оповещает о существующей «дыре» в защите программы владельцев, тем самым фактически «закрывая» её. В этом случае, наверное, нельзя обвинять хакера в преступлении.

Конечно, как в реальной жизни существуют преступники, так и в компьютерном мире имеются свои нарушители законов. Объектами их атак становятся банки различного масштаба, биржи, Интернет-магазины. А поскольку закон един для всех и нарушать его нельзя даже гениям, любое противозаконное деяние хакеров будет раскрыто, а виновный — наказан. По крайней мере, хочется в это верить.

В Германии однажды был пойман хакер, работавший в банке. Он написал программу, которая переводила на его денежный счёт доли пфеннигов (мелкая монета немецкой валюты, имевшая хождение до введения евро), остававшиеся после начисления процентов вкладчикам. За год на его счету скопилась довольно крупная сумма.

КОМПЬЮТЕРНЫЕ ПРЕСТУПЛЕНИЯ

Убытки от компьютерных ошибок и преступлений огромны.

По данным Американского национального центра информации по компьютерной преступности, только за один год она нанесла американским фирмам убытки в размере 500 млн долларов. В Германии компьютерная мафия похищает за год около 2 млрд евро. В конце XX в. во Франции аналогичные потери достигли 1 млрд франков в год. Количество подобных преступлений ежегодно увеличивается на 30—40%.

Сообщения об информационных преступлениях отрывочны. Пожалуй, никто в мире не имеет сегодня полной картины нарушений безопасности информации. Всего (по самым скромным подсчётам) ежегодные потери от компьютерной преступности в Европе и Америке составляют несколько десятков миллиардов долларов. Понятно, что пострадавшие государственные и коммерческие структуры не очень склонны афишировать «эффективность» своих систем защиты и последствия, причинённые нападениями.

Поэтому случаи преступлений становятся достоянием гласности далеко не всегда. Но и те факты, которые известны, производят сильное впечатление. И не зря Подкомитет ООН по преступности ставит эту проблему в один ряд с терроризмом и наркотиками.

Способны ли развитые страны оказать отпор этому бичу XXI в.?

Администрация президента США в 2000 г. предложила национальный план по борьбе с киберпреступлениями. Был создан новый Институт по защите информационных инфра-

ПРЕДНАМЕРЕННОЕ ВРЕДИТЕЛЬСТВО

Компьютерные преступления наносят подчас непоправимый вред компаниям и государственным структурам. Уволенные или озлобленные системные администраторы и программисты бывают опаснее, чем профессиональные шпионы конкурентов.

35-летний компьютерный оператор лишил своего работодателя полумиллиона фунтов стерлингов. Не будучи удовлетворён условиями работы и размерами вознаграждения, он просто отсоединил несколько кабелей от центрального сервера компании. После отключения кабелей компьютерная система стала периодически давать сбои, и только вызванный квалифицированный специалист смог обнаружить диверсию. Адвокат обвиняемого пытался доказать, что его подзащитный сошёл с ума от работы, однако судья приговорил вредителя к тюремному заключению.

В результате злонамеренных действий служащих одной компании, вероятно тоже недовольных начальством, на открытый сайт в Интернете попала конфиденциальная информация о счетах дилеров компании, включая номера банковских счетов и кредитных карточек. Сведения были закрыты лишь спустя несколько часов. Кто знает, сколько преступников успели воспользоваться этой информацией.

Нередко так ведут себя не только люди, но и целые компании. Подобные действия могут являться основой нечистоплотной борьбы с конкурентами.

Европейский союз в 2000 г. выдвинул обвинения против компании Microsoft, обвинив её в том, что операционные системы серии Windows 2000 построены таким образом, что могут устойчиво работать исключительно с продукцией Microsoft. Это, естественно, не только ущемляет интересы других компаний (Microsoft на европейском рынке получает до четверти своих ежегодных доходов), но и наносит ущерб конечным пользователям, заставляя их приобретать лишь продукты названной компании.

ПРОМЫШЛЕННЫЙ ШПИОНАЖ

В XXI столетии Европейский парламент снова обратил внимание на американский спутник ECHELON, управляемый спецслужбами США, Великобритании, Австралии, Канады и Новой Зеландии. Спутник способен перехватывать сигналы телефона, факса, электронной почты по всему миру и предназначен для борьбы с терроризмом и другими опасностями, угрожающими США и союзникам. Однако существует подозрение, что спутник используется также и для коммерческого шпионажа против бизнеса единой Европы в пользу США. Европейский парламент намерен выяснить, насколько это соответствует истине.

Ещё пример. Подразделение компании Johnson & Johnson, производящее продукты для диабетиков, было обвинено в поощрении промышленного шпионажа. Основанием послужило присуждение премий «Inspector Clouseau» и «Columbo» сотрудникам, добывшим наибольшее количество сведений о конкурентах.

структур, а финансирование научных исследований в этой области превысило 600 млн долларов. В том же году ФБР увеличило штат на 150 специалистов по компьютерам и юристов для предотвращения преступлений в Интернете. В свою очередь конгресс выделил ФБР 38 млн долларов на эти расходы.

ФБР использует современную систему автоматизированного просмотра электронной почты на наличие информации криминального или террористического содержания. Эта система позволила выявить и раскрыть около сотни преступлений.

Правительство Великобритании заявило, что Национальная служба по криминальным преступлениям должна создать специальную команду для борьбы с Интернет-преступлениями, такими, как обман, отмывание денег, нелегальные азартные игры, порнография.

В борьбе с компьютерными преступлениями Европейская комиссия предлагает привлекать Интерпол, ведь компьютерная преступность не имеет ни национальности, ни границ.

ПИРАТЫ

Как известно, программное обеспечение можно приобрести двумя путями.

Способ первый — купить лицензионную версию у представителя компании-производителя и пользоваться всеми возможностями такого приобретения: поддержкой, обновлениями. Кроме того, сознание, что он обладает легальной копией программы, даёт пользователю уверенность в качестве продукта.

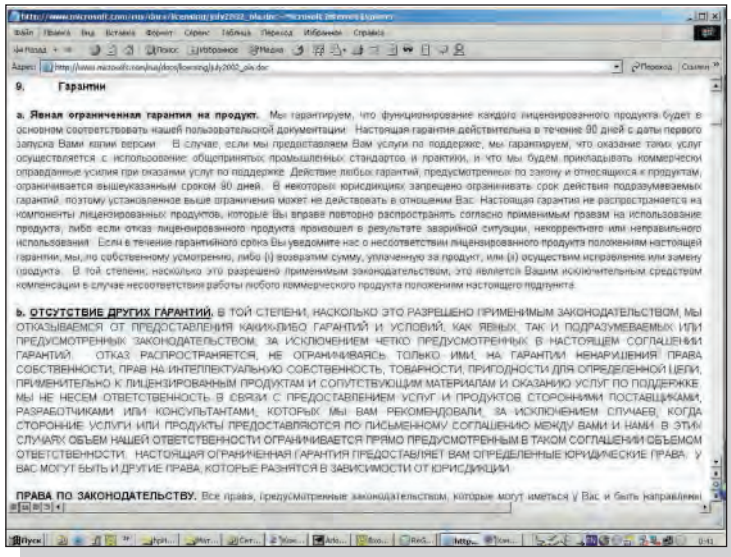
Способ второй — поехать на ближайший рынок, где торгуют дисками с программами, и купить их в любом количестве по цене порядка двух долларов за диск.

Каждый из способов имеет свои достоинства. В первом случае теоретически — это поддержка производителя программ, во втором — доступность.

Можно долго рассуждать о достоинствах легальных копий программ, но и здесь имеются некоторые нюансы, на которые стоит обратить внимание.

Посмотрим на всё с точки зрения конкретного потребителя.

Некто привёз домой вожделенную «коробочку», в которой хранятся один или несколько дисков с программой, инструкции, реклама других программ и т. д. Он убеждён, что при покупке продавец даёт гарантию работоспособности продукта, особенно если он дорогой. Остаётся только вставить



9. Гарантии

а. Явная ограниченная гарантия на продукт. Мы гарантируем, что функционирование каждого лицензированного продукта будет в основном соответствовать нашей пользовательской документации. Настоящая гарантия действительна в течение 90 дней с даты первого запуска Вами копии версии. В случае, если мы предоставляем Вам услуги по поддержке, мы гарантируем, что оказание таких услуг осуществляется с использованием общепринятых промышленных стандартов и практики, и что мы будем прикладывать коммерчески оправданные усилия при оказании услуг по поддержке. Действие любых гарантий, предусмотренных по закону и относящихся к продуктам, ограничивается вышеуказанным сроком 90 дней. В некоторых юрисдикциях запрещено ограничивать срок действия подразумеваемых гарантий, поэтому установленное выше ограничение может не действовать в отношении Вас. Настоящая гарантия не распространяется на компоненты лицензированных продуктов, которые Вы вправе повторно распространять согласно применимым правам на использование продукта, либо если отказ лицензированного продукта произошел в результате аварийной ситуации, некорректного или неправильного использования. Если в течение гарантийного срока Вы уведомите нас о несоответствии лицензированного продукта положениям настоящей гарантии, мы, по собственному усмотрению, либо (i) возвратим сумму, уплаченную за продукт, или (ii) осуществим исправление или замену продукта. В той степени, насколько это разрешено применимым законодательством, это является Вашим исключительным средством компенсации в случае несоответствия работы любого коммерческого продукта положениям настоящего подпункта.

b. ОТСУТСТВИЕ ДРУГИХ ГАРАНТИЙ. В ТОЙ СТЕПЕНИ, НАСКОЛЬКО ЭТО РАЗРЕШЕНО ПРИМЕНИМЫМ ЗАКОНОДАТЕЛЬСТВОМ, МЫ ОТКАЗЫВАЕМСЯ ОТ ПРЕДОСТАВЛЕНИЯ КАКИХ-ЛИБО ГАРАНТИЙ И УСЛОВИЙ, КАК ЯВНЫХ, ТАК И ПОДРАЗУМЕВАЕМЫХ ИЛИ ПРЕДУСМОТРЕННЫХ ЗАКОНОДАТЕЛЬСТВОМ, ЗА ИСКЛЮЧЕНИЕМ ЧЕТКО ПРЕДУСМОТРЕННЫХ В НАСТОЯЩЕМ СОГЛАШЕНИИ ГАРАНТИЙ. ОТКАЗ РАСПРОСТРАНЯЕТСЯ, НЕ ОГРАНИЧИВАЯСЬ ТОЛЬКО ИМИ, НА ГАРАНТИИ НЕНАРУШЕНИЯ ПРАВА СОБСТВЕННОСТИ, ПРАВ НА ИНТЕЛЛЕКТУАЛЬНУЮ СОБСТВЕННОСТЬ, ТОВАРНОСТИ, ПРИГОДНОСТИ ДЛЯ ОПРЕДЕЛЕННОЙ ЦЕЛИ, ПРИМЕНИТЕЛЬНО К ЛИЦЕНЗИРОВАННЫМ ПРОДУКТАМ И СОПУТСТВУЮЩИМ МАТЕРИАЛАМ И ОКАЗАНИЮ УСЛУГ ПО ПОДДЕРЖКЕ. МЫ НЕ НЕСЕМ ОТВЕТСТВЕННОСТЬ В СВЯЗИ С ПРЕДОСТАВЛЕНИЕМ УСЛУГ И ПРОДУКТОВ СТОРОННИМИ ПОСТАВЩИКАМИ, РАЗРАБОТЧИКАМИ ИЛИ КОНСУЛЬТАНТАМИ, КОТОРЫХ МЫ ВАМ РЕКОМЕНДОВАЛИ, ЗА ИСКЛЮЧЕНИЕМ СЛУЧАЕВ, КОГДА СТОРОННИЕ УСЛУГИ ИЛИ ПРОДУКТЫ ПРЕДОСТАВЛЯЮТСЯ ПО ПИСЬМЕННОМУ СОГЛАШЕНИЮ МЕЖДУ ВАМИ И НАМИ. В ЭТИХ СЛУЧАЯХ ОБЪЕМ НАШЕЙ ОТВЕТСТВЕННОСТИ ОГРАНИЧИВАЕТСЯ ПРЯМО ПРЕДУСМОТРЕННЫМ В ТАКОМ СОГЛАШЕНИИ ОБЪЕМОМ ОТВЕТСТВЕННОСТИ. НАСТОЯЩАЯ ОГРАНИЧЕННАЯ ГАРАНТИЯ ПРЕДОСТАВЛЯЕТ ВАМ ОПРЕДЕЛЕННЫЕ ЮРИДИЧЕСКИЕ ПРАВА. У ВАС МОГУТ БЫТЬ И ДРУГИЕ ПРАВА, КОТОРЫЕ РАЗНЯТСЯ В ЗАВИСИМОСТИ ОТ ЮРИСДИКЦИИ.

ПРАВА ПО ЗАКОНОДАТЕЛЬСТВУ. Все права, предусмотренные законодательством, которые могут иметься у Вас и быть направлены

Выдержка из лицензии ПО фирмы Microsoft (фотография с экрана).

диск в компьютер и запустить программу установки...

Первое, что покупатель увидит, — это так называемое лицензионное соглашение, в котором сказано, что программа распространяется по принципу As is — «как есть», со всеми ошибками и недоработками. То есть если в результате работы программы у пользователя что-то сломается или пропадут данные на компьютере, то попытка возместить ущерб через суд или иным способом вряд ли приведёт к каким-либо результатам. Отлично! Так за что же заплачена немалая сумма?

Допустим, у пользователя останется сознание того, что руку помощи протянет служба поддержки.

Но вот программа установлена, и тут обнаруживается: что-то идёт не так. Самое время обратиться в службу поддержки... А там отвечают (можно просто выбрать один из вариантов):

• «Установите Service Pack (заплатку)». Хорошо, что есть соединение с Интернетом, пусть даже по телефону. Конечно, три часа на скачивание этой «заплатки» — просто ничто, хотя родители, сестра, жена (нужное подчеркнуть) будут слегка недовольны тем, что телефон занят продолжительное время. Плюс ещё два часа в случае плохой связи. Плюс ещё три, если файл «приехал» с ошибкой. И так далее.

• «Ваше "железо" несовместимо с нашей программой». Например, блок питания. Или корпус. Или коврик для мыши явно не того оттенка. Как странно, а продавец не предупредил, да и на коробочке ничего такого не написано... А коробочка вскрыта, вернуть не получится. Можно, конечно, пообщаться с представителями фирмы, но надежды, что это поможет, почти нет, так как личные проблемы пользователя никого не интересуют.

• «Ждите новой версии нашей замечательной программы». Сколько ждать? И сколько платить за обновление? Где гарантии того, что новая версия их замечательной программы заработает как надо?

Главное «оружие» легальных копий — их поддержка производителем (по крайней мере, эти самые производители хотят, чтобы все так думали). Но порой российские реалии не дают ею воспользоваться, потому что:

• её просто нет, так как фирма-представитель только продаёт товар, а поддержкой занимается какое-нибудь европейское бюро где-нибудь в далёком Дублине;

• поддержка платная. Точнее, сначала она, конечно, бесплатная, а вот потом может быть всё что угодно. То есть первые два звонка в службу поддержки будут бесплатными, а за остальные — кругленькая сумма.

Получается, что разрекламированная поддержка программного обеспечения — порой лишь миф и трюк для привлечения покупателя.

Утверждается, что легальные покупки программ поддерживают индустрию программного обеспечения. Так ли это?

Представим себе ситуацию, что пиратов, взломщиков программ, распространителей нелегальных копий просто нет. Замечательно! А теперь предположим, что покупатель смог «наскрести» 500 долларов на компьютер, а вот на программы денег не осталось. Точнее, остались, конечно, но те программы, которые нужны для того, чтобы пользоваться компьютером, а не вести «войну» с настройками всю оставшуюся жизнь, стоят значительно дороже.

Компьютер желают иметь многие, и вся ситуация говорит о том, что если бы не было бесплатных пиратских ко-

пий, то тот информационный ажиотаж, который переживают Россия и другие страны с тяжёлым прошлым и не менее сложным настоящим, мог бы и не состояться. События, вероятно, развивались бы совсем по-другому. Компьютер не стал бы частью жизни обычного современного человека, а оставался уделом специалистов. И конечно, все продавцы легального программного обеспечения продавали бы что-нибудь другое. Например, пылесосы или сантехнику.

Как ни странно, Россия занимает не первое место по количеству пиратских копий программ. Лидером в этом отношении является Китай (70 % рынка).

Получается замкнутый круг. Потенциальный покупатель фирменного продукта на начальном этапе чаще всего пользуется программой благодаря пиратам, благо крупные фирмы не препятствуют этому, а производитель несёт убытки тоже благодаря пиратам.

Решение этой проблемы — в создании благоприятных условий для продаж легальных копий. Так, если бы лицензионная программа стоила не 200 долларов, а, например, 2, то пиратам, возможно, невыгодно было бы продавать всё это из-под полы. Сейчас функции удешевления программ выполняют как раз пираты, а крупные фирмы пока что находятся в раздумьях о том, как с этими пиратами бороться, вместо того чтобы самим создать информационный ажиотаж, снизив цены на программное обеспечение.

Вопрос о компьютерном пиратстве поднимается в средствах массовой информации уже давно — примерно с того времени, когда компьютер появился в домах, а носители данных стали достаточно дешёвыми, т. е. с начала 80-х гг. Тема эта затрагивается во множестве изданий, начиная от детского «Хакера» (новое воплощение «Мурзилки» и «Весёлых картинок» для поколения MTV) и заканчивая изданием для интеллектуальной элиты — журналом «Компьютерра». В каких-то публикациях пираты выглядят бравыми молодцами из кино про хакеров, а в каких-то — воплощением вселенского зла. Но ведь они в первую очередь люди, такие же, как мы. Пиратская деятельность — это целый мир. Мир со своими законами и правилами. Он живёт собственной жизнью, его обитатели рождаются и умирают, сталкиваются с действительностью и продолжают заниматься привычным ремеслом.

