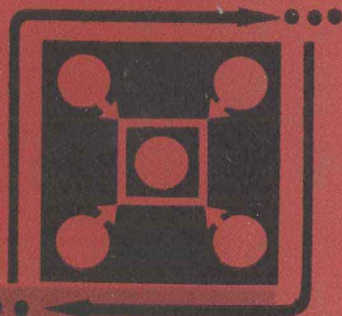


*В. П. Цымбал
С. П. Куценко*



ИНФОРМА- ЦИОННЫЕ СИСТЕМЫ И СТРУКТУРЫ ДАННЫХ

В. П. Цымбал, С.

ИНФОРМА- ЦИОННЫЕ СИСТЕМЫ И СТРУКТУРЫ ДАНЫХ

*Допущено Министерством высшего
и среднего специального образования СССР
в качестве учебного пособия
для студентов вузов,
обучающихся по специальности
«Организация
механизированной обработки
экономической информации»*

КИЕВ
ГОЛОВНОЕ ИЗДАТЕЛЬСТВО
ИЗДАТЕЛЬСКОГО ОБЪЕДИНЕНИЯ
«ВИЩА ШКОЛА»
1980

ББК 32.973-я 73
6Ф7.3
Ц94

УДК 681.518.3 (07)

Цымбал В. П., Куценко С. П. Информационные системы и структуры данных: Учеб. пособие для вузов. — Киев: Вища школа. Головное изд-во, 1980. — 240 с. — 30502. 1502000000.

Описаны принципы обработки данных в современных ЭВМ, способы организации данных в памяти ЭВМ и на различных физических носителях. Подробно и с большим количеством примеров представлены многочисленные способы сжатия информации.

Информационные системы рассматриваются как в общетеоретическом плане, так и в конкретном приложении к решению задач информационного обеспечения. Особое внимание уделено информационно-поисковым системам и языкам, организации баз данных, сетей информационных систем и индустрии информации. Изложены вопросы оценки экономической эффективности информационных систем.

Предназначено для студентов вузов, обучающихся по специальностям «Механизированная обработка экономической информации» и «Экономическая кибернетика». Может быть использовано также студентами университетов, изучающими вопросы, связанные со способами организации памяти ЭВМ, созданием информационных систем и сетей.

Табл. 15. Ил. 76. Список лит.: 42 назв.

Редакция литературы по кибернетике, электронике и энергетике

ПЕРЕЧЕНЬ ПРИНЯТЫХ СОКРАЩЕНИЙ

АИС — автоматизированная информационная система;
АИЦ — автоматизированный информационный центр;
АП — абонентский пункт;
АПД — аппаратура передачи данных;
АСОД — автоматизированная система обработки данных;
АС НТИ — автоматизированная система научно-технической информации;
АСУ — автоматизированная система управления;
АЦПУ — алфавитно-цифровое печатающее устройство;
БД — база данных;
БУ — блок управления;
ВИК — Всесоюзная информационная классификация;
ВИНИТИ — Всесоюзный институт научно-технической информации;
ВНИИКИ — Всесоюзный научно-исследовательский институт технической информации, классификации и кодирования Государственного комитета стандартов Совета Министров СССР;
ВЗУ — внешнее запоминающее устройство;
ВНТИЦ — Всесоюзный научно-технический информационный центр Государственного комитета Совета Министров СССР по науке и технике;
ВЦ — вычислительный центр;
ВУ — внешнее устройство;
ГАС НТИ — Государственная автоматизированная система научно-технической информации;
ГПНТБ — Государственная публичная научно-техническая библиотека СССР;
ГРНТБ — Государственная республиканская научно-техническая библиотека;
ГСВЦ — Государственная сеть вычислительных центров;
ЕС ЭВМ — Единая серия электронно-вычислительных машин;
ЗУ — запоминающее устройство;
ИПЯ — информационно-поисковый язык;
ИПС — информационно-поисковая система;
ИЛ — информационный листок;
ИРИ — избирательное распределение информации;
ИИС — интегрированная информационная система;
КОП — код операции;
КС — канал связи;
КТС — комплекс технических средств;
ЛО — лингвистическое обеспечение;
М — модем (модулятор-демодулятор);
МКИ — международная классификация изобретений;
МАРС НТИ — межотраслевая автоматизированная система научно-технической информации;
МБ — магнитный барабан;
МЛ — магнитная лента;
МД — магнитный диск;
МО — математическое обеспечение;

