

В.С. ВЕРНИКОВ

ИНФОРМАТИКА

ЭНЦИКЛОПЕДИЧЕСКИЙ
СЛОВАРЬ-
СПРАВОЧНИК



Воройский Ф.С.

**Информатика.
Энциклопедический
словарь-справочник:
введение в
современные
информационные и
телекоммуникационные
технологии в
терминах и
фактах.**



МОСКВА
ФИЗМАТЛИТ ®

УДК 802.0-323.2-82:002+002.5(038)

ББК 39.97я2

В 75



Издание осуществлено при поддержке
Российского фонда фундаментальных
исследований по проекту 05-07-950040

Воройский Ф. С. **Информатика. Энциклопедический словарь-справочник: введение в современные информационные и телекоммуникационные технологии в терминах и фактах.** — М.: ФИЗМАТЛИТ, 2006. — 768 с. — ISBN 5-9221-0717-8.

Энциклопедический словарь-справочник содержит более 18 тыс. русско- и англоязычных терминов, тематически систематизированных по следующим крупным разделам: I. Основы информационной технологии; II. Автоматизация информационных процессов и автоматизированные системы (АС); III. Техническое обеспечение АС; IV. Программное обеспечение АС; V. Мультимедиа, гипермедиа, виртуальная реальность, машинное зрение; VI. Сетевые технологии обработки и передачи данных; VII. Компьютерный и сетевой сленг; VIII. Пиктограммы, использующиеся в электронной почте; IX. Сокращения слов и выражений, использующиеся в интернете.

Словарные статьи носят расширенный характер и включают в себя справочные данные об объектах описания, а также ссылки на первичные документальные источники для более полного ознакомления с ними заинтересованных в этом пользователей.

Структура и содержание словаря позволяют использовать его для систематизированного изучения материалов по интересующим читателя тематическим разделам и подразделам, производить предварительную проработку решений, связанных с проектированием разнородных автоматизированных информационных и телекоммуникационных систем, а также готовить на его основе учебно-методические, обзорные, справочные и др. документы.

Словарь ориентирован на широкий круг пользователей, профессиональная деятельность или интересы которых связаны с современными информационными технологиями.

ISBN 5-9221-0717-8

© ФИЗМАТЛИТ, 2006

© Ф. С. Воройский, 2006

СОДЕРЖАНИЕ

Предисловие к энциклопедическому изданию словаря	7
Предисловие к третьему изданию словаря, о словаре-справочнике и его авторе	9
От автора	11
О пользовании словарем	13
I. Основы информационной технологии	15
1.1. Данные, информация, знания, логика	15
1.2. Информационные ресурсы, теория информации, информатика	19
1.3. Носители информации, документы, документация, издания	22
1.4. Принципы структурированного представления документов и данных	27
1.4.1. Информационные элементы и их виды	27
1.4.2. Запись, файл, массив, ключ	30
1.4.3. Структуры, модели данных и связанные с ними термины	34
1.4.4. Формат, поле данных и связанные с ними термины	45
1.5. Информационная технология	49
1.5.1. Общие понятия и термины	49
1.5.2. Обработка и переработка документов и данных	52
1.5.3. Ввод документов и данных в ЭВМ	58
1.5.4. Информационный поиск — общие понятия и термины	63
1.5.5. Индексирование, поисковый образ документов и запросов	66
1.6. Безопасность информационной технологии	74
1.6.1. Общие понятия и термины	74
1.6.2. Кодирование и декодирование документов и данных	83
1.6.3. Криптология и связанные с нею понятия	87
II. Автоматизация информационных процессов и автоматизированные информационные системы	93
2.1. Общие понятия и термины	93
2.2. Автоматизация информационных и библиотечных процессов	95
2.2.1. Термины, связанные с автоматизацией	95
2.3. Автоматизированные системы	98
2.3.1. Общие понятия и термины	98
2.3.2. Функционально-ориентированные автоматизированные системы	106
2.4. Лингвистическое и информационное обеспечение автоматизированных систем	117
2.4.1. Лингвистическое обеспечение — общие понятия и термины	117
2.4.2. Информационно-поисковые языки и словарные средства АИС	119
2.4.3. Метаданные и форматы АИС	128
2.4.4. Информационное обеспечение АИС	147