Россия в этом смысле не исключение. Здесь компьютерное пиратство приобрело совершенно особые формы, хотя бы потому, что в торговлю нелегальным программным обеспечением вовлечены гигантские силы, и, что печально, деятельность эта

ПИРАТСКИЙ СЛОВАРЬ

Warez. Происходит от слова software. То самое, вокруг чего и идёт вся пиратская «жизнь». Так называют любое коммерческое, закрытое, недоступное программное обеспечение, которое было получено нелегальными способами и позднее взломано, изменено или дополнено каким-то образом.
Scene. Сцена. Это волшебное слово для всех людей, так или иначе связанных с компьютерным андеграундом. Существует warez-сцена, cracking-сцена, demo-сцена и т. д. Сцену можно охарактеризовать как некую взаимосвязь событий, людей, компьютеров и программ. Это и конкретное место, где происходит подобное взаимодействие, и вся та деятельность, в которую вовлечены люди этого круга. В общем, сцена — это сцена.
Crack. Программа, производящая изменения в другой программе с целью получения нужного эффекта.
Key, Keygen. Ключ. Это уникальная последовательность символов, позволяющая использовать некое программное обеспечение в полном объёме или просто использовать. Keygen — это программа для генерации ключей. Очень удобная вещь.
NFO. Файл с описанием продукта группы, составленный по определённым правилам.
Release. Релиз. Конечный продукт работы группы. Программное обеспечение, ранее закрытое, а теперь всем доступное.

подпадает под влияние криминальных структур всевозможных уровней. Печально потому, что люди, связанные с компьютерным андеграундом, — это целое сообщество со своей культурой и идеологией.

Как же появляются взломанные программы и программы с бесплатными ключами, за которые при нормальном положении вещей нужно платить порой немалые деньги? Благодетелями в этом случае являются пиратские группы, или группировки, как их называют непримиримые противники пиратов.

Группа — это некое неформальное сообщество людей, как правило молодых и талантливых. Кто же туда входит?

Organiser (пишется именно так, а не Organizer). Человек, с которого начинается группа. Он находит остальных, формирует цели и задачи группы. Некое подобие менеджера или начальника. Можно сказать, отец родной для членов группы, так как без него «в товарищах согласья нет» и работа не движется.

Cracker. Наверное, второй по важности человек в группе. Он владеет технологиями reverse engineering (дословный перевод — «обратное проектирование») и может изменять и снимать защиту в коммерческих программах.

Supplier. Член группы, который обеспечивает её материалом — свежими версиями программ, у которых можно снять защиту, новыми играми и т. д. Supplier может работать продавцом программного обеспечения или иметь контакты с производителями либо тестерами программ.

Courier. Курьер. Рабочая лошадка. Один из тех, кто отвечает за имидж группы на сцене. Он несёт ответственность за своевременное появление релиза на серверах для дальнейшего распространения.

Trader. Тот, кто ведёт обмен свежим «софтом» с другими группами или частными лицами.

Людей, которые осуществляют одни и те же функции, может быть несколько; в то же время и один человек может исполнять несколько обязанностей. Всё зависит от размера группы.

Взаимоотношения в группе, как правило, неформальные. Случается, что участники группы даже не знают друг друга лично, а общаются исключительно по электронной почте или IRC (Internet Relay Chat). Очень часто группа имеет свой собственный IRC-канал. Такой способ общения нисколько не ухудшает общей продуктивности работы.

Как рождается релиз? Организатор выбирает направление деятельности (релиз игр, мультимедиа-программ или, например, офисных приложений). Так, группа CLASS специализируется исключительно на игрушках, а RADIUM — на музыкальных программах. Supplier добывает свежую версию программы, cracker снимает защиту. Далее программа тестируется, упаковывается по неким правилам, в частности, необходимо, чтобы некоторые фрагменты были заархивированы для удобства загрузки из Интернета. Обязательно пишется NFO-файл. В нём порой присутствует логотип группы, выполненный в жанре ASCII Art (новый вид искусства, возникший в результате популяризации компьютера, — рисование на экране ЭВМ с помощью символов, предназначенных для вывода текста), а также имеются краткое описание продукта, псевдонимы создателей релиза, инструкция по установке, информация о группе и какие-нибудь новости. Очень часто релиз сопровождается небольшим роликом, выполненным в жанре demo — эффекты и музыка в реальном времени плюс некие сведения о группе, приветы и название релиза. После этого трейдеры и курьеры распространяют релиз на серверах для общего доступа.

Стоп! А где же выгода? Коммерческой прибыли здесь нет. Что же получает команда? Наверное, в первую очередь моральное удовлетворение. Имя на сцене, почёт. Снятие особо сложной защиты (точнее, её корректное снятие) даёт много очков группе. Это означает, что релизы группы будут чрезвычайно популярны и она получит хорошую репутацию.

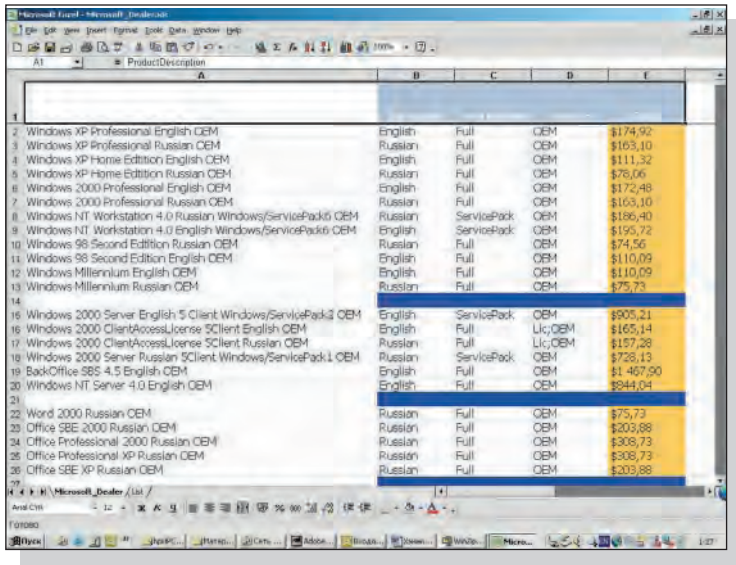
Какие программы являются предметом внимания для групп? Популярные. Именно популярные. Те, которые представляют хоть какую-то ценность, выражаемую не в деньгах, а в пригодности для работы.

Оценка, данная пиратами тому или иному продукту, дорогого стоит. Пираты, как правило, являются профессионалами в области компьютерных технологий, и их мнение — одно из самых авторитетных. Если программа попала в руки пиратов, это означает, что она уже представляет некую

ценность. Более того, действительно хорошая работа будет отмечена такой строчкой в NFO: If you like this programm, buy it! — «Если тебе нравится эта программа, купи её!». В истории компьютерного пиратства даже был беспрецедентный случай, когда группа LAXITY отказалась от взлома новых версий Bleem! (эмулятор Sony PlayStation), так как посчитала программу слишком хорошей, чтобы лишать автора возможности получения гонораров за неё. Это говорит само за себя. Лучшие должны получать прибыль. Аутсайдеры — думать, как стать лучше.

Почему пираты порой обращаются к пользователям с воззванием купить оригинал программы? Во-первых, считают, что автор заслуживает поддержки. А во-вторых, понимают, что если автор не получит причитающихся ему денег, то, возможно, полностью прекратит работу над новыми версиями. И что самое странное, всё это — и пиратство, и благородство — сосуществует и работает.

Чем западные пираты отличаются от российских? Наверное, мотивами. У первых сильны традиции — они ломают ради самого процесса; уровень жизни на Западе выше, и пользователь может позволить себе платить за нужное ему программное обеспечение.



	A	B	C	D	E
1	ProductDescription				
2	Windows XP Professional English OEM	English	Full	OEM	$174,92
3	Windows XP Professional Russian OEM	Russian	Full	OEM	$163,10
4	Windows XP Home Edition English OEM	English	Full	OEM	$111,32
5	Windows XP Home Edition Russian OEM	Russian	Full	OEM	$76,06
6	Windows 2000 Professional English OEM	English	Full	OEM	$172,48
7	Windows 2000 Professional Russian OEM	Russian	Full	OEM	$163,10
8	Windows NT Workstation 4.0 Russian Windows/ServicePack6 OEM	Russian	ServicePack	OEM	$186,40
9	Windows NT Workstation 4.0 English Windows/ServicePack6 OEM	English	ServicePack	OEM	$195,72
10	Windows 98 Second Edition Russian OEM	Russian	Full	OEM	$74,56
11	Windows 98 Second Edition English OEM	English	Full	OEM	$110,09
12	Windows Millennium English OEM	English	Full	OEM	$110,09
13	Windows Millennium Russian OEM	Russian	Full	OEM	$75,73
14					
16	Windows 2000 Server English 5 Client Windows/ServicePack3 OEM	English	ServicePack	OEM	$905,21
16	Windows 2000 ClientAccessLicense 5Client English OEM	English	Full	Lic;OEM	$165,14
17	Windows 2000 ClientAccessLicense 5Client Russian OEM	Russian	Full	Lic;OEM	$157,28
18	Windows 2000 Server Russian 5Client Windows/ServicePack1 OEM	Russian	ServicePack	OEM	$728,13
19	BackOffice SBS 4.5 English OEM	English	Full	OEM	$1 467,90
20	Windows NT Server 4.0 English OEM	English	Full	OEM	$844,04
21					
22	Word 2000 Russian OEM	Russian	Full	OEM	$75,73
23	Office SBE 2000 Russian OEM	Russian	Full	OEM	$203,88
24	Office Professional 2000 Russian OEM	Russian	Full	OEM	$308,73
25	Office Professional XP Russian OEM	Russian	Full	OEM	$308,73
26	Office SBE XP Russian OEM	Russian	Full	OEM	$203,88

Выдержка из прайс-листа ПО фирмы Microsoft (фотография с экрана).

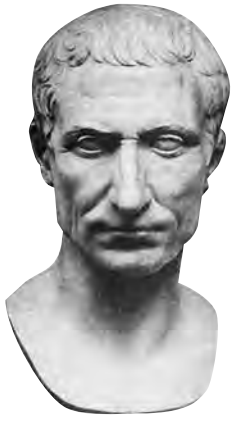
Российская сцена, точнее, её традиции молоды, и ломка осуществляется именно с целью последующего бесплатного использования. Но результаты те же — рост надёжности защиты ПО, бесплатная реклама хороших программ, выработка вкуса у массового пользователя.

Не надо смотреть на пиратство как на абсолютное зло. У медали две стороны. Крупным фирмам стоит реально оценивать российский рынок. Например, издатели игр уже делают это, продавая специальные, удешевлённые версии игр в России. Будем надеяться, что другие производители возьмут с них пример и повернутся лицом к конечным пользователям, а не к каким-то «крупным клиентам», фигурирующим в отчётах менеджеров по продажам.

ШИФРОВАНИЕ ИНФОРМАЦИИ

Поиски надёжных способов секретной передачи и хранения информации корнями уходят далеко в прошлое. Шифрование текстов — одна из самых любимых игр в истории человечества. Шифры использовались в военных целях, для передачи секретных сообщений между друзьями, для хранения тайного знания и в сотнях других случаев.

Первые зашифрованные сообщения использовались ещё в Древнем Египте. Способ шифрования был очень прост, сейчас он называется *шифрование простой подстановкой*. Каждый иероглиф исходного сообщения заменялся на другой иероглиф в зашифрованном сообщении. Соответствие иероглифов было взаимно однозначно, и, чтобы прочитать зашифрованное сообщение, требовалось выполнить обратную замену.



Гай Юлий Цезарь.

ШИФР ЦЕЗАРЯ

Примером наиболее простого шифра, относящегося к группе шифров простой подстановки, является шифр Цезаря. По свидетельству древнеримского историка Светония, Гай Юлий Цезарь (102 или 100—44 до н. э.) использовал его для тайной переписки. В шифре Цезаря каждая буква исходного сообщения сдвигается в алфавите на фиксированное число позиций вперёд, при необходимости переходя циклически на начало алфавита. Сам Цезарь использовал сдвиг на три позиции. В этом случае сообщение

ВОЗВРАЩАЙТЕСЬ В РИМ

шифруется так:

ЕСКЕУГЬГМХИФЯ Е УЛП

Здесь буква В шифруется буквой Е, отстоящей от буквы В на три позиции, буква О — буквой С и так далее (считается, что буквы Ё в алфавите нет). Последняя буква алфавита Я шифровалась бы при этом методе как В. Для расшифровки сообщения нужно сделать сдвиг на три позиции назад.

Шифр Цезаря определяется величиной сдвига. Поскольку число различных сдвигов на единицу меньше, чем число букв алфавита, разгадывание шифра Цезаря не представляет особого труда. Достаточно перебрать

всевозможные величины сдвига — от 1 до 31 в случае русского алфавита. Сообщение будет расшифровано, как только получится осмысленный текст.

ШИФРОВАНИЕ ПРОСТОЙ ПОДСТАНОВКОЙ

Более сложным является *метод простой подстановки*, при котором каждая буква исходного сообщения кодируется другим знаком, заданным таблицей кодировки.

Верхняя строка таблицы содержит все буквы алфавита. В нижней строке таблицы можно произвести любую перестановку букв алфавита; каждой такой перестановке соответствует определённый шифр.

Шифрование простой подстановкой описано в рассказе Артура Конан Дойла «Пляшущие человечки». Каждая буква сообщения изображалась определённой фигуркой пляшущего человечка. «Цель изобретателя этой системы заключалась, очевидно, в том, чтобы скрыть, что эти значки являются письменами, и выдать их за детские рисунки. Но всякий, кто сообразит, что значки эти соответствуют буквам, без особого труда разгадает их, если воспользуется обычными правилами разгадывания шифров». Какие же правила имел в виду Шерлок Холмс?

Изобретатели шифра «пляшущих человечков» — чикагские бандиты значительно облегчили задачу разгадывания шифра, снабдив флагами человечков, стоящих в начале и конце слов. Таким образом, сразу прояснялось разбиение зашифрованного сообщения на слова. А в любом языке всегда очень немного однобуквенных или двухбуквенных слов, и путём их перебора легко подобрать правильные буквы. Подставляя найденные буквы в другие слова, можно быстро отбросить неправильные варианты. Каждое угаданное слово проясняет соответствующие буквы во всём тексте и быстро приближает к разгадке всего шифра. Шерлок Холмс владел и дополнительной информацией, облегчившей задачу разгадывания шифра: он знал, что послания предназначены женщине по имени Илси, а письма в большинстве случаев начинаются с имени того, кому они адресованы. Таким образом, он сразу стал обладателем трёх букв. Затем Шерлок Холмс угадал слово «приходи», и дальнейшее не составило особого труда.

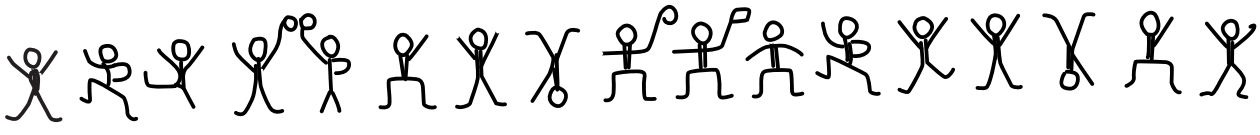
Кадр из кинофильма «Приключения Шерлока Холмса и доктора Ватсона».

ЧАСТОТНЫЙ АНАЛИЗ ТЕКСТА

Метод разгадывания шифра «пляшущих человечков», применённый Шерлоком Холмсом, не очень подходит для компьютера.

Имеется другой эффективный метод, основанный на знании вероятности, с которой каждая буква встречается в тексте. Он называется *методом частотного анализа*. Единственный недостаток этого метода — зашифрованное сообщение должно быть достаточно длинным, чтобы его можно было разгадать.

Проанализировав большой объём каких угодно текстов, можно для каждой буквы алфавита подсчитать, с какой примерно частотой встречается эта буква в любом тексте. Если учитывать пробелы между словами, то выясняется, что пробел встречается чаще всего с вероятностью примерно 0,167 (т. е. в тексте объёмом 1000 знаков имеется в среднем 167 пробелов). Наличие разделителей слов в зашифрованном тексте намного облегчает его расшифровку, поэтому при шифровании обычно исключают пробелы. Если не учитывать пробелы, то частотность букв русского алфавита примерно такова (приведённые данные являются результатом анализа нескольких художественных текстов



суммарным объёмом около 500 тыс. символов):

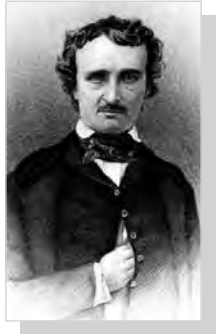
```
О 0,1163 А 0,0890 Е 0,0833 Н 0,0697
И 0,0599 Т 0,0585 Л 0,0532 С 0,0528
Р 0,0418 В 0,0402 К 0,0332 М 0,0311
Д 0,0302 П 0,0280 У 0,0279 Я 0,0218
Ь 0,0203 Ы 0,0185 Г 0,0184 Б 0,0173
З 0,0172 Ч 0,0143 Й 0,0121 Ж 0,0120
Х 0,0097 Ш 0,0087 Ю 0,0044 Щ 0,0031
Э 0,0029 Ц 0,0029 Ф 0,0014 Ъ 0,0002
```

Таким образом, в любом тексте на каждую тысячу букв приходится в среднем 116 букв О, 89 букв А, 83 буквы Е, 70 букв Н и т. д. Самыми редкими буквами являются буква Ф и Ъ: на 10 000 букв в среднем приходится всего лишь 14 букв Ф и всего 2 Ъ! (Однако если бы после Октябрьской революции не была проведена реформа правописания, то, скорее всего, Ъ был бы одной из наиболее часто встречающихся букв.)

Не менее важной характеристикой текста являются частоты, с которыми в тексте присутствуют различные двухбуквенные сочетания. Такие пары в криптографии называют *биграммами* или *диграммами* (не путать с диаграммами; «ди» — от *греч.* «двойной»). Вот список наиболее часто встречающихся в русских текстах биграмм в порядке убывания их вероятности:

```
ТО НА НЕ ПО НО ЛА СТ ОН
РА АЛ КО ГО КА ЛО НИ ОВ
```

На статистических характеристиках текста основан метод разгадывания любого шифра из группы подстановочных шифров. Подсчитав частоты всех букв и биграмм в зашифрованном сообщении, можно предположить, что наиболее частая в нём буква обозначает одну из наиболее часто встречающихся букв алфавита. Аналогично самая частая биграмма зашифрованного текста соответствует, скорее всего, одной из наиболее часто встречающихся биграмм в незашифрованных текстах. Конечно, помощь компьютера в статистическом анализе зашифрованного текста неоценима, но в принципе все эти подсчёты можно провести и без ЭВМ.