МПД — мультиплексор передачи данных;
 МФ — микрофиша;
 МТЦ НТИ — межотраслевой территориальный центр НТИ;
 НД — набор данных;
 НМБ — накопитель на магнитном барабане;
 НМД — накопитель на магнитном диске;
 НМЛ — накопитель на магнитной ленте;
 НТИ — научно-техническая информация;
 ОГСПД — Общегосударственная сеть передачи данных;
 ОЗУ — оперативное запоминающее устройство;
 ОНТИ — отдел научно-технической информации;
 ПОД — поисковый образ документа;
 ПОЗ — поисковый образ запроса;
 ППП — пакет прикладных программ;
 ПМФ — предмашинный формат;
 РАБД — распределенный автоматизированный банк данных;
 РБ — регистр базы;
 САЦ НТИ — сеть автоматизированных центров научно-технической информации;
 СВЦ — сеть вычислительных центров;
 СИФ — справочно-информационный фонд;
 СМО — система математического обеспечения;
 СПД — система передачи данных;
 ТЧ — тональная частота;
 УВВ — устройство ввода-вывода;
 УДК — универсальная десятичная классификация;
 УПДЛ — устройство подготовки данных на перфоленте;
 УП — указатель перехода;
 УПДК — устройство подготовки данных на перфокартах;
 УС — указатель списка (строки);
 ЦНИИПИ — Центральный научно-исследовательский институт патентной информации и технико-экономических исследований Государственного комитета по делам изобретений и открытий Совета Министров СССР;
 ЦНТИ — Центр научно-технической информации;
 ЦБТИ — Центральное бюро научно-технической информации;
 ЭС — элемент списка (строки);
 ЭПС — элемент подписка.

ПРЕДИСЛОВИЕ

Одними из важнейших задач, стоящих перед народным хозяйством нашей страны, являются обеспечение дальнейшего развития и повышение эффективности автоматизированных систем управления и вычислительных центров, объединение их в единую общегосударственную систему сбора и обработки информации для учета, планирования и управления. Для их успешного решения необходимо, в частности, чтобы будущие специалисты сферы учета, планирования и управления были знакомы с теорией и практикой построения информационных систем. С этой целью в программу вузов страны введены специальности «Автоматизированные системы управления», «Экономическая кибернетика», «Механизированная обработка экономической информации». С вводом этих специальностей появился и ряд новых дисциплин, одной из которых является «Информационные системы и структуры данных».

Информационные системы и структуры данных (ИСИСД) — дисциплина, предмет изучения которой составляют формы рациональной организации, представления, поиска и передачи данных внутри и вне ЭВМ, а также системы, в которых процесс управления основан на информации, полученной в результате обработки данных, циркулирующих в системе.

Эта дисциплина возникла на стыке экономики и кибернетики. На различных этапах построения ИСИСД имеют связь с рядом смежных с экономикой и кибернетикой наук и дисциплин (рис. 1). Основными из них являются: исследование старой информационной системы, построение математической модели, составление алгоритмов и программ, проигрывание модели, построение новой информационной системы.

Вопросы общей теории информационных систем, информационного поиска, оценки информационных потоков, структуры, организации памяти ЭВМ и другие не всегда вытекают один из другого, хотя каждый из них связан с задачами, объединяющими такое глобальное понятие, как «информационная система». Разделение материалов книги на две части предопределено тем, что структуры данных (вопрос сравнительно узкий) требуют углубления до взаимосвязи отдельных байтов в памяти ЭВМ. Информационные системы, наоборот, требуют

широты охвата материала, обобщений, ознакомления с существующими информационными системами и различиями между ними.