2.5. Персонал и пользователи автоматизированных систем	153
2.5.1. Разработчики и персонал АИС	153
2.5.2. Пользователи АИС	157
2.5.3. Сертификация специалистов в АИС	159
2.6. Процессы создания и эксплуатации автоматизированных систем	162
2.6.1. Проектирование автоматизированных систем	162
2.6.2. Жизненный цикл АИС и системная интеграция	165
III. Техническое обеспечение автоматизированных систем	169
3.1. ЭВМ, их виды и общая классификация	169
3.2. Архитектура, конфигурация, платформа ЭВМ	175
3.3. Персональные ЭВМ (ПК)	178
3.4. Портативные ПК и автономные цифровые устройства разного назначения	185
3.4.1. Виды портативных ПК	185
3.4.2. Воспроизводящие и записывающие цифровые устройства	188
3.5. Системный блок и элементы его конструкции	191
3.5.1. Процессоры, их виды и связанные с ними термины	192
3.5.2. Память ЭВМ — понятия и термины	202
3.5.3. Функциональные устройства памяти ЭВМ	208
3.5.4. Адаптеры, интерфейсы и связанные с ними термины	216
3.5.5. Платы, порты, шины, гнезда	224
3.6. Периферийные (внешние) устройства ЭВМ	233
3.6.1. Внешняя память ЭВМ, накопители и связанные с ними термины	233
3.6.2. Компакт-диски и связанные с ними термины	251
3.6.3. Устройства ввода данных, манипуляторы	260
3.6.4. Устройства вывода данных	271
3.6.5. Модемы, шифраторы, источники питания	286
3.7. PC-карты	289
3.8. Микроэлектронная база ЭВМ	294
3.9. Оптикоэлектронные устройства	299
IV. Программное обеспечение автоматизированных систем	303
4.1. Алгоритмы, программы, программирование	303
4.1.1. Общие понятия и термины	303
4.1.2. Языки программирования	307
4.1.3. Связанные с программированием термины	319
4.2. Общее программное обеспечение	327
4.2.1. Операционные системы	328
4.2.2. Сервисные средства общего программного обеспечения	338
4.3. Прикладное программное обеспечение автоматизированных систем	339
4.3.1. Общие понятия и термины	339
4.3.2. Прикладные программы	342
4.3.3. Компьютерные вирусы и антивирусы	346
4.4. Термины, связанные с работой программных средств	350
4.4.1. Некоторые общие понятия и термины	350

4.4.2. Архивация, сжатие-восстановление записей данных	352
4.4.3. Доступ, адрес и связанные с ними термины	364
V. Мультимедиа, гипермедиа, виртуальная реальность, машинное зрение	372
5.1. Системы мультимедиа и связанные с ними термины	372
5.2. Средства обеспечения музыкального и речевого сопровождения	375
5.2.1. Общие понятия и термины	375
5.2.2. Звуковые файлы, их стандарты и форматы	380
5.3. Машинная (компьютерная) графика	389
5.3.1. Общие понятия и термины	389
5.3.2. Графические файлы и их форматы	392
5.3.3. Технология машинной графики	400
5.4. Компьютерное видео, цифровое телевидение и анимация	408
5.4.1. Общие понятия и термины	408
5.4.2. Технология видео	412
5.4.3. Технология анимации	416
5.4.4. Цифровое телевидение	420
5.5. Виртуальная реальность, параллельные миры	424
5.6. Компьютерное зрение	427
VI. Сетевые технологии. Средства обработки и передачи информации	430
6.1. Общие понятия и термины	430
6.2. Локальные вычислительные сети	433
6.3. Распределенные вычислительные сети	441
6.3.1. Общие понятия и термины	441
6.3.2. Интранет	450
6.3.3. ETHERNET	455
6.4. Глобальные вычислительные сети, интернет	471
6.4.1. Общие понятия и термины	471
6.4.2. Web-технология	482
6.4.3. Технологии передачи данных по каналам Интернета	489
6.4.4. Сервисы и сервисные средства в интернете	499
6.4.5. Интегрированные службы цифровых сетей — ISDN	518
6.4.6. Сотовая связь и компьютерная телефония	520
6.4.7. Телекоммуникационное оборудование зданий	526
6.4.8. Разработки технических средств и комплексов, основанных на использовании телекоммуникационных технологий	532
6.4.9. Субъекты юридических отношений в интернете	533
6.5. Средства и технологии защиты вычислительных сетей	536
6.6. Основные стандарты сетей передачи данных	541
6.6.1. Стандарты ISO	541
6.6.2. Стандарты IEEE	543
6.6.3. Стандарты ITU-T	554
6.6.4. Другие стандарты и протоколы	560

VII. Компьютерный и сетевой сленг	565
VIII. Иконки и символы-смайлики для электронной почты	592
IX. Сокращения слов и выражений, используемых в Интернете	594
Список литературы	597
Англоязычный алфавитный указатель	644
Русскоязычный алфавитный указатель	708

ПРЕДИСЛОВИЕ К ЭНЦИКЛОПЕДИЧЕСКОМУ ИЗДАНИЮ СЛОВАРЯ

Сегодня общепризнанно, что информационные технологии и средства телекоммуникаций вошли во все сферы жизнедеятельности современного общества, а знания в области Информатики стали обязательной составляющей элементарной грамотности каждого его активного члена. Вместе с тем Информатика как наука развивается столь динамично, что никакие справочные издания и даже специальная литература в области информационных технологий не успевают зафиксировать огромное количество возникающих новых понятий и терминов, не говоря уж о необходимости более или менее однозначной трактовки этих терминов и отражения их в отечественных стандартах.