Эдгар Аллан По.

Знаменитый американский писатель Эдгар Аллан По также был увлечён криптографией. Сюжет его рассказа «Золотой жук» выстроен вокруг разгадывания секретного сообщения. Эдгар По верил, что раскрытие шифров и других загадок требует всего лишь применения логики и интеллекта. В 1893 г. в журнале «Alexander's Weekly Messenger» он бросил вызов читателям, утверждая, что разгадает любую криптограмму, использующую подстановочный шифр, которую те пришлют ему.

«Человеческая изобретательность не в силах состряпать шифр, который человеческая изобретательность не в силах была бы разгадать». Эти слова принадлежат Эдгару По. Однако современная криптография опровергла утверждение знаменитого писателя. Современные шифры, использующие кодирование с открытым ключом (см. статью «Современная криптография»), невозможно взломать даже с помощью самых мощных компьютеров.

ШИФР ВИЖЕНЕРА

Блезом де Виженером, придворным короля Франции Генриха III, в конце XVI в. был предложен весьма изящный метод шифрования. Иногда этот шифр называют также *шифром с перекрытием текста*. Для шифрования используются секретное слово или фраза. Нужно писать это секретное слово над исходным текстом, повторяя его, пока не кончится сообщение. Каждая буква исходного текста заменяется на отстоящую от неё в алфавите на несколько позиций. Величина сдвига задаётся буквой ключевого (секретного) слова, стоящей над данной буквой исходного текста. Для буквы А сдвиг вообще отсутствует, буква Б соответствует сдвигу на одну позицию вперёд, буква В — сдвигу на две позиции и т. д. Последняя буква — Я соответствует сдвигу на 31 позицию, поскольку в русском алфавите 32 буквы. То есть размер сдвига определяется порядковым номером буквы в алфавите, из которого вычтена единица.

В примере в качестве ключевого используется слово ХОЛМС. Пусть надо зашифровать сообщение

ПРИХОДИ НЕМЕДЛЕННО

Для этого пишется ключевое слово над шифруемой фразой:

ХОЛМСХО ЛМСХОЛМСХО
ПРИХОДИ НЕМЕДЛЕННО

Теперь каждую букву сообщения надо сдвинуть вперёд по алфавиту в соответствии с буквой ключевого слова, стоящей над ней. Например, буква Х является двадцать второй буквой алфавита и задаёт сдвиг на двадцать одну позицию вперёд. Вместо буквы П исходного текста получится буква Д зашифрованного сообщения. Вторая буква — Р исходного сообщения сдвигается в соответствии с буквой О ключевого слова на 14 позиций вперёд и заменяется на букву Ю. И так далее:

ДЮУБЯЩЦ ШСЭЪТЦСЮВЬ

РАЗГАДКА ШИФРА ВИЖЕНЕРА

На протяжении почти 300 лет шифр Виженера считался практически не взламываемым. Впервые метод разгадки шифра Виженера предложил в 1863 г. майор прусской армии Фридрих Касицкий. Его метод был основан на определении длины ключевого слова. Если нам известна длина ключевого слова, то можно разбить текст на несколько фрагментов, для каждого из которых применяется шифр Цезаря. В примере длина ключевого слова ХОЛМС равна пяти. Это означает, что первая, шестая, одиннадцатая, шестнадцатая и так далее буквы сообщения кодируются одним и тем же шифром Цезаря, соответствующим сдвигу вперёд на 21 позицию (поскольку первая буква ключевого слова — Х является 22-й буквой алфавита). Аналогично вторая, седьмая, двенадцатая, семнадцатая и так далее буквы кодируются шифром Цезаря со сдвигом на 14 позиций, соответствующим второй букве ключевого слова — О. Весь текст разбивается на пять фрагментов. К каждому из них можно применить метод частотного анализа, как в случае обычных подстановочных шифров. Таким образом, разгадывание шифра Виженера сводится на 90 % к нахождению длины ключевой фразы. Конечно, этот метод применим, только если текст достаточно большой, а ключевая фраза значительно короче его.

Метод Касицкого основывается на определении расстояний между повторяющимися биграммами в зашифрованном тексте. В случае когда в исходном тексте одна и та же биграмма повторяется на расстоянии, кратном длине ключевой фразы, она встретится на тех же позициях и в зашифрованном тексте. Подсчитав расстояние между повторяющимися биграммами и найдя все делители этого числа, можно получить набор чисел — кандидатов на длину ключевой фразы.

Пусть надо зашифровать четверостишие, используя ключевое слово БАРТО:

Наша Таня громко плачет:
Уронила в речку мячик.
Таня, Танечка, не плачь,
Не утонет в речке мяч!

Для начала надо переписать стихотворение, удалив пробелы, знаки препинания и построчное членение и разбив полученную последовательность букв на пятёрки для удобства подсчёта позиций:

НАШАТ АНЯГР ОМКОП ЛАЧЕТ УРОНИ
ЛАВРЕ ЧКУМЯ ЧИКТА НЯТАН ЕЧКАН
ЕПЛАЧ ЬНЕУТ ОНЕТВ РЕЧКЕ МЯЧ

Применив метод шифрования Виженера, получим:

ОАИТА БНПХЮ ПМЪАЭ МАЗЧА ФРЮЯЦ
МАТВУ ШКГЮН ШИЪДО ОЯВТЫ ЖЧЪТЫ
ЖПЫТЕ ЭНХЕА ПНХДР СЕЗЬУ НЯЗ

Зашифрованный текст содержит три повторяющиеся биграммы: МА в позициях 16 и 26, ТЫ в позициях 44 и 49 и НХ в позициях 57 и 62. Биграмма МА повторяется в зашифрованном тексте на расстоянии в десять позиций, биграммы ТЫ и НХ — на расстоянии в пять позиций. Поскольку десять кратно

Блез де Виженер.

Фридрих Касицкий.

пяти, то, скорее всего, длина ключевого слова равна пяти. Если длина 5 не подойдёт, то можно будет затем проверить и длину 10.

Определив 5 как наиболее вероятную длину ключевого слова, мы сделали самый значительный шаг к разгадке шифра Виженера.

Метод Касицкого требует некоторого везения. Биграмма в исходном тексте должна повториться на расстоянии, кратном длине ключевого слова. Если длина текста не очень большая, то этого может и не случиться. Кроме того, зашифрованный текст может содержать случайные повторяющиеся биграммы, расположенные на расстояниях, не кратных длине ключевого слова. Хотя их вероятность ниже, чем у «регулярных» биграмм, в случаях коротких текстов они могут помешать расшифровке. В 1920 г. американец Вильям Ф. Фридман (уроженец России) предложил другой, более простой и в то же время более продуктивный метод определения длины ключевой фразы. Этот метод является одним из самых ярких достижений классической криптографии.

Метод Фридмана основан на подсчёте так называемого «индекса совпадения». Зашифрованное сообщение переписывается с циклическим сдвигом текста на несколько позиций.

Вильям Ф. Фридман.

Далее полученные таким образом сообщения переписываются одно над другим и подсчитывается число совпавших букв в верхней и нижней строках. Вычисляется индекс совпадения, равный отношению числа совпадений к длине сообщения.

Для текстов на русском языке индекс совпадения составляет в среднем примерно 6 %. Но для абсолютно случайной последовательности русских букв индекс совпадения значительно ниже! Поскольку всего в русском алфавите 32 буквы, то вероятность совпадения составляет 1/32, или 0,031, — чуть больше 3%.

На этом соображении и основан метод Фридмана. Зашифрованный текст записывается со сдвигами 2, 3, 4 и так далее, и для каждого сдвига вычисляется индекс совпадения. Если индекс совпадения колеблется между 0 и 5 %, то, скорее всего, размер сдвига не соответствует длине ключевого слова. Длина ключевого слова найдена, как только индекс совпадения скачком возрастёт до 6 % и более.

В нашем примере зашифрованный текст выглядит следующим образом:

```
ОАИТАБНПХЮПМЪАЭМАЗЧАФРЮЯ
ЦМАТВУШКГЮНШИЪДООЯВТЫЖЧЪ
ТЫЖПЫТЕЭНХЕАПНХДРСЕЗЬУНЯЗ
```

Циклически сдвигая его и подсчитывая индекс совпадения, получим следующие результаты:

Сдвиг	Число совпадений	Индекс
2	0	0,000
3	5	0,068
4	2	0,027
5	8	0,110 — это длина ключа!
6	1	0,014
7	1	0,014
8	2	0,027

При сдвиге 5 индекс совпадения оказался значительно выше, чем при любых других сдвигах. Следовательно, длина ключевого слова, вероятнее всего, равна пяти.

Понять, почему индекс совпадения резко возрастает, когда величина сдвига становится кратной длине ключевого слова, совсем нетрудно. В случае

любого подстановочного шифра, в частности шифра Цезаря, индекс совпадения для зашифрованного текста такой же, как и для исходного текста. При сдвиге текста, зашифрованного методом Виженера, на длину ключевого слова мы сравниваем между собой буквы, которые шифруются одним и тем же шифром Цезаря. Следовательно, индекс совпадения получится в точности таким же, как и у незашифрованного текста при сдвиге на длину ключевого слова.

НЕМЕЦКАЯ ШИФРОВАЛЬНАЯ МАШИНА «ЭНИГМА»

В период Первой мировой войны и после неё широкое распространение получили механические и электромеханические шифровальные машины. С одной из них — немецкой шифровальной машиной «Энигма» связано наиболее крупное достижение практической криптографии.

«Энигма» использовалась в Германии с 1928 г., весь период после прихода к власти Гитлера и годы Второй мировой войны. В Шифровальном бюро Варшавы в 1932 г. начались работы над раскрытием тайны «Энигмы». Возглавлял группу молодой польский математик Мариан Решевский, выпускник математического факультета университета в Познани. Группа имела в своём распоряжении устаревшую коммерческую шифровальную машину, купленную в Германии. Конечно, эта модель была очень далека от современных для той поры немецких военных шифровальных машин и принесла мало пользы. Поэтому главным моментом в работе учёных для решения задачи «Энигмы» было применение математики (теории групп и теории перестановок).

В январе 1933 г. было теоретически воссоздано устройство машины, что позволило позже построить её реальную модель; были определены также методы восстановления ключей к шифрам на основе перехваченных сообщений.

Позднее, в 1939 г., материалы по «Энигме» были переданы во Францию и Англию. Англичанам удалось работы, раскрыть усовершенствования, которые были внесены в конструкцию последних немецких машин, и систему кодов, используемую Германией. В этой работе, выполнявшейся большой группой учёных в местечке Блетчли в 70 км от Лондона, участвовал знаменитый математик Алан Тьюринг, широко известный как автор воображаемой «машины Тьюринга». Работа сохранялась в глубокой тайне, и немцы даже не подозревали, что все их секретные сообщения становятся известны антигитлеровской коалиции. Многие историки, изучающие Вторую мировую войну, убеждены, что это значительно ускорило падение фашистской Германии и сохранило тысячи жизней. Информация об «Энигме» была закрытой и после окончания войны; она была опубликована только 30 лет спустя, по истечении срока давности военных секретов.

Мариан Решевский.

Воистину во второй половине XX в. произошла революция в криптографии. В современной схеме кодирования с открытым ключом ни кодировочная машина, ни ключ шифрования не являются секретом! Для шифрования сообщения вообще не нужно знать секретов (секретный ключ нужен только для дешифровки), и шифр практически нельзя взломать. Так что история, подобная раскрытию тайны «Энигмы», в наши дни вряд ли возможна.

СОВРЕМЕННАЯ КРИПТОГРАФИЯ

С появлением компьютеров криптография полностью изменилась. Классические методы шифрования не выдержали проверки компьютерами и теперь уже не используются. Одной из главных идей, лежащих в основе современных методов кодирования, является применение шифрования

УСТРОЙСТВО «ЭНИГМЫ»

«Энигма» относилась к классу электромеханических шифровальных машин. Её конструкция была основана на системе из трёх роторов, осуществлявших замену двадцати шести букв латинского алфавита. Каждый ротор имел 26 входных контактов на одной поверхности и 26 выходных контактов — на другой. Внутри каждого ротора проходили провода, связывавшие входные и выходные контакты между собой. Выходные контакты первого ротора соединялись с входными контактами второго ротора. Когда оператор нажимал на какую-либо букву на клавиатуре машины, электрический ток подавался на входной контакт первого ротора, соответствующий этой букве. Ток проходил через первый ротор и поступал на выходной контакт, соответствующий какой-либо другой букве. Затем ток проходил последовательно через второй и третий роторы и подавался на неподвижный рефлектор (от *лат.* reflecto — «обращаю назад», «отражаю»). В конструкции рефлектора 26 контактов разбивались на пары, контакты внутри каждой пары были соединены между собой. Таким образом рефлектор заменял букву на парную ей.

Шифровальная машина «Энигма».

Ток, прошедший через рефлектор, подавался назад, на систему роторов. Он вновь проходил через три ротора, но в обратном порядке. В конце концов на световом табло «Энигмы» загоралась одна из 26 лампочек, соответствовавшая зашифрованной букве.

Самым важным свойством «Энигмы» являлось вращение роторов. Первый ротор после каждой преобразованной буквы поворачивался на одну позицию. Второй ротор поворачивался на одну позицию после того, как первый ротор совершал полный оборот, т. е. после 26 преобразованных букв. Наконец, третий ротор поворачивался на одну позицию после того, как второй ротор совершал полный оборот, т. е. после $26 \cdot 26 = 676$ зашифрованных букв.

Благодаря рефлектору «Энигма» на каждом шаге осуществляла перестановку букв внутри пар, и если, к примеру, буква N заменялась на S, то при том же положении роторов буква S заменялась на N (ток шёл по тем же проводам, но в обратную сторону). Этим объяснялась особенность «Энигмы»: для расшифровки сообщения достаточно было вновь пропустить его через машину, восстановив предварительно начальное положение роторов.

Таким образом, начальное положение роторов играло роль ключа шифрования. Начальное положение роторов устанавливалось в соответствии с текущей датой. Каждый оператор имел специальную книгу, задававшую положение роторов для каждого дня. В этом заключалась очевидная слабость данной системы шифрования: достаточно было завладеть книгой и машиной, чтобы раскрыть все секреты.

не к буквам алфавита, как в классических шифрах, а к двоичным кодам. Это позволяет шифровать не только тексты, но и любые файлы (изображения, компьютерные программы и т. п.).

Каждый байт имеет 256 значений, что недостаточно для надёжного шифрования. Поэтому шифруются не отдельные байты, а блоки, состоящие из нескольких подряд идущих байтов. В большинстве схем шифрования каждый блок шифруется независимо от остальных. Чем больше размер блока, тем надёжнее шифр.

Все современные методы шифрования разделяются на две группы: *симметричное кодирование,* или кодирование с секретным ключом, и *асимметричное кодирование,* или кодирование с открытым ключом.

СХЕМА DES. КОДИРОВАНИЕ С СЕКРЕТНЫМ КЛЮЧОМ

В кодировании с секретным ключом один и тот же ключ используется как для шифрования, так и для дешифровки сообщения. Поэтому такое кодирование называют *симметричным.*

Кодируются блоки текста, состоящие из фиксированного числа идущих подряд байтов. Ключ также представляет собой несколько байтов. Например, в схеме DES (Data Encryption Standard — «стандарт шифрования данных») размер блока равен 8 байт, или 64 бит. Длина ключа также равна 8 байт. В них, правда, используются лишь 56 бит: старший бит каждого байта служит для проверки чётности и в шифровании не участвует.

Шифрование состоит в преобразовании входного слова из 64 двоичных разрядов в зашифрованное выходное слово той же длины. Полное описание схемы DES достаточно сложно и занимает несколько десятков страниц. Вот её основные идеи:

1. Нередко в разных областях программирования, и в частности в шифровании используется побитовая операция «исключающее ИЛИ» (общепринятое обозначение — XOR). Математики называют её сложением по

модулю 2. Операция задаётся следующим образом:

$$0 \oplus 0 = 0, \quad 0 \oplus 1 = 1,$$
$$1 \oplus 0 = 1, \quad 1 \oplus 1 = 0.$$

Символом $\oplus$ обозначается сама операция XOR. Когда её применяют к двум двоичным словам, она выполняется по отдельности для каждой пары битов в соответствующих позициях. Операция XOR обладает замечательным свойством обратимости:

$$(X \oplus Y) \oplus Y = X.$$

Прибавляя Y к X, шифруют сообщение X, а для дешифровки достаточно ещё раз выполнить ту же самую операцию.

2. Преобразование в схеме DES разбивается на элементарные шаги, которые называют раундами. Таких раундов 16. В каждом раунде используется свой 48-битный ключ K, составленный определённым образом из битов исходного ключа. Исходное 64-битное слово разбивается на 32-битные половины L (левую) и R (правую). К ним применяется следующее преобразование Фейнстеля, названное так в честь Хорста Фейнстеля, разработавшего для фирмы IBM один из первых методов симметричного кодирования «Люцифер»:

$$L' = R;$$
$$R' = L \oplus f(R, K),$$

где L' и R' — новые значения левой и правой половин слова.

Ключевая идея Фейнстеля состоит в том, что такое преобразование обратимо независимо от вида функции f:

$$R' \oplus f(R, K) = L;$$

и преобразование

$$R = L';$$
$$L = R' \oplus f(L', K)$$

является обратным к преобразованию Фейнстеля, поэтому оно используется при дешифровке.

3. Наиболее важным моментом в схеме DES является построение функции f, которая в каждом раунде применяется к ключу K данного раунда и к правой половине шифруемого слова. Построение функции f основывается на идеях Клода Шеннона, изложенных в 1949 г. в журнале «Bell System Technical Journal». От функции f требуется, чтобы её аргумент и результат были как можно меньше связаны друг с другом. Небольшое изменение аргумента должно приводить к существенному изменению результата, которое будет выглядеть абсолютно случайным.

Клод Шеннон предложил для построения подобной функции несколько раз последовательно применять преобразования типа «перемешивание» и «запутывание» (diffusing и confusing). Конкретное воплощение этой идеи было предложено в конце 60-х гг. XX в. Хорстом Фейнстелем. Перемешивание состоит в простой перестановке битов слова. Для запутывания применяются так называемые S-ящики (*англ.* substitution — «подстановка»). Длинное двоичное слово разбивается на группы по четыре бита. Значение этих четырёх битов рассматривается как одно двоичное число в диапазоне от 0 до 15, и оно подаётся на вход S-ящика. S-ящик заменяет это число на другое число, которое выдаётся на выходе S-ящика.