Главу 14 и введение написал д-р экон. наук С. П. Куденко, главы 1—13 и 15 — канд. техн. наук В. П. Цымбал.

Авторы выражают благодарность О. Н. Алексееву, Г. Я. Береговенко, А. М. Коптеву и Б. В. Тимонькину за помощь в подборе материала для написания некоторых глав.

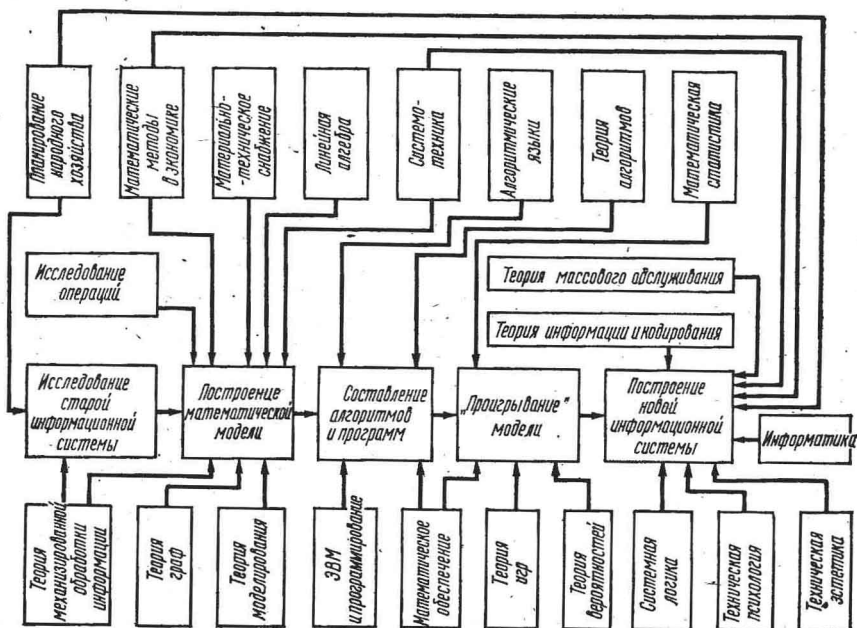


Рис. 1. Связь дисциплины «Информационные системы и структуры данных» с другими дисциплинами.

Особую признательность авторы выражают профессорам С. И. Волкову и Г. В. Лавинскому за рецензирование, а также за ценные замечания и советы, которые способствовали значительному улучшению данного пособия.

Отзывы и пожелания просим направлять по адресу: 252054, Киев-54, ул. Гоголевская, 7, Головное издательство издательского объединения «Вища школа».

ВВЕДЕНИЕ

В развитии производства большую роль играют, как известно, численность работающих и капитальные вложения. Однако дальнейшее расширение масштабов производства не может осуществляться только за счет этих факторов. Часто вводимые в производство новое оборудование, приборы, инструменты не всегда базируются на прогрессивной технологии. Поэтому данный элемент производительных сил не может быть определяющим. В этой связи становится очевидным, что необходимо обратить внимание на другие элементы производства, которые недостаточно оценивались ранее и неэффективно использовались на более ранних стадиях развития народного хозяйства.

К традиционным элементам производства обычно относят средства производства, при помощи которых создаются материальные блага, а также людей, приводящих в движение эти средства и осуществляющих производство материальных благ. Кроме того, в современном производстве все большую роль играют и другие элементы, которые не вписываются в характеристики ни орудий, ни предметов труда. К ним относят энергию, технологию, информацию и технику научных исследований.

В современных условиях информация приобретает особое значение в системе производительных сил. Она становится одним из главных ресурсов производства наравне с сырьем, машинами, энергией, рабочей силой. Современный период развития общественного производства — это период, когда уровень организованности общества и его потенциальные возможности определяются уже не только энерго- и фондовооруженностью, но и информационной вооруженностью (ресурсами и потенциалами).