Обратно оно представляется также в виде четырёх битов. Таким образом, каждый *S*-ящик осуществляет фиксированное отображение четырёхбитных слов. Оно должно быть по возможности близким к случайному. Сложность дешифровки в значительной степени основывается на сложности отображений, осуществляемых *S*-ящиками.

В схеме DES используют несколько *S*-ящиков, выбранных на базе ключа *K* текущего раунда и битов правой половины слова (*R*). Преобразования, сходные с преобразованием Фейнстеля, применяются практически во всех современных схемах симметричного кодирования, а не только в DES.

Схема DES успешно применялась на практике в течение последних 20 лет и зарекомендовала себя как достаточно надёжная. До сих пор не найдено никаких существенно лучших методов её взламывания, чем полный перебор всех ключей. Вместе с тем ключ в схеме DES достаточно короткий, всего 56 бит, и перебрать все ключи вполне возможно. На современном персональном компьютере для этого понадобится порядка 1000 лет, что в теории алгоритмов вовсе не считается серьёзным временем! Специально же сконструированный многопроцессорный комплекс взламывает DES за несколько часов. Поэтому период широкого использования схемы DES подходит к концу. В настоящее время используется её модификация 3DES, которая состоит в трёхкратном применении алгоритма DES. В схеме 3DES ключ в два раза длиннее, поэтому подобрать его путём полного перебора нереально. К тому же активно ведётся поиск нового стандарта симметричного шифрования, который сможет полностью заменить DES.

КОДИРОВАНИЕ С ОТКРЫТЫМ КЛЮЧОМ

Подлинной революцией в криптографии явилась разработка схем асимметричного шифрования, или шифрования с открытым ключом. Если современные алгоритмы симметричного кодирования в той или иной мере напоминают классические шифры, и особенно роторные шифровальные машины начала XX в., то никаких прототипов схем кодирования с открытым ключом в истории не существовало.

Идея кодирования с открытым ключом очень проста: для шифрования и для дешифровки сообщения используются два разных ключа. Ключ, с помощью которого сообщение шифруется, называется открытым (public key). Ключ для дешифровки называется секретным или приватным (private key). Причём, зная открытый ключ, нельзя с его помощью восстановить секретный ключ.

Представим дверной замок с двумя разными ключами. Один служит только для запирания двери, другим её можно лишь отпереть. Похожим образом устроена схема шифрования с открытым ключом. Открытый ключ известен всем, и любой может зашифровать сообщение («запереть дверь»). Но только обладатель секретного ключа способен дешифровать сообщение («отпереть дверь»).

В одном из фильмов про Джеймса Бонда, «Из России с любовью», основной целью «агента 007» является захват секретной советской шифровальной машины «Лектор». Подобный сюжет в настоящее время был бы невозможен: шифровальная машина,

◄
Джеймс Бонд с машиной «Лектор».

т. е. компьютер плюс алгоритм шифрования, не представляет никакого секрета. Открытый ключ шифрования можно опубликовать даже в телефонной книге. Ценность для английской контрразведки мог бы представлять лишь секретный ключ дешифровки, но агент, передающий зашифрованные сообщения в Москву, им не обладает (секретный ключ находится в Москве).

Чрезвычайно важно дополнительное свойство наиболее распространённой в настоящее время схемы кодирования с открытым ключом RSA: процедуры шифрования и дешифровки взаимно обратны. Это означает, что схему кодирования с открытым ключом можно применять в обратном направлении: шифровать сообщение с помощью секретного ключа, а дешифровать — с помощью открытого. Таким образом можно подтвердить личность отправителя сообщения, ведь послать его может только обладатель секретного ключа; зато прочитать сообщение сумеет кто угодно, применив для дешифровки открытый ключ. На этом свойстве основано построение так называемой электронной подписи сообщения, удостоверяющей личность отправителя.

СХЕМА ШИФРОВАНИЯ С ОТКРЫТЫМ КЛЮЧОМ RSA

Схема шифрования с открытым ключом RSA, опубликованная в 1977 г., названа в честь её создателей — Роналда Ривеста, Ади Шамира и Леонарда Адлемана. Она была разработана в научных кругах и, в отличие от многих других алгоритмов шифрования, никогда не имела статуса государственного секрета. Основана она на простых и красивых результатах элементарной теории чисел, полученных ещё древними греками (расширенный алгоритм Евклида), китайцами (китайская теорема об остатках) и европейскими математиками XVII—XVIII вв. (малая теорема Ферма, теорема Эйлера).

Первая идея схемы RSA — это рассмотрение блоков текста, состоящих из нескольких последовательных байтов, как элементов кольца вычетов по модулю m. Объединяя двоичные записи всех байтов блока, получают длинное двоичное слово, которое трактуется как двоичная запись целого числа. Это число, в свою очередь, рассматривается как элемент кольца вычетов по модулю m. Таким образом, блоки текста отождествляются с элементами кольца $\mathbf{Z}_m$. Шифрование состоит в преобразовании элементов $\mathbf{Z}_m$.

Число m в схеме RSA достаточно большое, не менее 200 десятичных знаков. Оно равняется произведению двух больших простых чисел:

$$m = pq.$$

Число m является открытым, но его представление в виде произведения двух простых чисел сохраняется в тайне. Зная разложение m на множители, можно вычислить функцию Эйлера:

$$\phi(m) = (p-1)(q-1).$$

Далее выбирается произвольное натуральное число e, взаимно простое

Пьер Ферма.

Роналд Ривест, Ади Шамир и Леонард Адлеман.

с $\phi(m)$. Для него с помощью расширенного алгоритма Евклида вычисляется обратный элемент d в кольце вычетов по модулю $\phi(m)$:

$$de \equiv 1 \pmod{\phi(m)}.$$

Открытым ключом является пара чисел (m, e). Секретным ключом является пара чисел (m, d).

Процедура шифрования E состоит в возведении исходного числа s в степень e в кольце вычетов по модулю m:

$$E(s) \equiv s^e \pmod{m}.$$

Здесь через s обозначен блок исходного сообщения, который рассматривается как элемент кольца $\mathbf{Z}_m$.

Процедура дешифровки D состоит в возведении в степень d в кольце вычетов по модулю m:

$$D(t) \equiv t^d \pmod{m}.$$

Здесь через t обозначен блок зашифрованного сообщения.

E и D — две взаимно обратные процедуры.

Применим к зашифрованному сообщению

$$t \equiv s^e \pmod{m}$$

процедуру дешифровки:

$$D(t) \equiv t^d \equiv (s^e)^d \equiv s^{ed} \pmod{m}.$$

Поскольку числа e и d по построению являются взаимно обратными элементами кольца вычетов по модулю $\phi(m)$, то

$$ed \equiv 1 \pmod{\phi(\mathrm{m})},$$

т. е. разность $ed-1$ делится на $\phi(m)$.

Для некоторого натурального числа k

$$ed - 1 = k\,\phi(m),$$
$$ed = k\,\phi(m) + 1.$$

Следовательно,

$$s^{ed} = s^{k\phi(m)+1} \equiv s \pmod{m}.$$

Последнее равенство справедливо в силу теоремы Эйлера. Таким образом, применив к исходному сообщению s сначала шифрующую, а затем дешифрующую процедуры, снова получено исходное сообщение s.

Итак, как шифрование, так и дешифровка состоят просто в возведении в степень в кольце вычетов. Числа, которые участвуют в шифровании, должны быть достаточно большими — не менее 200 десятичных знаков.

Вот для иллюстрации пример с маленькими числами.

Пусть $m = 11 \cdot 13 = 143$. Функция Эйлера: $\phi(m) = 10 \cdot 12 = 120$.

Выбрав $e = 113$, получим $d = 17$ — обратный к e элемент в кольце $\mathbf{Z}_{120}$:

$$113 \cdot 17 = 1921 = 120 \cdot 16 + 1 \equiv$$
$$\equiv 1 \pmod{120}.$$

Пара (143, 113) составляет открытый ключ, пара (143, 17) — секретный ключ. Шифрующая процедура E состоит в возведении в степень 113 по

КОЛЬЦО ВЫЧЕТОВ ПО МОДУЛЮ *m*

Основным объектом в элементарной теории чисел является кольцо вычетов по модулю m, которое обозначается $\mathbf{Z}_m$. Кольцом в алгебре называется множество, элементы которого можно складывать и умножать. Например, кольцом является множество целых чисел $\mathbf{Z}$. Пусть m — некоторое натуральное число. Кольцо $\mathbf{Z}_m$ строится следующим образом: всё множество целых чисел $\mathbf{Z}$ разбивается на m подмножеств, каждое из них составляют числа, разность которых кратна m. Нулевое подмножество состоит из чисел $\{\dots -m, 0, m, 2m, \dots\}$, первое подмножество — из чисел $\{\dots -m+1, 1, m+1, 2m+1, \dots\}$, m–1-ое — из чисел $\{\dots -1, m-1, 2m-1, 3m-1, \dots\}$. Эти подмножества в математике называются классами. Указанные классы являются элементами кольца $\mathbf{Z}_m$, всего их m штук. Операции сложения и умножения классов проводятся с их представителями — по одному от каждого класса. И тот класс, в который попадает сумма (произведение), по определению является суммой (произведением) двух исходных классов.

Тот факт, что два числа — x и y принадлежат одному и тому же классу (т. е. их разность делится на m), записывается в математике следующим образом: $x \equiv y \pmod m$. Читается: «x сравнимо с y по модулю m».

Система представителей каждого класса называется системой остатков. Математики чаще всего рассматривают неотрицательную систему остатков $\{0, 1, 2, \dots, m-1\}$, но в программировании более удобна так называемая симметричная система, состоящая из отрицательных и неотрицательных чисел, минимальных по абсолютной величине. Так, для кольца $\mathbf{Z}_5$ симметричная система остатков состоит из чисел $\{-2, -1, 0, 1, 2\}$. Компьютер работает не с целыми числами, а с элементами кольца вычетов по модулю $m = 2^{32}$. Те числа, которые программисты называют положительными и отрицательными, являются на самом деле элементами симметричной системы остатков этого кольца.

Важность колец $\mathbf{Z}_m$ на практике вытекает из того, что эти кольца содержат конечное число элементов и при вычислениях в них числа не растут до бесконечности. Вычисления с «настоящими» целыми числами проводятся крайне редко. Объясняется это тем, что при умножении целых чисел результат очень быстро растёт.

Например, произведение двух 100-значных десятичных чисел состоит уже из 199 или 200 цифр, возвести же 100-значное число в 100-значную степень вообще нереально: результат превышает даже количество элементарных частиц во Вселенной!

модулю 143, дешифрующая процедура D — в степень 17 по модулю 143. Зашифруем произвольное сообщение, например, $s = 123$:

$$E(123) \equiv 123^{113} \pmod{143} \equiv 41.$$

Тогда 41 — это закодированное сообщение. Применив к нему декодирующую процедуру, получим:

$$D(41) \equiv 41^{17} \pmod{143} \equiv 123,$$

т. е. исходное сообщение.

Насколько сложна задача восстановления секретного ключа по открытому ключу? Открытый ключ представляет собой пару чисел (m, e), секретный — пару (m, d), где d — обратный элемент в кольце вычетов по модулю (m). Следовательно, для вычисления d нужно знать функцию Эйлера $\phi(m)$. Задача вычисления функции Эйлера эквивалентна задаче о разложении числа m на множители. То есть для определения открытого ключа надо разложить на множители число m.

Как оказалось, задача разложения числа на множители очень сложна. Она привлекала внимание многих математиков, начиная с Эйлера, который разложил на множители пятое число Ферма

$$F_5 = 2^{2^5} + 1 = 4\,294\,967\,297,$$

до этого ошибочно считавшееся простым. Эйлер использовал изящную идею — найти два квадрата, совпадающих по модулю m, которая и по сей день лежит в основе многих современных алгоритмов разложения. Тем не менее у всех известных алгоритмов время работы очень быстро растёт с увеличением длины записи числа m, что ставит для всех них естественный предел применимости. В настоящее время он составляет примерно 100 десятичных знаков. При этом рост быстродействия компьютеров почти ничего не даёт. Грубо говоря, если компьютеры станут в миллион раз быстрее, будет возможно раскладывать 120-значные числа вместо 100-значных, но всё равно машине никогда

Необходимо ли для взламывания открытого ключа разложить на множители число m, до сих пор не доказано (не известно, эквивалентно ли взламывание RSA задаче о разложении числа на простые множители).

АЛГОРИТМ ЕВКЛИДА

Одним из краеугольных камней элементарной теории чисел является алгоритм Евклида: даны два целых числа n и m, нужно найти их наибольший общий делитель $d = \text{НОД}(m, n)$. Для этого пара чисел (n, m) последовательно заменяется на пару (m, r), где r — остаток от деления n на m, т. е. $n = qm + r$. Поскольку общие делители пар (n, m) и (m, r) совпадают, то $\text{НОД}(n, m) = \text{НОД}(m, r)$. С другой стороны, $\text{НОД}(m, 0) = m$. Поскольку при замене (n, m) на (m, r) второе число уменьшается, то через конечное число шагов оно станет равным нулю, при этом первое число пары будет равно НОД исходной пары чисел.

Алгоритм Евклида очень быстрый, число шагов в нём примерно равно количеству битов в двоичной записи числа, т. е. для 100-значных десятичных чисел алгоритм Евклида завершается примерно через 330 шагов. На практике чрезвычайно важен расширенный алгоритм Евклида, позволяющий выразить $d = \text{НОД}(n, m)$ в виде линейной комбинации чисел n и m:

$$d = un + vm,$$

где u и v — некоторые целые числа. Расширенный алгоритм Евклида позволяет вычислить обратный к n элемент в кольце вычетов по модулю m. Элемент u называется обратным к n, если их произведение равно единице в кольце $\mathbf{Z}m$. Например, в кольце $\mathbf{Z}_5$

$$2 \cdot 3 = 6 \equiv 1 \pmod 5,$$

следовательно, число 2 является обратным к числу 3 в кольце $\mathbf{Z}_5$. Число n обратимо в кольце вычетов по модулю m тогда и только тогда, когда n взаимно просто с m, т. е. $\text{НОД}(n, m) = 1$. Применив расширенный алгоритм Евклида, можно найти такие числа u и v, что

$$1 = un + vm.$$

Но $vm \equiv 0 \pmod m$, следовательно,

$$1 \equiv un \pmod m,$$

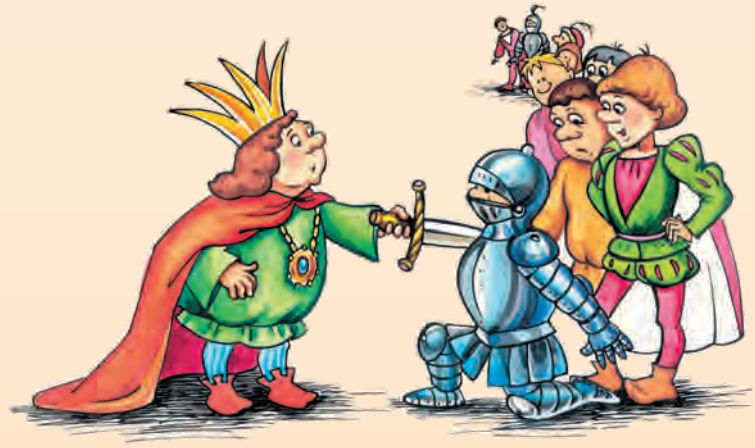
т. е. элемент u является обратным к элементу n в кольце вычетов $\mathbf{Z}_m$.

БЫСТРОЕ ВОЗВЕДЕНИЕ В СТЕПЕНЬ

Ещё один алгоритм, который постоянно используется на практике, — это алгоритм быстрого возведения в степень. Как правило, алгоритм этот применяется не для целых чисел, а для элементов колец вычетов (целые числа невозможно возводить в большие степени, поскольку они при этом очень быстро растут). Идея этого алгоритма состоит в том, что мы сводим возведение в степень к операциям возведения в квадрат и умножения. Эти операции выполняются в кольце $\mathbf{Z}_m$. Пусть надо возвести элемент s в степень e. Используется следующий алгоритм:

```
алг цел быстрое возведение в
| степень (цел s, e)
| дано: число s, показатель
| степени e
| надо: возвести s в степень e
| в кольце вычетов
| нач цел p
| p:=1
| цикл пока e > 0
| | если e/2*2=e | e-чётное
| | | то e:=e/2; s:=s*s
| | | иначе e:=e-1; p:=p*s
| | всё
| кц
| знач:=p
кон
```

Справедливость алгоритма вытекает из того, что величина ps^e остаётся неизменной после каждого выполнения тела цикла. Вначале она равна s^e, поскольку $p = 1$. По окончании алгоритма $e = 0$, следовательно, $ps^e = p$, т. е. в p содержится искомая степень первоначального значения s. Алгоритм также работает очень быстро, количество шагов не превосходит удвоенного числа битов двоичной записи числа e. В частности, если e — 100-значное десятичное число, то число шагов алгоритма примерно равно 600.



МАЛАЯ ТЕОРЕМА ФЕРМА И ТЕОРЕМА ЭЙЛЕРА

Подлинной жемчужиной теории чисел является малая теорема Ферма.

Пусть m — простое число (т. е. число, делящееся без остатка только на себя и на единицу). Тогда для всякого целого числа b, не делящегося на m, справедливо сравнение

$$b^{m-1} \equiv 1 \pmod{m}.$$

Например, пусть $m=7$, $b=2$, тогда

$$2^6 = 64 = 9 \cdot 7 + 1 \equiv 1 \pmod{7}.$$

Существует также другая форма малой теоремы Ферма. Пусть m — простое число. Тогда для всякого целого числа b справедливо тождество

$$b^m \equiv b \pmod{m}.$$

Леонард Эйлер в XVIII в. обобщил малую теорему Ферма для случая произвольного числа m. По определению, функцией Эйлера (m) называют количество обратимых элементов кольца Z_m.

Теорема Эйлера: пусть m — произвольное натуральное число, отличное от единицы, b — целое число, взаимно простое с m, тогда

$$b^{\phi(m)} \equiv 1 \pmod{m}.$$

Функция Эйлера легко вычисляется с помощью другой классической теоремы — китайской теоремы об остатках, которая здесь не приводится. В схеме RSA задействован частный случай теоремы Эйлера для $m = pq$, где p и q — два разных простых числа. В этом случае

$$\phi(m) = (p-1)(q-1).$$

Существует следующее усиление теоремы Эйлера: пусть $m = pq$, где p и q — разные простые числа. Тогда для любого целого числа b и для любого натурального k справедливо тождество

$$b^{k\phi(m)+1} \equiv b \pmod{m}.$$

В этом случае уже не требуется, чтобы число b было взаимно просто с числом m.

не разложить 400-значное число. Многие математики убеждены, что для задачи разложения чисел не существует быстрых алгоритмов, хотя строго это не доказано.

Для многочленов, в отличие от чисел, имеется быстрый алгоритм разложения на множители. Это только подчёркивает, насколько сложными объектами являются целые числа. В отличие от разложения на множители задачи проверки простоты числа и генерации больших простых чисел имеют красивые алгоритмические решения, найденные в 70-х гг. XX в. Это позволяет генерировать сколь угодно длинные ключи в схеме RSA.

Шифрование и дешифровка в схеме RSA происходят всё же значительно медленнее, чем те же операции с использованием алгоритмов симметричного шифрования. Поэтому шифрование с открытым ключом обычно применяют лишь в наиболее важных случаях. Прежде всего это процедура входа в сеть и идентификации пользователя, в которой схема кодирования с открытым ключом заменяет примитивную передачу пароля. Кроме этой возможности удостоверения пользователя и электронной подписи имеется способ обмена секретными ключами по открытому каналу связи. Обычно после входа пользователя в сеть система выделяет ему одноразовый ключ для симметричного шифрования (3DES и т. п.), и вся передающаяся по сети информация шифруется с помощью симметричного ключа.

При использовании кодирования с открытым ключом можно не опасаться хакеров, ведь для определения ключа надо решить чисто математическую задачу разложения числа на множители, на что никакой хакер не способен.

Леонард Эйлер.

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

ИНФОРМАЦИОННОЕ И ПОСТИНФОРМАЦИОННОЕ ОБЩЕСТВО

Человечество ныне переживает переход от одной эпохи к другой. Индустриальное общество сменяется информационным, когда доступ к информации упрощается, а компьютеры используются в качестве универсальных приборов для её хранения, переработки и передачи.

Всегда интересно заглянуть за поворот и узнать, что ждёт впереди, что придёт на смену информационному обществу?

Существует мнение, что эпоха информационного общества уже завершена, практически не начавшись. Бурное развитие средств коммуникации и новейших технологий создали ситуацию, при которой новая информация теряет актуальность быстрее, чем её удаётся применить. Руководитель разведывательной службы одного из высокоразвитых государств отметил, что если недавно основная проблема состояла в добывании данных, то сейчас на первом месте находится их переработка. Конец информационного общества грозит обвалом мировой финансовой системы, так как деньги делает тот, кто знает, откуда они берутся, т. е. владеет информацией. Другой фактор, предрекающий глобальный хаос, — кризис управленческой технологии. Для действий по традиционной схеме «исследование — анализ результатов — синтез управленческих решений» может просто не хватить времени. Что приведёт к погрешностям и ошибкам в решениях глобального характера. В результате потери контроля над ситуацией восторжествует всемирная супертирания с

угнетением и уничтожением значительной части человечества. Невесёлая перспектива!

В качестве аргументации против такого взгляда можно сказать, что человечество пока ещё в состоянии «переварить» нарастающие объёмы информации, активно используя изобретённые компьютеры. И глобальные управленческие решения, судя по печальным событиям в Ираке и Киргизии, 2003—2004 г., по-прежнему принимаются людьми. А при наличии большего объёма достоверной информации решения скорее будут более правильными, чем при отсутствии данных.

Другая точка зрения состоит в том, что сейчас наблюдается не избыток, а дефицит информации. Несмотря на лавинообразное увеличение её объёма и скорости информационных потоков, в конечном счёте вся она доставляется человеку по двум каналам — слуховому и зрительному. В информационном обществе эти каналы чрезмерно загружены. Будущее постинформационное общество предполагает не отказ от информации, а перенос акцентов на другие каналы. По крайней мере, органов чувств у человека больше двух. (Правда, возникает вопрос: почему тогда это общество именуют постинформационным, а не мегаинформационным?)

Ключевой фигурой в ряду профессионалов завтрашнего дня станет интеллектуал — человек, производящий информацию, а само производство информации займёт важнейшее место в индустрии. Мир глобализируется: у него будут единые финансы, общие информационное поле, мораль, стиль жизни и т. п. Вместе с тем каждая страна продолжит жить сама по себе, сохраняя свою самобытность. Данный парадокс нового тысячелетия и приведёт к формированию постинформационного общества, когда внутренние импульсы и все созданные технологии смогут способствовать адаптации человека к текущей ситуации.

В целом большинство размышлений о постинформационном обществе (за исключением тех немногих, в которых предрекается конец света) сводится к тому, что прогнозируются некоторые, часто несущественные изменения в рамках информационного общества. Сам переход невозможно точно зафиксировать, так же как нельзя обозначить чёткую границу между юностью и зрелостью. В 15 лет человек ещё юн, а 40 — это, вне всякого сомнения, зрелый возраст. Так и общество: в XV в. оно не было индустриальным, а в XIX в. — информационным. Информационное общество — зрелость человечества. По аналогии с человеком ему предсказывают упадок и смерть или долгую старость, отягощённую болезнями. Но высказывают и оптимистическое предположение: зрелость не сменится увяданием, т. е. информационное общество не состарится никогда. Как, собственно, и заканчиваются сказки: «И они жили долго и счастливо и умерли в один день».

Трагедию, которая произошла с Останкинской телебашней летом 2001 г., и следующие дни можно рассматривать как один из возможных путей в постинформационное общество. Когда погасли голубые экраны телевизоров у миллионов москвичей, многие действительно ощутили так называемый информационный голод. Правда, это был скорее «телевизионный» голод, так как радиостанции с успехом компенсировали нехватку новостного потока.

ПРОИЗВОДИТЕЛИ МАССОВОЙ ИНФОРМАЦИИ

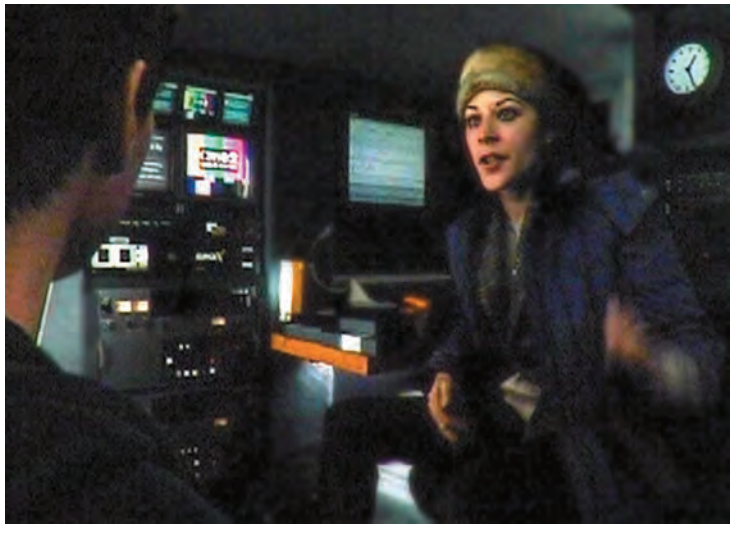
Часто спорят, средства массовой информации только сообщают новости и проводят расследования или сами производят ту самую информацию, которая до неузнаваемости меняет окружающий мир. Недаром их не жалуют и боятся — ведь это четвёртая власть.

Что такое *средства массовой информации?* Это прежде всего издания, которые поставляют нам сведения об окружающем мире. Главное в них — то новое, что произошло с миром, обществом, конкретными людьми, известными персонажами, культурой, наукой за небольшой отрезок времени: неделю, день, час, минуты. А значит, главное для СМИ настоящего и будущего — это оперативная передача информации тем, кто в ней нуждается.

С каждым годом система и технологии СМИ меняются всё быстрее и быстрее. Мир становится единым и всё более открытым, потому все эти обстоятельства будут влиять на развитие *информационного поля*. Никакие, даже самые смелые, фантазии не могут предугадать будущего.

Средства массовой информации условно делятся на *традиционные* и *синтетические*.

К традиционной группе можно отнести:



В блокбастере 2005 г. Стивена Спилберга «Война миров» есть эпизод, в котором, находясь в окружении инопланетян в условиях, когда большинство СМИ уничтожены, оставшись чудом в живых, корреспондент одного из телеканалов сетует, что не получилось сенсационного репортажа. Поистине новость дороже жизни.

• офлайновые, или «бумажные», — газеты, журналы и прочие печатные издания;

• электронные — телевидение и радио;

• Интернет-издания.

Синтетические СМИ, как это следует из названия, соединяют все эти технологии воедино.

Газеты и журналы мало меняются со временем — они будут существовать, как и прежде, поскольку читать текст, напечатанный на газетной бумаге, удобно и приятно. Это станет даже восприниматься как элемент престижа или консерватизма для серьёзных бизнесменов или молодых людей, которые хотят выглядеть респектабельно. Кроме того, у газеты и журнала есть ещё одно неоспоримое преимущество: они дожидаются своего читателя. Человеку не надо в спешном порядке усваивать какую-то информацию, комментарий к событиям одномоментно, здесь и сейчас, когда у него нет времени. Он может узнать и подумать об этом в другой день, на досуге, в выходной. Он имеет возможность развернуть газетную полосу и, читая, отгородиться от мира, оставшись наедине с газетой и чашкой кофе в руке.

Другое дело, что появятся дайджесты — издания с подборками публикаций на конкретную тему. Одно и то же событие с разных точек зрения будет подано журналистами разных изданий. Например, про создание новой ракеты напишут авторы научного издания, литературного альманаха, разнополярных политических изданий и молодёжной газеты. Таким образом, читатель сможет получить самые разносторонние оценки одного и того же явления. Поменяется тематика изданий: если сейчас в основном российские газеты пишут о политике и здоровье, то в будущем основными темами станут наука и культура. Изменится карта планеты. Мир сделается единым, политика перестанет интересовать всё общество, но проблема освоения недр земли, развитие высоких технологий, знания в самых акту-

альных областях, культура, мораль — те вопросы, которые всегда волнуют общество. Конечно, газета не успевает за сменой новостей, но эти вопросы не требуют постоянного обновления информации на её полосах.

Если судьба газет в большей или меньшей степени оптимистична, то это вряд ли можно сказать про так называемые толстые литературные журналы. Они сократятся, поскольку новые романы, повести и рассказы будут сразу издаваться в виде книг и параллельно продаваться в Интернете. И два этих формата окончательно погубят традицию семейного чтения толстых журналов.

Глянцевые журналы с картинками, где текста меньше, чем фотографий, тоже останутся на своих местах, потому что после работы (например, с компьютером) у любого человека будет внутренняя потребность посмотреть на красивые иллюстрации и полистать страницы с лёгким, ненавязчивым текстом.

Отношение к радио и телевидению изменится, поскольку мало кто захочет щёлкать кнопками пульта, переключая каналы. Потребитель информации станет её непосредственным заказчиком. Он начнёт лично программировать свой радио- и теледень. Например, будет применяться технология, по которой можно заказать себе экономические новости одной радиостанции, новости спорта другой, новости культуры третьей и т. д.

Музыку также будут подбирать на свой вкус: приёмник станет по определённой программе «выдёргивать» музыкальные композиции с волн других радиостанций. Если слушатель захочет принять участие в заинтересовавшем его ток-шоу, то ему достаточно голосом дать соответствующую команду, и приёмник, настроившись на определённую частоту, сделает его участником этой передачи.

Кроме того, можно настроить свой приёмник так, чтобы он принимал только позитивную (или отрицательную) информацию, передавал голоса ведущих только с определённым тембром. Естественно, что в этой ситуации усилятся «боевые действия» между радиостанциями и слушателями. Люди начнут настаивать на своём праве выбора, а радиостанции всеми правдами и неправдами — бороться за клиентов. Поэтому «зашедший на волну» будет фиксироваться и «приклеиваться» к ней. Специальные электронные «крючки» не дадут ему уйти на другую радиостанцию. Победит тот, чей «крючок» сильнее. Многие станции будут

незаконно использовать *нейролептическое программирование* слушателей в этих же целях. Подобное попытаются искоренить, но даже угроза закрытия станции не решит этой проблемы.

Та же ситуация сложится и на телевидении: зритель сам выбирает понравившегося диктора и заказывает, чтобы все новости с разных каналов читал именно он. Компьютерные технологии позволят смотреть по одной программе все фильмы или передачи, которые имеются в видеотеках всех каналов телевидения. Ну и, конечно, можно будет выбирать концовку фильма по своему желанию.

Сетка программ телевидения станет очень подвижной. Она будет прерываться в момент важных и экстренных событий: картинка автоматически будет переноситься не в студию, а на место происшествия. Зритель сможет использовать технологию детализации. Он сам будет выбирать в кадре тот фрагмент, который его интересует, укрупнять его, монтировать со следующим, с лицом ведущего и т. д. Зритель сможет стать и участником ток-шоу, соединяясь со студией по цифровому каналу и посылая в неё своё виртуальное изображение через Интернет.

Интернет-издания в будущем не будут иметь себе равных среди офлайновых и электронных СМИ. Среди них разгорится жесточайшая борьба за читателя. Расширение тем, многообразие ссылок и перекупка журналистов — это только немногие формы борьбы за популярность. Человек, попавший на определённый сайт, не должен уйти с него в поисках необходимой информации. Он должен увязнуть в сетях этого СМИ и сделать его стартовой страницей. Там окажутся не только собственные материалы журналистов того или иного издания, но и автоматические ссылки по конкретной теме на все СМИ и иные источники, размещённые в Интернете.

Интернет-издания станут синтетически сочетать в себе офлайновые и электронные технологии, такие, как текст, звук и видео. Многие материалы можно будет увидеть и услышать. Уже сейчас существуют радиостанции, которые можно не только услышать, но и увидеть в Интернете, поскольку специальные проекты снимаются на цифровые видеокамеры. Очень скоро будет сниматься и транслироваться весь эфир — интервью с гостями, работа диджея, появятся в Сети и видеоклипы.

Также в Интернете можно будет увидеть и все телепрограммы и фильмы всех телеканалов. Поэтому, попав на сайт, можно будет прочитать-увидеть-услышать и обсудить на форуме любой сюжет (как в онлайновом режиме, так и в архивно-ссылочном). Многие теле- и радиозвёзды перейдут в Интернет, поскольку смогут здесь делать в звуке и живом изображении практически то же самое, что они

ИНФОРМАЦИОННЫЕ ВОЙНЫ

В будущем самыми жестокими боями без правил в электронных СМИ будут информационные войны.

Аксиомой станет выражение: «Кто владеет информацией, правит миром». Информация, скорость её получения и обработки, напрямую будет связана с успешностью в бизнесе. Вся деловая активность сосредоточится в сфере купли-продажи акций предприятий и поставок. Поэтому конкурирующие организации будут использовать все имеющиеся каналы получения информации, и неоценимую помощь в этом оказывают все СМИ, в которых информация присутствует как в чистом, подробном виде, так и в аналитических материалах, упоминаниях в иных контекстах, намёках и в случайных цитатах. СМИ будут ещё больше гоняться за эксклюзивной, особенной, недоступной конкурентам информацией, чтобы разместить её у себя, опередить и выиграть зрителя-слушателя. При этом политики, представители власти будут использовать СМИ в своих целях активно и беззастенчиво. Естественно, что журналистика даёт представления и знания человека об окружающем мире, и сведения об этом мы можем получать только из СМИ. Часто нам кажется, что мнение по поводу какого-то факта мы формируем сами, исходя из собственных аналитических способностей. Однако это не так. Мы пользуемся информацией, которую нам кто-то предоставляет, и она может быть как субъективной, так и объективной. Она может быть дозированной и содержать необходимые информатору акценты и сведения. Например, СМИ могут сообщать только о повышении жизненного уровня в Европе и Америке. И никто и никогда не узнает о голоде где-нибудь в Африке, потому что об этом никто не сообщил. Мир огромен, и обычный человек не может знать, что происходит не только на другом конце света, но и в соседнем городе, если у него нет там знакомых.

Выигрывать войны будет тот, кто в своих интересах будет использовать средства информации, формируя общественное мнение. Если страна X будет бомбить страну Z, то для агрессора главными станут не боевые действий, а сражения на страницах газет и журналов, на ТВ-экранах и в радиоэфире. Если о реальной войне нельзя не сообщать, то войну можно назвать «мелкими столкновениями конкурирующих бандитских группировок». Ну а если всё-таки приходится говорить о полномасштабных военных действиях, то будут смещены оценочные характеристики. Никто не узнает правды, если страну-агрессора назвать государством, борющимся с «осью зла». Бандитов, наркодельцов и торговцев людьми, прикрывающихся националистическими лозунгами, можно именовать борцами за свободу и независимость, а тех, кто пытается противостоять им, клеймить как организаторов геноцида. С другой стороны, жестокость военных или полицейских сил по отношению к простым людям можно легко скрыть за рассказами о борьбе с бандитскими формированиями и их пособниками, пугая население уже случившимися или ожидаемыми терактами. Политики, крупные бизнесмены и целые государства будут всё больше прибегать к информационным и пропагандистским ресурсам журналистики.

На этом фоне сами войны будут просто играми в песочнице.

делали раньше, но расширят аудиторию. Можно будет увидеть, как автор сам читает свой материал. Кроме того, в онлайновом режиме можно будет увидеть и процесс работы редакции. Это окажется любопытно как обычным читателям, так и студентам факультетов журналистики, для которых такие наблюдения станут мастер-классом профессионалов.