Накопленная в результате непрерывной работы, постоянного познания и исследовательской деятельности информация позволяет все более полно познавать глубины законов природы, изучать их, осваивать и использовать для совершенствования средств и условий труда, что неизбежно приводит к повышению его производительности. Наряду с этим информация выступает и в качестве своеобразного средства управления. Таким образом, информация выступает, с одной стороны, как непереносимое условие функционирования всех элементов

производства и предприятий в целом, а с другой, — становится предметом и продуктом труда отдельной категории работников.

В этой связи информация, как и орудие, и предмет труда, становится неотъемлемым и специфическим элементом производства. Если энергия, предметы и орудия труда являются основными условиями процесса производства, то информация оказывается наиболее активным, организующим элементом, приводящим производство в движение и направляющим его. Еще более организующее значение приобретает информация в масштабах всего народного хозяйства. В данном случае она является средством, при помощи которого все отрасли народного хозяйства связываются в единую систему.

Если рассматривать информацию с точки зрения ее значимости в развитии общественного производства и организации систем ее сбора и переработки, то в первом приближении весь объем информации народного хозяйства может быть подразделен на три характерных вида:

1) научно-техническая и другая информация, сосредоточенная в специальных библиотеках и в различных коллективах людей;

2) опосредованная информация, т. е. информация, воплощенная в средствах и предметах труда, методах организации и управления, технологических процессах;

3) информация, характеризующая развитие народного хозяйства и используемая для координации производственно-хозяйственной деятельности отдельных его звеньев (этот вид информации называют еще управленческой, экономической, производственно-экономической и т. д.).

Научно-техническая информация содержит все необходимые сведения о средствах и предметах труда, технологических процессах, организации производства и все то, без чего немислимо современное производство материальных благ. Роль научно-технической информации в развитии народного хозяйства трудно переоценить.

Одной из основных особенностей такого рода информации является ее преемственность и интернациональный характер. Поэтому наука как непосредственная производительная сила строит свое здание на фундаменте информации, накопленной предыдущими поколениями. Преемственность и интернациональный характер научно-технической информации способствуют ускорению темпов развития науки. «Всеобщим трудом является всякий научный труд, всякое открытие, всякое изобретение. Он обуславливается частью кооперацией современников, частью использованием труда предшественников»¹.

Преемственность научно-технической информации неразрывно связана с ее интернациональным характером. Поэтому, как правило, научное открытие и изобретение представляют собой результаты дея-

¹ Маркс К., Энгельс Ф. Соч. 2-е изд., т. 25, ч. 1, с. 116.

тельности ученых разных стран мира, что может иллюстрироваться развитием таких наук, как кибернетика, биология, физическая химия, бионика, инженерная психология, космическая медицина и многих других.

Непрерывное накопление научно-технической информации является одним из основных факторов ускорения развития науки. В статье «Наброски критики политической экономии», опубликованной в 1844 г., Ф. Энгельс отмечал: «...наука движется вперед пропорционально массе знаний, унаследованных ею от предшествующего поколения...»¹.

Подчеркивая важность научно-информационной деятельности как специфической разновидности научной работы, В. И. Ленин писал: «Надо внимательно следить за всей соответствующей литературой на всех языках, переводя или, по крайней мере, реферируя сколько-нибудь ценное...»².

Совершенствование системы научно-технической информации стало одним из главных факторов развития науки в нашей стране. В частности, XXV съезд КПСС поставил задачу: «Улучшать деятельность научно-технических обществ, совершенствовать систему научно-технической информации.

Обеспечивать комплексное планирование, финансирование и стимулирование научно-технического прогресса»³.

Значительную часть накопленной человечеством информации представляет *опосредованная информация*. Причем накопление информации вне человека приобретает все большие темпы и масштабы.