АУДИО XXI ВЕКА

ПЕРЕЗАПИСЬ = КОПИРОВАНИЕ

Переход от кассет и виниловых пластинок к CD и мини-дискам улучшил качество звучания, в результате появилась возможность осуществлять перезапись со скоростью, намного превосходящей скорость воспроизведения. При перезаписи аудиокассет даже с удвоенной скоростью качество звука неминуемо ухудшается. Этого не происходит, если копировать цифровое представление звука. Скорость копирования зависит только от устройств, которые используются при этом. Например, CD на компьютере теоретически может достичь скорости чтения в 50 раз выше (в действительности около десяти раз) скорости обычного воспроизведения, а при записи на CD — в десятки раз выше. То есть 74-минутный CD удастся переписать за несколько минут. В сущности, 6 ч музыкальных записей, хранящихся в формате MP 3 (поток в 256 кбит/с), копируются за 6 мин.

Итак, копирование записей благодаря компьютерам очень упростилось. Теперь каждый может организовать у себя дома производство компакт- и мини-дисков. И оборудование, которое потребуется для этого, недорого. А как же авторские права? Законодательно оформленное запрещение «дикого» копирования вряд ли кого-то остановит. Необходимо вообще пресечь незаконное производство записей.

На практике, при наличии в каждом доме компьютера, подключённого к Интернету, это сделать не удаётся.

Механизм мини-дискового плеера.

В мини-дисковых устройствах установлен так называемый запрет на серийное копирование. Оборудование позволяет записывать как по цифровому (оптическому) каналу, так и по обычному, аналоговому. Если диск записан при помощи цифрового канала или несёт заводскую фонограмму, то переписать его можно только по аналоговому каналу (записывающая дека «знает», что пишет), а воспроизвести по цифровому. Тогда при перезаписи качество звука теряется. Напротив, если диск записан по аналоговому каналу, то его можно переписать по цифровому. При помощи компьютера этот запрет легко обойти.

ИСТОРИЯ NAPSTER

Летом 1999 г. 19-летний Шон Фэннинг создал программу Napster, позволяющую при помощи Всемирной паутины обмениваться музыкальными файлами в формате MP 3. В системе использовались специально выделенные машины Сети — серверы Napster, которые обеспечивали поиск MP 3-файлов на компьютерах подключённых пользователей, а сам обмен происходил напрямую. Существенная доля записей, циркулирующих в Napster-среде, распространялась в обход закона об авторских правах. За считанные дни Napster приобрёл огромную популярность у студентов американских университетов, но свободный обмен файлами закончился менее чем через полгода.

7 декабря Ассоциация индустрии звукозаписи Америки (RIAA) подала в суд на компанию Napster, обвинив последнюю в «прямом и косвенном нарушении авторских прав». Napster попал на первые полосы газет.

При рассмотрении иска адвокаты защищали свою позицию тем, что на серверах Napster не хранится пиратских материалов. В апреле 2000 г. группа «Metallica» подала в суд на Napster за нарушение их авторских

прав. Тогда же рэпер Андре Янг подал иск как на Napster, так и на всех пользователей, которые обменивались пиратскими записями его песен. Napster пришлось прекратить доступ к своим серверам более чем полумиллиону пользователей (любители metallica и поклонники Dr. Dre), чьи IP-адреса удалось выявить по заказу истцов. В ответ возмущённые пользователи направили жалобы адвокатам истцов. Только в течение одной недели было подано более 30 тыс. таких жалоб.

Правда, у Napster появились и сторонники: рэперы Chuck D и Ice-T назвали эту систему «радио двадцать первого века». Она даёт музыкантам возможность напрямую общаться со своей аудиторией. Соответственно увеличиваются объёмы продаж их записей.

12 июня 2000 г. RIAA потребовала закрыть Napster.

ПРАВО ЧИТАТЬ, СЛУШАТЬ, СМОТРЕТЬ

Napster представляет собой гораздо меньшую опасность, чем принято думать в RIAA. Запрет Napster не прекратил неконтролируемый обмен файлами в Сети. В ней живут и прекрасно себя чувствуют десятки версий Gnutella и подобных программ, которые существовали параллельно с Napster. Так что конфликт компании Napster и махины американской индустрии звукозаписи не сводится только к проблеме нарушения авторских прав.

Виниловая пластинка, CD и мини-диск.

Ричард Столмен, «отец» GNU, написал очень удачный фантастический рассказ «Право читать», в котором за несколько лет до истории с Napster поднимались те же проблемы.

Наличие систем, подобных Napster, — свершившийся факт, и запретить их, разумеется, нереально. Можно, конечно, организовать сколь угодно сложные методы защиты, издать сколь угодно суровые законы, только это приведёт к обратному результату. Скорее всего, объёмы продаж упадут, а защита будет вскрыта. Если оборудование ограничивает свободу граждан в том, что слушать, смотреть и читать, то таким оборудованием перестают пользоваться. Получается, что владеть авторскими правами в XXI в. — самое выгодное

> «С первого класса в школах внушали, что поделиться книгой — ужасное преступление, сравнимое с морским пиратством. Шансов ускользнуть от бдительного ока SPA — Software Protection Authority (Служба защиты программного обеспечения) — практически не было. Дэн знал, что каждая книга имела контрольный монитор, который сообщал, кто и когда её читает, в Центр лицензирования (Central Licensing). Как только компьютер войдёт в Сеть, Центр может это засечь, а Дэн, как владелец компьютера, понесёт самое тяжёлое наказание за то, что не предупредил преступления».
>
> *Р. Столмен. «Право читать»*

Некоторые университеты закрыли доступ к Napster, обнаружив, что 20 % трафика составляют MP3-файлы. Возмущённые студенты организовали кампанию протеста и сбор подписей против закрытия доступа к Napster.

«Дэн знал, что её (героини. — *Прим. ред.*) семья принадлежит к среднему классу и с трудом оплачивает обучение, так что на книги и вовсе не хватает. Чтение чужих книг для неё могло быть единственным способом завершить образование. Ситуация была знакомой: он сам влез в долги, оплачивая статьи, которые приходилось читать (10 % этих средств получали авторы; поскольку Дэн мечтал об академической карьере, он надеялся, что на его собственные исследования будут ссылаться достаточно часто и он сможет вернуть долг).

Позднее Дэн узнал, что было время, когда каждый мог пойти в библиотеку и бесплатно получить журнальную статью и даже книгу. В те годы существовали независимые студенты, которые могли читать тысячи страниц, не прибегая к правительственным библиотечным грантам».

Еще одна программа — Gnutella, которую выпустила компания Nullsoft, автор MP3-плеера Winamp. В соответствии с идеологией GNU вместе с программой распространялись и её исходные тексты. Gnutella, в отличие от Napster, позволяла, не подключаясь к серверам системы, создавать собственные сети и обмениваться файлами внутри них. Используя Gnutella, можно организовать обмен любыми файлами. Кроме этой программы была выпущена целая серия подобных продуктов.

предприятие. Однако, если некто может подать иск на врача, который не вылечил больного, то почему нельзя подать в суд на автора книги, если эта книга не понравилась читателю. Ведь проценты автору выплачены сполна.

Сражение за право читать идёт уже сегодня. Причём борются не за абстрактные свободы, а за адаптацию существующих законов, традиций и норм, сложившихся в течение нескольких столетий, к новой реальности, возникшей с появлением Интернета.

В этих новых условиях воротилам индустрии звукозаписи и видеозаписи, а также и книгоиздателям, видимо, придётся в корне пересмотреть способы своего взаимодействия с аудиторией. Возможно, авторы и компании откажутся от получения сверхприбылей, и тогда стоимость CD сравнится со стоимостью их распространения по Сети, что в конечном счёте только увеличит авторский гонорар. Ведь именно так и борются с пиратством, распространяя легальную продукцию по цене ниже контрафактной.

Аккорд — одновременное сочетание трёх и более звуков, как правило расположенных по терциям. На нотном стане ноты, обозначающие терцию, лежат на соседних линиях или в соседних промежутках между линиями, например до-ми-соль. Терция — интервал между нотами в 1, 1/2 или 2 тона (малая и большая).

ПРОГРАММИРОВАНИЕ МУЗЫКИ

Ещё в начале 80-х гг. в некоторых моделях популярных электронных клавишных инструментов — электроорганах — появилась очень удобная функция, когда композитор мог не только извлекать электронный звук, но сразу распечатывать ноты сыгранной мелодии. В те времена это было пределом мечтаний многих композиторов — ноты без карандаша и бумаги!

Первые инструменты с автоаккомпанементом произвели настоящий фурор. Теперь практически каждый мог выступать с персональными концертами, не привлекая дополнительно музыкантов. В СССР в конце 80-х гг. существовал огромный спрос на исполнителей популярной музыки, обладающих домашним синтезатором с автоаккомпанементом. Тут сыграл определённую роль дефицит композиторов, пишущих поп-музыку. Появилась масса групп и псевдокомпозиторов, придумывающих и исполняющих простейшие мелодии. Они сочиняли мелодии всего в три-четыре такта с использованием авто-

аккомпанемента, записывали и успешно продавали новоиспечённые хиты. Вероятно, сказались бреши в массовой музыкальной культуре россиян, ведь долгое время поп-музыка была фактически под запретом.

В отличие от оркестровки, когда произведение (для фортепиано) «раскладывается» на инструменты оркестра, при аранжировке к основной музыкальной теме композитора могут быть добавлены штрихи, характерные жанру композиции.

Аранжировка строится на том, что в любом музыкальном стиле есть набор некоторых правил, ограничивающих композитора. Существуют общие правила для разных стилей, так, практически любой аккорд имеет терцовую основу. Но, например, в хард-роке используют так называемый двойной бас (добавление к основной мелодии двузвучий электрогитары). Разные стили, как правило, имеют и различный ритмический рисунок. Например, смещение в такте акцента на некоторую долю вперёд или назад даёт характерную блюзовую окраску произведению. Это смещение, которое может быть и в басу, и в ударной группе, даёт ощущения ожидания очередной музыкальной фразы блюза, создаёт его неповторимый рисунок. При этом характер мелодии (мажор, минор) и темп произведения не являются атрибутами стиля.

Как при традиционной, так и при компьютерной аранжировке стили представляют собой наборы шаблонов, приёмы аранжировки, характерные для конкретного стиля. Они хорошо известны, и их изучают в процессе специального музыкального образования будущие композиторы и аранжировщики. Эти приёмы частично формализованы и используются в синтезаторах с автоаккомпанементом и в специальных программах.

Первые опыты компьютерной аранжировки хотя и давали правильные с точки зрения теории музыки композиции, но были уж очень просты, не учитывали характер мелодии. Одна из популярных программ Band-In-a-Box (фирмы PG Music Inc.) предоставляла музыканту удобный простой интерфейс, возможность исполнения мелодии без midi-клавиш, прямо на клавиатуре компьютера: нижний ряд клавиш имитировал белые, а следующий ряд — чёрные. Оркестровка позволяла использовать пять инструментов и даже монтаж произведения с голосом вокалиста. Исполнению готовых композиций можно было подыгрывать, причём имелся интересный режим (вероятно, для начинающих исполнителей), когда неправильно сыгранная нота автоматически корректировалась до ближайшей правильной.

Синтезатор.

Более совершенные программы позволяли в созданных сэмплах (звуках инструмента) на ходу менять высоту, темп, частотную характеристику

Сейчас часто происходит подмена понятий, и электронная музыка понимается слишком широко, а именно как любая композиция, сыгранная при помощи электронных инструментов. На самом деле если музыкальное произведение может быть сыграно на «живых» инструментах без заметной смены звучания, то вряд ли стоит его относить к электронной музыке. Электронными являются те композиции, которые немыслимо создать или воспроизвести без помощи компьютера и электронных инструментов.

В современных студиях электронной музыки отсутствуют не только привычные инструменты, но даже midi-клавиатуры.

мелодии. Так можно было напеть популярную музыкальную фразу на одном по высоте тоне, соблюдая только ритмический рисунок. Затем выбрать стиль и сыграть мелодию, используя этот сэмпл в качестве нового звучания инструмента. Например, «Эх, дубинушка, ухнем» в стиле регги.

Теперь, когда имеется огромный выбор сэмплов, можно на ходу менять сам ритмический рисунок и программировать практически любые композиции разных исполнителей в каком хочешь стиле. Компьютерная аранжировка неудобна при записи с вокалистом, так как компьютерное исполнение слишком правильно соблюдает ритм и не следит за голосом. То, что может «запрограммировать» на компьютере один композитор, может сделать и другой. Правда, поражает воображение огромное количество материала, который используется при аранжировке. Если выбран индивидуальный «почерк», то можно даже на компьютере с автоаранжировкой, создать неповторимое произведение.

Синтезатор-фортепиано Yamaha 2000 pro.

Так, популярная программа DJ Dance (фирмы Ejay) фактически позволяет собирать композицию на 32-дорожечном микшере. Набор сэмплов для прогрессивной танцевальной музыки исчисляется десятками тысяч, и можно создавать свои сэмплы.

Разнообразный набор эффектов, драм-машина и многое другое позволяют получить специфическую музыкальную студию у себя дома. Надо только иметь чуть-чуть музыкального слуха, немножко вкуса, и при полном незнании музыкальной грамоты можно создавать неповторимые композиции.

Сейчас многие композиторы активно применяют компьютеры в своей творческой работе, причём на вопрос, что может делать композитор при помощи компьютера, часто отвечают — практически всё. Вот насколько компьютер удобен для композитора. Однако надо сразу оговориться, что всё, кроме фантазии, творчества и воображения. К счастью, пока машина не даёт сама себе заданий на написание той или иной композиции.

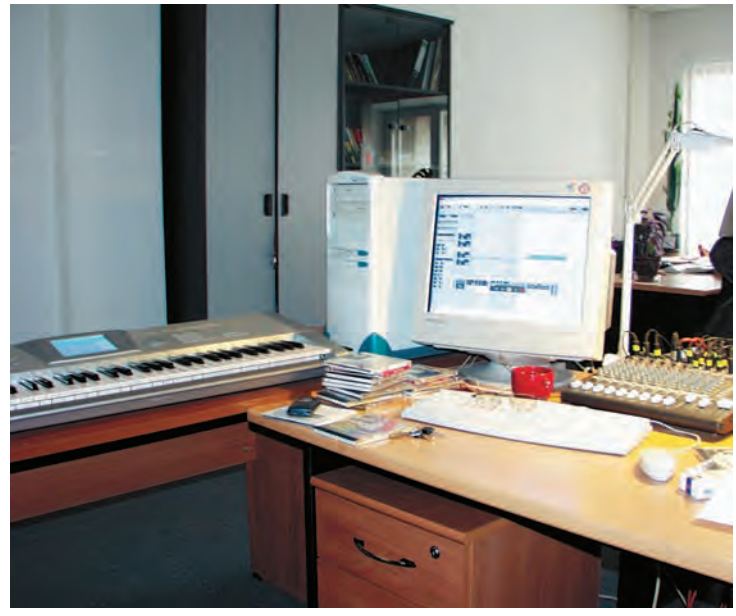
При этом компьютер можно использовать на протяжении всего процесса создания нового произведения. В начале просто как «пишущую машинку» для композитора. Потом, в процессе аранжировки, «окраски» основной темы в различные жанры, компьютер фактически даёт новое звучание композиции: от традиционных стилей до элитарной электронной музыки. Затем машина выполняет аудиомонтаж, например при подготовке CD.

Если речь идёт об электронной композиции, то отдать монтаж «на сторону», как правило, немыслимо для автора, так как монтаж может изменить авторское звучание электронной музыки.

При столь развитой вычислительной технике композитор имеет воз-

можность работать у себя дома, а не в студии. Для создания нового произведения используются разные инструменты и предустановленные сэмплы в синтезаторе либо на компьютер с CD ROM загружаются новые (профессионалы имеют обширную базу сэмплов, занимающую сотни CD ROM, содержащую тысячи различных звучаний саксофона, альта и т. п.), в том числе из Интернета. Можно, например, включить в композицию партию для клавишных, которая будет звучать точно, как рояль Steinway в Карнеги-холле.

Сейчас, взяв с собой в путешествие ноутбук с midi-интерфейсом, композитор может организовать себе студию в любой точке мира, что-то дописать, доделать и даже создать, летя в самолёте, а затем воспроизвести на концерте, подключив свой компьютер к аппаратуре на сцене.



Рабочее место современного композитора.

КИНО И ФОТО В НОВОЙ ЭПОХЕ

ЦИФРОВЫЕ АКТЁРЫ

Голливуд, как известно, «фабрика грёз». Как ни ругают американские киностудии за бесконечную погоню за прибылью, надо признать, что американское кино всё же не случайно лидирует в прокате. Индустрия кино развита в США лучше, чем где-либо в мире.

В конце 80-х гг. как в России, так и за рубежом предвидели появление симбиоза компьютерных игр и кинофильмов. С одной стороны, это была всё та же игра, в которой человек мог активно влиять на происходящие на экране компьютера события. С другой стороны, предполагалось, что это всё-таки кино, где играют настоящие актёры, только практически каждый эпизод имеет несколько вариантов.

«Ходилки» и «стрелялки» знакомы каждому, но до появления качественного изображения на экране ЭВМ и носителей информации, способных вместить весь обильно ветвящийся сюжет, о таких киноиграх можно было лишь мечтать. Всё изменилось к середине 90-х гг. XX в., когда стали появляться первые подобные игры на компакт-дисках.

Настоящие актёры с экрана монитора разговаривали с игроком и друг с другом, бегали, стреляли и даже умирали в игре (а в кино умирают только один раз). Так появилось (естественно, в Голливуде) новое амплуа — digital actor (*англ.* «цифровой актёр»). Одной из первых в этой серии была игра «Critical Path» («Опасный путь»), выпущенная в середине 90-х гг. XX в.

DVD и видеокассета.

Вот что говорит актриса Эйлен Вейсингер о своей первой роли в игре: «До этой роли я использовала компьютер только для написания текстов. Когда мне сказали, что буду сниматься в интерактивной игре, я подумала, что игрок просто будет управлять моими движениями, как в играх про карате. Меня посадили за компьютер и показали на примере игры "Шерлок Холмс", как всё будет выглядеть».

Технология blue screen одинакова для съёмок кино и игр. Актёр играет сцену перед голубым экраном, а оператор тщательно следит, чтобы на экране не появлялись тени. Если актёр делает неверное движение, например рукой, не видя препятствия, то в компьютерной сцене там может располагаться стена и герой просто проткнёт её. Затем отснятый материал цифруют, вводя в компьютер, заменяя синий цвет на белый и заполняя сценами из кинофильма или игры. В традиционном кинематографе для обработки сцен комбинированных съёмок часто обходятся без компьютера и используют чисто аналоговую аппаратуру, которая «вырезает» героя из синего экрана и накладывает на отснятый материал. Синий (а часто и зелёный, особенно в виртуальных телевизионных студиях) цвет экрана используется потому, что в самом изображении человека этот цвет практически не встречается и его легко выделить.