Развитие и совершенствование производства непрерывно связано с конкретизацией информации в средствах и предметах труда. Опосредованная информация создается в процессе творческой деятельности участников трудового процесса и концентрируется в конечном счете в средствах и предметах труда, в новых изобретениях, рационализаторских предложениях и т. п. В данном случае орудия и предметы труда как бы аккумулируют, втягивая в себя, определенный запас лучшей и ценной информации и хранят ее в себе.

Если оценивать количество информации, заключенное в орудиях труда, по количеству элементов, из которых они состоят, то нетрудно заметить, что с развитием общества они усложняются. С ростом сложности орудий труда увеличивается их информационное содержание, что в конечном счете ведет к расширению разнообразия связей и отношений между их элементами. Количество информации в производстве растет по меньшей мере пропорционально квадрату промышленного потенциала.

¹ Маркс К., Энгельс Ф. Соч. 2-е изд., т. 1, с. 568.

² Ленин В. И. — Полн. собр. соч., т. 45, с. 25.

³ Материалы XXV съезда КПСС. М.: Политиздат, 1976, с. 171.

Таким образом, специфической особенностью опосредованной информации является то, что она, возникая вначале абстрактно, конкретизируется в новых средствах производства, рационализаторских предложениях, изобретениях, нововведениях и, увеличивая информационное содержание промышленного потенциала, стимулирует рост управленческой информации.

Управленческая информация необходима для организации и управления производством, выбора и обоснования наиболее экономичных вариантов производства материальных благ, координации деятельности всего народного хозяйства с целью повышения эффективности общественного производства.

Рост масштабов производства, всестороннее его развитие и совершенствование, непрерывный процесс прогрессирующего разделения труда и специализации приводят к значительному увеличению объемов управленческой информации. В настоящее время ежемесячный объем этой информации только в промышленности составляет около 10 млрд. документо-строк, или свыше 1200 млрд. байт. Сбором и обработкой управленческой информации занимается свыше 15 млн. человек.

Вопросами управления народным хозяйством и отдельными его звеньями занимается специфическая отрасль народного хозяйства — «Управление». Кроме того, управлением отдельными предприятиями и организациями занимаются многочисленные специальные службы, которые не относятся к отрасли «Управление», а включаются в состав тех отраслей, к которым относится то или иное предприятие или организация. На предприятиях и в организациях сбором и обработкой информации занимается также многочисленный персонал, не относящийся к управленческому аппарату. Так, исследования и анализ рабочего времени работников различных профессий машиностроительных предприятий показали, что многие из них расходуют значительную часть своего рабочего времени на операции, связанные со сбором и обработкой информации. Только на оформление различного рода первичных документов расходуется: мастерами — 50, контролерами — 45, распределителями работ — 60, технологами — 80, нормировщиками — 50% рабочего времени. Таким образом, если речь идет о численности персонала, занятого управленческими работами, то необходимо иметь в виду и ту категорию лиц, которая занята организацией и обслуживанием трудовых процессов непосредственно на рабочих местах.

В течение ряда лет многочисленная армия работников управленческого аппарата и другого персонала, занимающегося документированием данных о результатах трудовых процессов на конкретных рабочих местах, осуществляла сбор и обработку информации, в основном, вручную или с помощью примитивных механических средств (конторских счетов, арифмометров и т. д.).

Следовательно, увеличение объема информации приводило к зна-

чительному росту управленческого персонала. Однако на данном этапе развития общественного производства даже безграничное увеличение численности аппарата управления не может решить проблему рациональной организации управления народным хозяйством в целом и отдельными его звеньями. Это связано с тем, что труд каждого работника, обрабатывающего информацию, нуждается в руководстве и координации с трудом других. Поэтому имеется некоторый критический предел численности управленческого персонала учреждения, после превышения которого полезная мощность учреждения по обработке информации уже не увеличивается, и учреждение начинает работать само на себя. Эти обстоятельства в значительной степени сказались на необходимости поиска новых, более совершенных путей сбора и обработки информации.

В нашей стране осуществляется ряд важных мер, направленных на совершенствование методов и техники управления с целью приведения их в соответствие с современным уровнем развития производительных сил. Этому уделено большое внимание в постановлении ЦК КПСС и Совета Министров СССР «Об улучшении планирования и усилении воздействия хозяйственного механизма на повышение эффективности производства и качества работы» (июль 1979 г.).

Одним из важных вопросов является внедрение вычислительных машин и других технических средств сбора и обработки информации в сферу управления. Индустриализация сбора и обработки информации в течение многих лет осуществлялась путем создания специфических информационных систем и предприятий по переработке информации.

Число информационных и информационно-вычислительных сетей, систем и предприятий индустрии информации растет буквально с каждым днем, растут капиталовложения, увеличивается количество публикаций, посвященных теории и практике информационного поиска, исследованию и обработке информационных потоков, всевозможному преобразованию информации.

Темпы роста индустрии информации, в частности автоматизированных информационных систем, ставят перед народным хозяйством задачу подготовки специалистов, способных решать многочисленные проблемы, связанные с индустриализацией информационных процессов.

Часть I

СТРУКТУРЫ ДАННЫХ

Глава 1

ОБЩИЕ ПОНЯТИЯ, ОПРЕДЕЛЕНИЯ, ТЕРМИНОЛОГИЯ

Информация, обрабатываемая вычислительными машинами, представляется через организованное некоторым образом множество элементов. Эти элементы представлены конкретным набором качественных признаков, которые могут нести в себе определенную смысловую информацию. Эта информация доступна тому, кто знает структурные правила, которым подчиняются качественные признаки, организованные по определенному закону в конкретные данные. Так, качественные признаки «г», «д», «о», «р» могут быть организованы в слова «огород» и «город». Если «город» читать справа налево, то получится слово «дорог».

Если в качестве структурного закона построения чисел из алфавита (0, 1, 2, ..., 9) положить, что крайняя справа цифра означает единицы, следующая — десятки и так далее, то любые комбинации элементов данного алфавита могут быть однозначно декодированы, так как известна структура данных, составленных из этого алфавита.

Данные, поступающие для обработки в вычислительную машину, в свою очередь, могут быть представлены в форме различных структур, организованных по определенным законам: в виде линейного списка элементов, в виде таблицы, матрицы, в виде древовидной, иерархической или более сложной мультисвязной структуры.

Структура данных — это множество элементов с заданной на нем системой отношений. Относительно ЭВМ структура данных представляет собой способ организации данных в ее памяти.

Рациональное использование ЭВМ во многом зависит от правильного понимания структурных зависимостей между данными внутри и вне машины.

Элементом данных вне ЭВМ является символ — структурно неделимая единица данных, которая обычно представлена одним качественным признаком (например, буквой или цифрой). Символы используются для образования семантически и синтаксически значимых понятий.

Символы, объединенные в совокупность, имеющую самостоятельное логическое значение, называются *словом*. Слова представляют собой неделимые осмысленные понятия.

Совокупность данных о некотором объекте или явлении из множества интересующих нас объектов или явлений, которая содержит законченную информацию, называется *записью*. Запись может как описывать объекты системы и их взаимодействия, так и быть самостоятельным документом.

Организованная система семантически связанных записей, обладающих общей структурой, образует *массив* данных. Обычно массивы представляют собой информационные структуры, состоящие из однородных записей, упорядоченных по определенным качественным признакам. Если обрабатываемые записи неоднородны и неупорядочены, термин массив лучше не употреблять. Более грамотно назвать такую группу записей *набором данных*, а если она поименована, то *файлом*.

Наименьшим неделимым элементом данных внутри ЭВМ является один двоичный разряд — *бит*. Значение двоичного разряда «0» или «1» фиксируется одним из двух возможных устойчивых состояний дискретных элементов, применяемых в данном типе ЭВМ: открыт — закрыт триод, намагничен — ненамагничен ферритовый сердечник и так далее (8 бит составляют 1 *байт*, 2000 байтов — 1 *страницу*, 500 страниц — 1 *том*)¹.