В процессе съёмок актёры снимаются не только в массе дублей, но и во многих сценах, которые похожи одна на другую, ведь речь идёт о различных финалах одного и того же эпизода. Актёр умирает сотни раз за игру, причём каждый раз по-разному. Цифровым актёрам трудно вживаться в образ, так как сюжетная линия неединственна. Кроме того, актёр должен сыграть роль так, чтобы играющим хотелось отождествить себя с героем.

«Моя героиня имеет очень сложный характер. Мне хотелось сыграть её как можно правдоподобнее, при этом так, чтобы играющий действительно захотел спасти её. Я добавила ей немного юмора», — вспоминает Эйлен Вейсингер.

Не все популярные актёры согласятся участвовать в интерактивном кино или игре. Многие относятся к новой актёрской специальности как к менее творческой. Игры напоминают собой «мыльные оперы», когда сценарий на сегодняшнюю съёмку готов только вчера и актёры не знают судьбы своего персонажа до конца съёмок.

Однако менее столетия назад так же относились и к профессии киноактёра, считая игру в фильмах искусственной из-за неизбежного обилия дублей на съёмках.

Фактически первый компьютерный анимационный фильм сняла студия Диснея, а назывался он «История игрушек». За ним последовали «Муравей Анц» и «Жизнь жуков». Но это были мультфильмы.

«Конечно, я расстроена, что играющие ожидали увидеть главным героем мужчину. В процессе съёмок я представляла игроков двуполыми существами. Зато я довольна своим костюмом. Мне достался лёгкий комбинезон с топиком и автоматом через плечо. Меня однажды захотели одеть в короткие шорты, но я отказалась, решив, что моя героиня не кошечка. Она что-то среднее между Сарой Коннер из ''Терминатора'' и француженкой Никитой».

Из рассказа актрисы Эйлен Вейсингер

КОМПЬЮТЕРНОЕ КИНО

В течение столетия киноплёнка оставалась основным носителем при производстве и показе фильмов. Но, похо-

Кадр из фильма «Последняя фантазия. Духи».

же, в XXI в. всё изменится: наступает эпоха цифрового кинематографа. И дело даже не в том, что некоторые кинематографисты уже снимают на мини-камеры формата DV (digital video), а материал затем переносится на киноплёнку и показывается в кинотеатрах. И не в том, что производство «Истории игрушек-2» целиком было цифровым и в большинстве случаев при прокате использовались цифровые проекторы. Дело в том, что в XXI в. появилось компьютерное кино, которое выглядело совсем не как мультфильм, а как настоящий кинофильм, в котором играли не люди, а созданные на компьютере персонажи. Одним из первых опытов стал фантастический боевик «Последняя фантазия. Духи» японского режиссёра Хиронобу Сакагучи, появившийся на экранах в 2001 г. Персонажи «Последней фантазии...» созданы при помощи компьютеров, причём в них всё правдоподобно, вплоть до пор на лице. Вымышленный мир достоверно воссоздан с помощью новейших разработок в области спецэффектов.

События в фильме разворачиваются в далёком 2065 г., когда Землю захватили инопланетяне, пожирающие души людей. Некоторые оставшиеся на Земле существа могут сопротивляться вторжению фантомов, но для полной победы «внутренние духи» необходимо собрать в одном живом существе — докторе Аки Росс. Кроме того, ей и её товарищам нужно остановить военных, которые хотят попросту уничтожить Землю.

Выход фильма с бюджетом почти 150 млн долларов любители анимационного кино восприняли на ура. Некоторые герои получились более естественными, чем другие, но ещё никогда раньше модели персонажей не были столь похожими на живых актёров. Когда смотришь фильм, забываешь о том, что перед тобой творение программистов. «Последняя фантазия...» стала компьютерным прорывом в кинематографе.

Американскую версию «Последней фантазии...» озвучивали такие звёзды, как Алек Болдуин, Доналд Сазерленд, Джеймс Вудс (его голосом говорил злодей). С изображениями компьютер уже справляется, чего нельзя сказать о голосе. Здесь пока нужны живые актёры.

Вскоре после выхода на экраны фильма «Последняя фантазия...» был продемонстрирован визуальный образ героини, но не с помощью цифрового видео, а непосредственно на компьютере, оснащённом графическим процессором nVidia Quadro DCC. Машина в реальном времени генерировала кадры. При этом только для прорисовки причёски Аки процессор за секунду обрабатывал более 1,5 млн элементов изображения, что превышает число волос на голове у человека.

Кадр из фильма «Последняя фантазия. Духи».

Появление цифровых актёров приведёт к большим изменениям у настоящих работников этой профессии. Гильдия актёров (Screen Actors Guild — SAG) серьёзно обеспокоена тем, что в скором времени обычные актёры, особенно статисты, могут остаться без работы. На роль партнёрши Аль Пачино в фильме «Симона» режиссёр Эндрю Николь разыскивал цифровую актрису, так как предложенные живые его не устроили. Впоследствии шум слегка улёгся, когда оказалось, что сценарный замысел перепутали с режиссёрским запросом. Главная героиня должна быть просто ненастоящей.

Не исключено, что незавидная судьба ожидает настоящих звёзд Голливуда. Так, образ реального актёра может быть использован для создания его цифрового дублёра. Известный актёр Чарли Шин предложил образовать комитет, который защитил бы его коллег от копирования.

В конце 2002 г. на экраны вышла очередная компьютерная лента «Элизиум». Этот анимационный фильм уже неуместно называть мультипликационным, так велико ощущение реальности происходящего на экране, хотя авторами не ставилась задача заменить настоящих живых исполнителей компьютерными.

Однако, возможно, недалёк тот день, когда «Оскара» за лучшую роль (мужскую или женскую) получит цифровой актёр. А вручать желанную статуэтку будут... компьютеру.

Карта памяти современного сотового телефона.

ОТ ЧЁРНО-БЕЛОЙ ФОТОГРАФИИ К МОНТАЖНЫМ СТУДИЯМ

Цифровые технологии незаслуженно часто обвиняют в бездушности. Фотохудожники XX в. использовали аналоговые широкоплёночные фотоаппараты, тщательно выбирали сюжет фотографии, место и освещение. Для проявки плёнки изготавливали собственные составы реактивов. Процесс печати был продолжителен и требовал специальных условий (проходил в темноте). Но вся цепочка от съёмки до выставки содержала в разной степени элементы творчества. Сейчас технология изменилась.

Большинство операций производится при помощи компьютера, а фотограф определяет лишь сюжет и объект съёмки. Можно сказать, что всё остальное делается компьютером, а не художником, которому остаётся только «нащёлкать плёнку» на одноразовом фотоаппарате и отдать её в студию на обработку. Машина всё выполнит сама. И в результате получатся фотографии неплохого качества. При старой технологии гарантированного результата добиться было довольно сложно.

Но осталось немало художников, использующих аналоговые фотоаппараты и раритетную технологию и считающих, что при цифровой творчество сведено на нет.

На самом деле новая технология даёт художнику ещё больше простора для творчества. Просто место тёмной комнаты, химикатов и фотоувеличителя заняли компьютер и программы обработки изображений. И неважно, использует настоящий художник мольберт и палитру или компьютер и принтер.

АВТОМОБИЛЬ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

По некоторым прогнозам, век двигателя внутреннего сгорания скоро завершится. Так, в США через пару десятков лет в результате реализации широкомасштабного проекта «Автомобиль Свободы» три ведущие корпорации Daimler Chrysler, General Motors и Ford должны будут разработать и осуществить массовое производство автомобилей, использующих вместо бензиновых моторов электродвигатели. Предполагается устанавливать моторы, работающие за счёт окисления водорода. Этим будет обеспечиваться экологическая чистота автомобилей. Мало того, в результате значительно снизится потребление нефти, а значит, и зависимость государства от её импорта.

Информационные технологии лишь косвенно влияют на смену типа силового агрегата в автомобиле. Основные изменения будут связаны с повсеместным внедрением электронных коммуникационных устройств в транспортные средства.

Компания Daimler Chrysler подготовила концептуальную машину с оборудованием, позволяющим не только информировать о пробках на дорогах и расстоянии до того или иного пункта, но и получать почту через Интернет и просматривать сайты. На панели перед водителем расположен компьютер с голосовым управлением, а также два терминала с сенсорными экранами для пассажиров. Предполагается, что во время движения не только водителя, но и пассажиров будет интересовать содержимое сайтов.

Помимо Daimler Chrysler два других гиганта американской автомобильной индустрии — General Motors и Ford ведут разработки в области беспроводных телекоммуникаций и онлайнового сервиса в применении к автомобилям. Как считают специалисты, такие интегрированные средства обработки и передачи информации сделают машины более привлекательными для автовладельцев.

Коммуникацию можно обеспечивать не только с внешним миром, но и внутри автомобиля. Так, если в центральном блоке использовать кодированный сигнал для световых приборов и передавать его по одному проводу, например на задние фонари, где стоит декодирующий блок, то вместо толстого пучка проводов, идущего по машине к багажнику, понадобится только два провода: по сигнальному будут передаваться управляющие импульсы, а силовой подведёт питание к фонарям и декодирующему блоку. (Как правило, минусовый провод в автомобиле заземлён на железный кузов, поэтому не потребуется третьего провода.) Сигналы, представляющие собой короткие импульсы, станут кодировать включение и выключение соответствующей лампочки в фонарях (три импульса — включить стоп-сигнал, четыре импульса — выключить стоп-сигнал, пять импульсов — включить мигание левого поворотника, шесть — выключить и т. д.).

Другой сферой приложения информационных технологий являются системы управления двигателем, трансмиссией, бортовые вычислители — информационные компьютеры, электронные средства охраны автомобиля и т. п.

По прогнозам, в 2010 г. доход от информационных систем в автомобилях вырастет до уровня 50 млрд долларов США в год с суммарного мирового объёма этой отрасли в 4 млрд долларов на начало XXI в.

Соединённые Штаты Америки потребляют 25 % добываемой в мире нефти.

Салон современного автомобиля «форд».

Системы управления и диагностики агрегатов автомобиля появились ещё в XX в. и распространены в так называемых моторах с инжектором. Иногда приходится слышать от автовладельцев жалобы на то, что у их машин «полетел» компьютер. Правда, эта часть автомобиля, как правило, самая надёжная. В XX столетии системы управления подачей топлива в моторе использовали в качестве элементной базы не микросхемы, а «россыпь» элементов: транзисторы, сопротивления и др. Часто поломка компьютеров связана с условиями эксплуатации автомашин, ведь перепады температуры могут быть от −50°C до + 50°C. Такая компоновка компьютера неудобна, так как перепрограммирование сводится к перестройке блока, часто с заменой элементов, когда не обойтись без паяльника. При использовании интегральных схем программа, в частности управляющая зажиганием и корректирующая угол его опережения в процессе работы, целиком помещалась в память управляющего процессора и учитывала тысячи значений. При изготовлении такого блока для нового мотора не требовалось перепроектировки всего блока, более того, блок оставался тем же самым, только в нём менялись программа или её данные. Таким образом, снижались финансовые затраты при производстве двигателя.

Всё вышесказанное относится и к другим системам, например блокам управления автоматической коробкой передач, кондиционером, трансмиссией полноприводных автомобилей.

Состояние мотора также можно оценивать при помощи компьютера. Для регулярной смены масла владельцу не придётся запоминать последние показания одометра, компьютер сам известит о необходимости заехать на станцию техобслуживания, причём заранее предупредит об этом. Если условия эксплуатации были тяжёлыми, то, основываясь на показании специальных масляных датчиков, ком-

Концепт-кар — образец машины недалёкого будущего, изготовленный фирмой Daimler Chrysler.

Концепт-кар «Лада» Автосалон-99. Москва.

пьютер может предупредить о внеочередной замене. Правда, бортовые компьютеры XX в. были не очень информативными. Они, как правило, лишь показывали расход топлива и оценивали, сколько горючего потребуется, чтобы преодолеть конкретное расстояние.

Более важен для автомобилиста маршрутный компьютер, который подскажет, как добраться до места назначения. При этом удобно знать положение самого автомобиля, что определяется при помощи GPS (*англ.* Global Position System) — системы глобального позиционирования. Персональный электронный штурман при помощи 24 спутников, находящихся на геостационарных орбитах, позволяет определять местоположение, направление и скорость движения, планировать маршрут в поездках и путешествиях. Программное обеспечение обычно состоит из небольшой информационной системы, где на фоне карт показаны положение автомобиля, предполагаемый маршрут движения, расстояние до гостиниц, автозаправок и т. п. Такие системы автоматически прокладывают маршрут до нужного места, отыскивая в базе данных улицу, гостиницу.

Для работы подобных систем необязательно иметь встроенный в автомобиль компьютер или, например, ноутбук. Существуют мини-системы для PDA, которые представляют собой электронную карту (на ней указаны все объекты вплоть до каждого дома) с базой данных в десятки тысяч адресов, обеспечивающую выбор оптимального маршрута движения, поиск объекта и получение подробной информации о нём. Такую систему вместе с GPS-приёмником можно брать с собой, выходя из автомобиля, чтобы прогуляться по городу пешком.

Геоинформационная система, содержащая описания большинства дорог США и приёмник GPS, подключаемый к ноутбуку, занимает четыре CD ROM-диска.

GPS-навигатор.

ЭЛЕКТРОННЫЙ ДОМ

В американском фантастическом фильме «Вспомнить всё» жена главного героя управляет квартирой голосом: включает и выключает свет и телевизор, тренируется под руководством голограммы теннисиста. Герой актёра Брюса Уиллиса в кинокартине «Пятый элемент» живёт скромно, тем не менее получает электронную почту и пользуется видеотелефоном.

Что такое электронный, или, как его ещё часто называют, «цифровой», дом? Многие знают об этом чуде техники из художественных произведений. Любой писатель-фантаст (или режиссёр), создавая книгу (фильм) о будущем, считает своим долгом рассказать и об устройстве быта. Причём этот быт должен поразить читателя (зрителя) удобством и совершенством.

Электронный дом обязательно включает средство связи — как правило, нечто среднее между видеотелефоном и телевизором. В нём действует система безопасности: дом «узнаёт» хозяина, например по отпечаткам пальцев. Домашние приборы выполняют голосовые команды. И естественно, в фантастических произведениях показаны страшные эпизоды: домашние компьютеры выходят из-под контроля владельцев и с помощью охранных систем захватывают в заложники людей.

Как же обстоит дело в реальности? В начале января 2001 г. в Лас-Вегасе (США) на очередной выставке бытовой электроники CES 2001 главное внимание участников было уделено концепции цифрового дома. Её суть состоит в том, что домашние приборы, включая компьютеры, телевизоры, музыкальные центры, холодильники, микроволновые печи и другие приборы должны представлять собой общую, надёжную, быстродействующую систему.

Лидер производителей микропроцессоров — фирма Intel предложила объединить бытовые устройства вокруг мощных домашних персональных компьютеров. Intel назвала свою технологию «эрой новых возможностей» (Extended PC Era). С развитием электроники бытовые устройства вместе с домашними компьютерами создадут единую среду, благодаря которой расширятся возможности и домашних приборов, и персонального компьютера. ЭВМ будет эффективно управлять бытовой техникой. Уже сейчас нетрудно обновлять программное обеспечение многих устройств. Так, в некоторые модели стиральных машин можно добавлять новые программы или корректировать старые (особенно, если в процессе массовой

Выдающиеся писатели-фантасты братья Стругацкие рисовали в своих произведениях самые разные по характеру картины мира будущего. В повести «Понедельник начинается в субботу» они с юмором изобразили институт магии и чародейства. В произведениях «Стажёры» и «Полдень XXII века» описан быт жителей Земли, подчинённый общей идее строительства лучшего мира.

ЧУДО-ПЕЧЬ

Речь пойдёт о самом любимом нами помещении в доме — кухне. Чего только не изобретали учёные мужи, чтобы облегчить тяжёлый труд матерей и жён. Это и знакомые всем холодильники и плиты, и пока ещё редкие в России посудомоечные машины, и микроволновые печи. Вот об этих печках и поговорим. Современные микроволновые печи могут самостоятельно приготовить множество блюд, надо только выбрать программу и нажать на кнопку. Словом, всё компьютеризировано, чего же ещё желать?

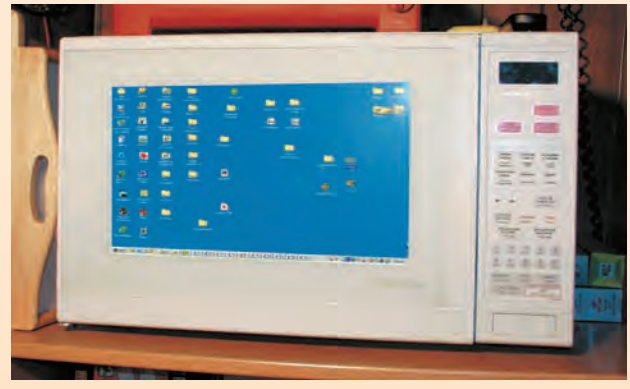
Но представим на минуту, что печь подключена ко Всемирной сети, а также снабжена лазерным сканером, таким же, каким пользуется кассир супермаркета для выяснения стоимости продукта, когда пробивает чек. Итак, мы дома кладём продукты в холодильник, пронося их мимо сканера. Наш компьютеризованный помощник сразу узнаёт название продукта, его вес и срок хранения (в запечатанном и открытом виде) и всё это запоминает. Ну вот, холодильник полон. Теперь примемся за готовку.