В практике работы с внешними запоминающими устройствами том — это отдельная физическая единица любого ВЗУ.

Томом может быть катушка магнитной ленты, пакет дисков, рулон перфоленты, колода вводимых за один раз перфокарт. В одном томе может быть несколько файлов и наоборот — один файл может размещаться в нескольких томах. Поэтому понятие «том» идентифицирует внешний носитель, а не содержащиеся на нем данные. Сменные ленты или пакеты дисков называют *сменными томами*. Емкость сменных томов зависит от вида физического носителя. Размер страниц является параметром системы и не зависит от содержания и длины записей, составленных программистом, т. е. данные логически не зависят от размера страниц, который является фиксированной величиной только в пределах данной системы. Что касается сменных пакетов дисков, то уже в настоящее время их емкость меняется от 7 до 100 Мбайт.

Для определения объемов ОЗУ пользуются еще одной структурной единицей *K* (кэ), равной $2^{10} = 1024$ ячейкам. ЗУ средних размеров содержит 64...256 *K*.

При изучении и описании наборов данных различают физический и логический уровни организации данных. Логический уровень отражает способ организации данных в представлении программиста либо пользователя, а физический — способ записи и хранения данных на физических носителях с помощью ЭВМ.

Сообщения, существующие на логическом уровне, становятся объектом обработки при вводе их в машину. Взаимосвязанные объекты

¹ В зависимости от типов решаемых задач размеры страниц могут быть от 256 до 4096 байтов, поэтому приведенное деление носит условный характер.

образуют структуру данного типа. Компонентами структуры являются поля. Если для данного множества объектов существует одна (или несколько) операция, которая может быть выполнена для всех объектов, входящих в множество, то относительно этой операции они принадлежат к одному *виду*.

Вид объекта идентифицируется в зависимости от используемого алгоритмического языка. Примеры идентификаторов вида:

ц е л — целочисленный объект;

в е щ — вещественно-числовой объект;

с т р о к — строка символов;

DECIMAL — объект представлен десятичным числом;

BINARY — объект представлен двоичным числом.

При хранении объектов в памяти ЭВМ часто возникает необходимость в неоднократном обращении к ним. В таком случае целесообразно присваивать объекту имя (метку), а обращение осуществлять с помощью ссылок на это имя. Имена конкретно представляются адресами ячеек памяти и являются идентификаторами места. Операция, устанавливающая ссылку на именуемый объект, называется *присваиванием*.

Объекты, к которым по той или иной причине нельзя обратиться с помощью идентификатора, зря занимают место в памяти и являются информационным мусором. Операция «Ведение файла» предусматривает уборку мусора, которая сопровождается пометкой освобождающихся мест, перераспределением памяти и исправлением ссылок.

Абстрактно представленные структурные элементы позволяют изучать способы организации данных в памяти ЭВМ независимо ни от организации набора данных на логическом уровне, ни от конкретного физического носителя.

Содержанием структурного элемента, независимо от уровня организации, могут быть не только данные в общепринятом смысле слова, но и программы. Хранение программ в памяти ЭВМ значительно ускоряет процесс обработки информации, так как машина может приступать к выполнению очередной команды, не дожидаясь выдачи результата выполнения предыдущей, и так до тех пор, пока не будут обработаны все данные, поступающие на вход машины.

Структуры данных различаются характером взаимосвязи структурных элементов в памяти ЭВМ в процессе решения той или иной задачи.

В большинстве структур эта взаимосвязь задается с помощью адресации структурных элементов.

На логическом уровне описания наборов данных (НД) различают следующие структурные элементы.

Поле — наименьший идентифицируемый элемент. Оно может представлять собой произвольную группу битов или байтов, которым присвоен идентификатор, позволяющий отыскать это поле в памяти ЭВМ.

Приведенное формальное определение не следует рассматривать буквально, так как записью может быть и отдельный знак, и целый документ.

Запись — несколько полей, объединенных общим названием либо смысловым содержанием.

Логическая запись, или сегмент (гл. 5), — порция данных, которая передается одним оператором при обработке данных или обмене между различными уровнями памяти ЭВМ.

Группа записей, или агрегат данных, — совокупность элементов данных, рассматриваемых как единое целое и имеющих общий идентификатор.

Атрибуты, используемые в ЭВМ для идентификации полей, записей и агрегатов данных, называют ключами.

Дата		
Год	Месяц	День
← Поле →	← Поле →	← Поле →
← Запись →		

Наименование записи «Дата» может служить одновременно ключом этой записи. При поиске выходят на группу записей с общим ключом этой записи «ДАТА», а затем осуществляют сравнение по полям.

Массив на логическом уровне представляет собой упорядоченную совокупность логически однородных записей. Массивы и файлы могут быть представлены как в виде списков, так и в виде таблиц.

Файл — поименованная совокупность логических записей. Файл может состоять из одного или нескольких массивов. Упорядоченность записей — необязательное условие для файла.

Справочник квартирных телефонов

Ивакин Ю. И. 72-17-86
Иваненко Б. В. 76-86-72
Иванов И. И. 68-23-11
Ивченко А. В. 21-72-00
Ивченко В. А. 29-98-25

Ералаш

Иваненко Б. В. 76-86-72
Ивакин Ю. И. 72-17-86
Сукно красное — 28 р.
Сатин синий — 14 р.
Ивченко А. В. 21-72-00

В приведенных выше примерах фрагмент телефонного справочника будет файлом в том случае, если название «Справочник квартирных телефонов» реально существует в программе как идентификатор этого набора данных. Произвольный набор записей, объединенный под именем «ералаш», также может называться файлом, если «ЕРАЛАШ» — реально существующий в программе идентификатор этого произвольного набора данных, который никак не может быть назван массивом, так как записи в нем и неоднородны, и неупорядочены.

Наборы данных могут быть представлены как в виде списков (только что рассмотренные примеры), так и в виде таблиц.

Таблицей обычно называют такую организацию набора данных, при которой каждый элемент однозначно определен одним или несколькими аргументами. Говорят еще, что таблица — это набор *кортежей*, а сам кортеж — это строка таблицы. Набор значений элементов данных одного типа, т. е. один столбец таблицы, называют *доменом*.

Таблица 1

Величины					Шопка массива
Название и порядок расположения величин					
Ключ 1	Ключ 2	Ключ 3	...	Ключ n	Кортеж
учетный номер	вид материи	цвет		цена за 1 м. руб.	
21011	Сукно	011		28	
21012	Сатин	100		14	
21013	Полотно	001		15	
21014	Штпель	100		9	
Домен					Тело массива

Записи, организованные в таблице, обычно состоят из величин и их названия, так как величина 14 (табл. 1) может означать и размер, и номер, и вес, и т. д.

Величины представляют собой содержание обрабатываемого массива и могут быть представлены цифрами (ключ 1), буквами (ключ 2), кодовыми комбинациями (ключ 3, табл. 2). Названия величин выносят в шапку массива, а в записях оставляют их значения. Шапка массива указывает порядок представления данных в массиве. Множество всех названий величин составляет *номенклатуру величин*, а множество записей — *номенклатуру записей*. Массивы данных, представленных на бумаге, нас интересуют как объекты преобразования при решении задач обработки данных в информационной системе.

Таблица 2

Код цвета	Цвет
001	Белый
010	Черный
011	Красный
100	Синий
101	Зеленый
110	Желтый
111	Коричневый

Значения величин с точки зрения их важности при решении конкретной задачи могут быть первостепенными, второстепенными и т. д. При перевозке важнее знать массу и габариты детали, а при комплектации — ее тип. Величины, которые при решении конкретной задачи

играют главную роль, называются *ключевыми*. Ключевым признакам присваивают идентификаторы-ключи, позволяют выделить определенные записи в НД (например, название материала позволяет различить