Выберем блюдо — печка может приготовить любое кушанье из поваренной книги! При этом смотрим на стекло печки — это цветной жидкокристаллический экран компьютера, дающий изображение блюда. Или «листаем» длинный список блюд, произнося: «Ещё, ещё...». Или, чтобы не отвлекать разговорами с микроволновкой домашних, по старинке жмём на кнопки. Простите за прозу, готовим «суп гороховый». Значит, ищем на букву «с» — «суп», находим среди супов «гороховый», для наглядности смотрим фото, потом выбираем рецепт — столько-то того, столько-то этого... И, о чудо, печь знает, что все компоненты из рецепта есть в холодильнике. Выложили продукты — варим. Помощник не дремлет, дзинь, нам сообщение от печки: «Напоминаю вам, что молоко, давно находящееся в холодильнике, завтра испортится, истекает срок хранения». Надо что-то сделать — вылить или выпить. Налили стакан молока. Несём пустой пакет к сканеру. Говорим: «Пусто». Помощник вычёркивает его из списка продуктов. Достаём ещё не вскрытый пакет. Снова проносим мимо сканера. Печь говорит: «Срок хранения истекает через четыре месяца». Открываем пакет — и к сканеру. Говорим: «Открыли». А помощник в ответ: «Срок использования — две недели при хранении в холодильнике».

А вот и суп сварился. Попробовали. Вкусно.

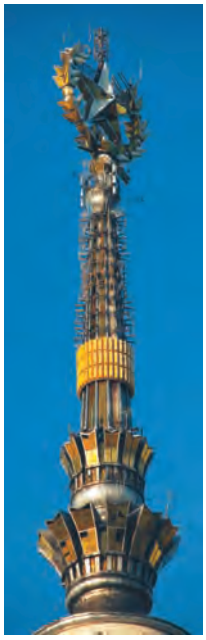
Теперь другой рецепт. Кстати, поищем что-нибудь экзотическое в Интернете. Что там можно найти на его просторах? И что для этого есть у нас в холодильнике? Но даже не будем его открывать, а поговорим с печью, полистаем список на экране. Выбираем продукты и спрашиваем печь. Что из этого можно приготовить? «Ага, вот список... большой... А из этого? Ничего нельзя, жаль, а если в магазин сходить? И что-то прикупить?» Да, теперь можно. Выбираем рецепт, и печка сама заказывает в ближайшем супермаркете недостающий продукт — через Интернет, естественно. Вот и посыльный. «Заказывали?» — спрашивает. «Да», — отвечаем. Начали готовить!

А пока микроволновая чудо-печь стряпает, мы смотрим на её экране кинофильм, полученный по трансляции из Интернета. Как на экране телевизора. Только вся реклама в перерывах — от фирмы — изготовителя микроволновой печи, и промотать вперёд ролики нельзя. Прошло немного времени — и вкусный обед на столе. Пообедали. Поставили грязную посуду в посудомойку. А вдруг она тоже подключена к Интернету?



эксплуатации были обнаружены программные ошибки).

Корпорация Microsoft в концепции нового домашнего управляющего компьютера предложила использовать вычислительные средства ЭВМ для бытовых приборов, которые сами не обладают вычислительной мощью.



Шпиль главного здания МГУ, увенчанный антеннами.

Человек сможет управлять компьютером отовсюду с помощью специальных сенсорных панелей.

Группы ЭВМ объединяются в сети, как правило, при помощи проводов. Но в наши дни всё чаще используют беспроводные соединения, особенно в локальных сетях, в пределах одного помещения или нескольких комнат. Высказывается даже предположение, что в наступившем десятилетии такая технология будет главной. Вероятно, средства беспроводной коммуникации станут основой функционирования домашней сети, в которую войдут все домашние приборы.

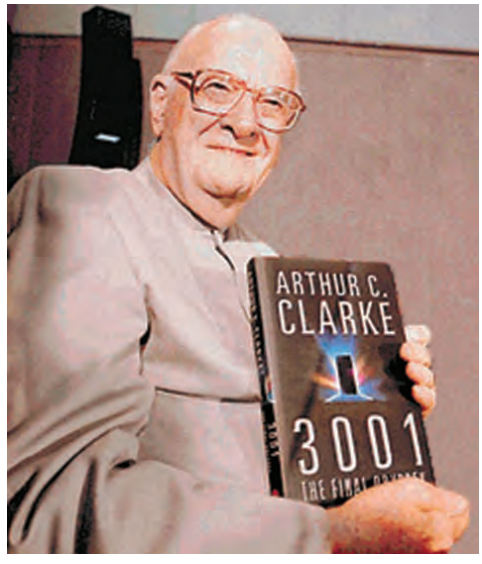
Управлять электронным домом можно будет откуда угодно через PDA (электронную записную книжку), Интернет или даже сотовый телефон. Регулирование климата (температуры помещения) позволит сэкономить немало денег, например, когда хозяева уезжают в отпуск. Ведь если в доме нет животных или теплолюбивых растений, то нет и необходимости поддерживать обычную температуру в жилых помещениях. А перед возвращением хозяев температура будет автоматически доведена до нормы. Если вы решите в отпуске поснимать на цифровые видео- и фотокамеры, материал будет пересылаться через Интернет на домашний компьютер, так что необходимость возить с собой массу карт памяти для хранения записей отпадёт. Вероятно, такая же перспектива ждёт и MP 3-плеер, и магнитолу в автомобиле.

Цифровые дома уже продаются в Европе, но пока они остаются скорее красивыми символами будущего, нежели атрибутом современной жизни. Одна из основных возможностей современного электронного дома, которая рекламируется продавцами, — дистанционное включение и выключение света в комнатах и манипуляции с занавесками (так хозяева показывают потенциальным ворам, что они дома). И, конечно, 50-дюймовый плазменный телевизор в гостиной с возможностью покупок online.

Но всё быстро меняется. В 2002 г. известный энтузиаст информационных технологий из США Джозеф Джейкобс ввёл себе под кожу микрочип, чтобы лучше «понимать» компьютер. Не исключено, что подобная технология вскоре позволит человеку управлять цифровым домом.

Примерно те же задачи сейчас выполняют обычные автомобильные сигнализации: они позволяют владельцу дистанционно запускать мотор автомобиля, прогревать салон, контролировать температуру в машине. Ведь так приятно, выйдя из дома, сесть в тёплый автомобиль, когда за окном 20 градусов мороза.

ПРОГНОЗЫ НА БЛИЖАЙШЕЕ СТОЛЕТИЕ



Артур Кларк на презентации своей новой книги.

Человек всегда стремился узнать, что его ждёт в будущем. Всматриваясь в звёзды, раскидывая карты, он надеялся найти ответы. Часто в качестве прорицателей и предсказателей будущего выступают писатели и учёные.

Имя Артура Кларка знакомо даже тем, кто научной фантастикой не интересуется. Славу ему принесли не только многочисленные романы и выступления на страницах известных изданий, но и фильм Стэнли Кубрика «2001 год: Космическая одиссея», поставленный по сценарию Кларка. Писатель просит не называть его пророком, хотя значительное число его предсказаний сбылось. А вот некоторые из его прогнозов на будущее

...10-е гг. XXI в. Электронный мониторинг практически вытесняет из

общества профессиональную преступность...

...20—30-е гг. XXI в. Искусственный интеллект достиг уровня развития человеческого мозга. С этого момента на Земле началось сосуществование двух разумных форм жизни, причём вторая эволюционировала со скоростью, недоступной биологической эволюции. К ближайшим звёздам отправились первые корабли, снабжённые системами искусственного интеллекта. Появились клонированные динозавры из генерированных компьютером ДНК.

...40—50-е гг. XXI в. Полностью усовершенствован автономный, регенерируемый, мобильный дом... «Универсальный репликатор», основанный на нанотехнологиях, может создать объект любой сложности при наличии сырья и информационной матрицы. Бриллианты и деликатесная еда могут быть сделаны в буквальном смысле слова из грязи. В результате за ненадобностью исчезают промышленность и сельское хозяйство, а вместе с ними и недавнее изобретение человеческой цивилизации — работа! Взрывное развитие искусств, развлечений, образования. Огромные территории планеты, которые больше не нужны для производства продовольствия, возвращаются в изначальное состояние, молодые люди могут дать волю своим агрессивным инстинктам, участвуя в большой охоте с самострелами...

Многие прогнозы уже сбылись: фирма LG выпустила стиральную машину, подбирающую оптимальный режим стирки и сообщающую, сколько времени она будет длится; микроволновые печи имеют встроенные программы приготовления блюд, при этом в нужное время оповещающие хозяйку, когда положить нужный продукт, а когда блюдо надо помешать; автомобильная компания Volvo уже выпускает автомобили со встроенной системой Volvo on call: при возникновении проблем водителю нужно лишь нажать нужную кнопку, и система сама соединит его с оператором, который автоматически получит координаты нахождения автомобиля, а в случае аварии со срабатыванием подушки

Кнопка вызова экстренной помощи на панели современного автомобиля.

безопасности система самостоятельно пошлёт оператору сигнал SOS, и тот вышлет на место аварии помощь. Этаже система может использоваться и как система сигнализации и как спутниковая система обнаружения при угоне автомобиля. Современные КПК и мобильные телефоны легко воспроизводят видео и изображают из себя «офис в кармане».

* * *

Пребен Мейер, руководитель отдела развития компании TeleDanmark, в одном из интервью прогнозирует быстрое распространение «думающей техники» в домах. Он убеждён, что в ближайшие 20—25 лет компьютеры практически полностью возьмут на себя заботу о жилище: стиральные машины самостоятельно будут определять нужный режим стирки, микроволновые печи — готовить обед, холодильники — заказывать продукты. Кроме того, автомобили смогут сами резервировать время в автомастерской, а тяжёлые электронные аппараты будут заменены лёгкими, как бумага, экранами на стенах квартир.

* * *

По прогнозам немецкого физика Харольда Группа, опросившего более 900 учёных, с XXI в. начнётся активное строительство домов, оснащённых солнечными батареями на крышах и фасадах. Прогресс дойдёт до такого уровня, когда всю работу по дому будут делать компьютеры, даже мусорное ведро само станет сортировать отходы.

Уже в первом десятилетии XXI в. операции с применением скальпеля безвозвратно исчезнут. Хирург-микроробот будет проникать во все органы

человека и производить необходимые манипуляции — удалять тромбы, вводить микродозы лекарства... Мини-компьютер заменит химика-эксперта, выдавая полную информацию о любом веществе. Спустя ещё десять лет Европа станет единой страной с единым правительством и единой компьютерной системой управления. Армия будет состоять в основном из компьютерщиков... А к концу второй декады на Земле появятся первые киборги — люди, сращённые с компьютером. Это вызовет не только переворот в науке и технике, но и новые нравственные проблемы, породит споры и глубокие разногласия в обществе.

* * *

Поражённый церебральным параличом английский учёный и мыслитель Стивен Хокинг, которого порой называют вторым Эйнштейном, в книге «Краткая история времени» пишет: «Человечеству необходимо улучшить умственный и физический уровень, чтобы справиться со всё более усложняющимся миром и решать новые проблемы, связанные, в частности, с космическим перелётами».

По мнению Хокинга, человеку необходимо становиться всё более и более совершенным, чтобы «бежать впереди электронных умников». «В данный момент компьютеры имеют преимущество в скорости, но у них нет интеллекта. Однако скорость и сложность компьютеров удваивается каждые 18 месяцев, и в будущем они смогут достичь уровня сложности мозга человека», — убеждён Стивен Хокинг. Также, по его прогнозам, в истории наступит новая эра — период планируемой эволюции, когда на эмбриональной стадии смогут определять и выбраковывать патологические отклонения в развитии человека, и благодаря этому люди будущего будут обладать отменным здоровьем и жить до 120 лет.

Стивен Хокинг.

* * *

По прогнозам японских учёных, в первом десятилетии нового века появится карманный компьютер, в корпусе которого будут соединены факс и видео. Компьютеры смогут мгновенно переводить на все языки мира не только тексты, но и разговорную речь. Благодаря компьютерам синтезируют искусственную кровь, будут побеждены СПИД и атеросклероз. В дальнейшем люди научатся использовать искусственные глаза и мозг.

* * *

По прогнозам журнала «Forbes ASAP», компьютер ближайшего будущего представляет собой устройство, через которое «будет работать всё, начиная от охранной сигнализации и

заканчивая холодильником». Кремний, который был основой для компьютерной индустрии, уже в 10-х гг. XXI в. перестанет удовлетворять нужды компьютерщиков из-за своих размеров, маленькой скорости и слишком большого выделения тепла при работе.

Уже сейчас сотрудники Станфордского университета разрабатывают оптоэлектронное устройство, которое использует скорость и пропускную способность оптических коммуникаций. В результате, по словам производителей новой техники, компьютер станет более дешёвым, более компактным, более производительным, менее энергоёмким. С его помощью можно будет управлять всеми приборами в доме, а на работе использовать в качестве огромного интерактивного экрана. Своего истинного хозяина компьютер будет узнавать, например по отпечатку пальца. Управлять таким компьютером легко: для этого достаточно просто разговаривать с ним. Для тех, кто привык к клавиатуре, она будет представлена в виртуальном виде — и нажимать кнопки можно будет пальцами прямо на экране. Память такого компьютера голографическая, она позволяет хранить сколь угодно большие объёмы информации. Скорость обмена данными между процессором и памятью сравнится со скоростью процессора.

* * *

По прогнозам Виктора Феллера, одного из талантливых футурологов, в ближайшее время произойдёт внедрение в медицину компьютерных технологий нового поколения. Вследствие этого произойдёт развитие биотехнологий, появятся искусственные органы: сердце, печень, почки, будут созданы компьютерные системы, вживляющиеся в организм человека для выявления и уничтожения там вирусов.

* * *

Участник проекта «Видение будущего» Экспертного центра Philips в Эйндховене (Нидерланды) Деррик Де Керкхоув говорит: «Я полагаю, в будущем мы будем иметь множество приборов, управляемых движением руки. Нет большей иллюзии власти, чем возникающая, когда вы делаете знак рукой и что-то немедленно происходит. Люди таковы. Распознавание с голоса существует и совершенствуется уже сегодня. Высококачественные системы распознавания будут устанавливаться в каждой машине, которую мы где-либо используем, даже на велосипеде. Я думаю, что в этом смысле и почерк сможет вновь занять подобающее место, констатировать факт вашего личного присутствия. Мне также кажется, что появятся

лучшие способы преобразования ручного письма в печатное».

По мнению этого футуролога, в скором будущем люди будут в основном путешествовать виртуально, ведь это намного дешевле и более доступно для многих, кто не имеет возможности осуществить такое путешествие физически. Кроме того, виртуальное путешествие может оказаться и более познавательным, так как всегда даёт нужное объяснение по интересующим вопросам.

Деррик Де Керкхоув также считает, что чем больше техника будет усложняться внутри, тем больше она должна упрощаться с внешней стороны. Людям почти не придётся ходить на работу, они будут посещать офис два или три раза в неделю, в остальное время можно выполнять задания дома. Оплату также станут высылать по Сети.

* * *

По мнению президента компании Microsoft Билла Гейтса, в будущем новые информационные технологии обогатят досуг, активизируют культурную жизнь, расширят доступ к разнообразной информации. Появившаяся возможность работать дома или в удалённых офисах ослабит нагрузку на городские структуры. Многие изделия будут выпускаться не в материальной форме, а в виде последовательности битов, в связи с этим снизится расход природных ресурсов. Новые технологии помогут обрести независимость, шире обмениваться опытом и покупать товары, подогнанные под запросы. Мощь цифровых технологий и универсальность их применения вызовут новые трудности в сохранении личных, коммерческих и государственных тайн. Многие профессии и целые отрасли индустрии будут забыты, но на их месте возникнут другие, ранее неизвестные. Произойдёт активнейшее развитие сети Интернет. Стоимость вычислительной техники и средств связи снизится настолько, что цена информации будет небольшой. Прибыль от рекламы позволит распространять многие материалы и совсем бесплатно (например, доступ к правительственной информации, консультативным медицинским служ-

бам, электронным доскам сообщений и некоторым учебным пособиям). У компаний, обслуживающих Сеть, появится стимул провести волоконно-оптические линии в удалённые регионы с высоким уровнем доходов. Электронные доски сообщений и другие сетевые форумы создадут превосходные условия для контакта «один на один» и по принципу «один со многими» или «многие со многими». Единомышленники смогут встречаться в компьютерных сетях, дебатировать, учреждать партии или объединения без особых усилий. Сеть сблизит или рассеет людей на миллионы сообществ. Информационная магистраль позволит всё делать по-новому: проводить свободное время, искать информацию, общаться друг с другом.

* * *

Развитие техники и информационных технологий, по мнению футурологов из американской консультационной компании Global Business Network, сильно отразится на жизни людей, освободив их от бытовых забот. Роботизация войдёт во все сферы деятельности человека. Эти полуразумные подвижные «домашние животные» будут готовить, стирать, убирать и т. д. Компьютеризация достигнет таких масштабов, что сделается основой технического прогресса. Персональные компьютеры станут миниатюрными, в них совместятся десятки функций, поэтому они будут выполнять обязанности советчиков, секретарей. Связь станет универсальной и дешёвой; транспорт — не только более комфортабельным, скоростным и экономичным, но и более компьютеризованным. Исчезнет такая профессия, как шофёр. Произойдёт слияние в одно целое ТВ, радио, магнитофона, пейджера, телефона. Кино и телевидение сумеют передавать запахи. Основными увлечениями населения станут спорт и компьютеры.

* * *

Прогноз английского политолога Джонатана Миллера в газете «Sunday Times» таков: в 2010 г. некая Кира Алисен, жительница Санкт-Петербурга, создаст компьютерную программу Infinity, которая позволит интегрировать все несовместимые между собой компьютеры. Лидерство в информатике навсегда перейдёт к России. Компания, созданная Кирой, станет номером один в мире.

* * *

Не так уж важно, насколько исследования, проведённые учёными, отражают реальную картину будущего. Цель исследований — своевременное распознавание технологических тенденций, которые будут играть особенно важную роль в развитии человеческого общества и сохранении природы. Вот что сказал об этом Деррик Де Керкхоув: «Будущее невозможно предвидеть, но его реально экстраполировать. Наша система прогнозов проста: мы мысленно продолжаем те направления, в которых движется сегодня человечество, и делаем вывод, в какой срок оно достигнет той или иной точки».

СОДЕРЖАНИЕ

ТЕОРИЯ ПРОГРАММИРОВАНИЯ

ЯЗЫКИ ПРОГРАММИРОВАНИЯ

ПРОГРАММИСТЫ И ПРОГРАММИРОВАНИЕ

ХРАНЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ

ХРАНЕНИЕ ИНФОРМАЦИИ

ОПЕРАЦИОННАЯ СИСТЕМА КОМПЬЮТЕРА

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