Появление профессионально составленного словаря, содержащего достаточно большое количество истолкованных терминов, в том числе и из смежных областей знания, — явление достаточно редкое. Несомненно, одним из таких ярких явлений стало издание в 1998 г. профессором Ф. С. Воройским первого тематически организованного систематизированного толкового словаря-справочника "Информатика", который существенно отличается от традиционных толковых словарей и открывает собой серию справочных изданий нового жанра.

Отличительные особенности таких изданий — форма представления материала (с возможностью чтения по разделам и отдельным терминам); полнота и оперативность отображения новых достижений науки и техники в выбранной тематической области; стиль изложения словарных статей, доступный разным категориям читателей; емкое и профессиональное изложение материала; а также наличие в словарных статьях ссылок на использованную литературу для предоставления читателям возможности обращения к первоисточникам для более полного ознакомления с интересующей их информацией. В силу перечисленных особенностей, начиная с первого издания, этот словарь-справочник автоматически стал "Введением в современные информационные и телекоммуникационные технологии в терминах и фактах" и успешно используется в учебных заведениях различного уровня, начиная от старших классов средней школы, при организации учебного процесса с применением информационных технологий по самым разным предметам и дисциплинам.

Основной материал издания дополняют и оживляют словарь компьютерного сленга, перечни пиктограмм и сокращений, используемых в электронной почте и Интернете.

Три предыдущих издания словаря-справочника, выпущенные соответственно в 1998, 2001 и 2003 гг., мгновенно нашли своего читателя и пользовались громадным спросом. При подготовке очередных изданий текст словаря существенно дополнялся автором: уточнялись некоторые переводы и толкования, изменялась структура отдельных подразделов, вводились новые словарные статьи и группы терминов.

В энциклопедическом издании отражены те изменения в предметной области, которые произошли за последние три года. В результате общее число терминов составило около 17 тыс., соответственно дополнены и переработаны более 4 тыс. статей, существенно увеличено количество ссылок на первоисточники.

Изменены состав и структура ряда разделов, в частности, выделены новые самостоятельные разделы.

Отличительными особенностями энциклопедического издания также являются более полный охват и подробное описание телекоммуникационных стандартов, протоколов и новых направлений в развитии вычислительной техники и средств программного обеспечения; кроме того сохранены описания объектов, относящихся к информационным технологиям, которые сыграли значительную роль в их развитии в прошлом, а сегодня уже стали историческими фактами. Эти особенности дали полное основание автору назвать предлагаемое читателю издание "энциклопедическим".

Работа автора над отслеживанием изменений в предметной области продолжается буквально ежедневно и, надеемся, будет оперативно отображаться в новых изданиях словаря-справочника.

Доктор физико-математических наук, профессор

А. М. Елизаров

ПРЕДИСЛОВИЕ К ТРЕТЬЕМУ ИЗДАНИЮ СЛОВАРЯ: О СЛОВАРЕ-СПРАВОЧНИКЕ И ЕГО АВТОРЕ

Совсем немного в отечественной и мировой издательской практике найдется литературных источников, относящихся к каким-либо областям науки и техники, которые вызывали бы одинаковый интерес как непрофессионалов, так и самых продвинутых специалистов. Два предыдущих издания словаря-справочника, выпущенных в 1998 и 2001 гг. и быстро реализованных без широкой и шумной рекламной компании, убедительно доказывают, что он относится именно к такому классу литературы. Сказанное подтверждает и значительное число предложений, поступивших Ф. С. Воройскому в последние годы, о размещении его словаря-справочника на Веб-сайтах различных организаций, в том числе и в коллекции "Избранные справочники" широко известного в нашей стране и за рубежом справочно-энциклопедического проекта **"Рубрикон"** (<http://www.rubricon.com>).

По отзывам читателей, в основе успеха словаря-справочника — помимо актуальности темы для России, быстро набирающей темпы автоматизации информационных процессов — лежат такие его особенности, как форма представления материала (можно читать по разделам и отдельным терминам); емкое, профессиональное изложение материала; наличие справочных данных о важнейших объектах и явлениях, связанных с Информатикой и современной информационной технологией; полнота и оперативность отображения новых достижений науки и техники в весьма широкой тематической области и, наконец, стиль изложения словарных статей, доступный разным категориям читателей.

Надеемся, что и третье издание словаря-справочника, существенно дополненное и переработанное, вызовет положительный отклик читательской аудитории. Представляем его автору.

Воройский Феликс Семенович, кандидат технических наук, профессор, капитан 2 ранга, инженер в отставке.

В 1953 г. окончил Высшее военно-морское училище связи им. Попова (г. Петродворец Ленинградской обл.) по специальности "Офицер радиотехнической службы корабля". С 1953 по 1958 г. служил начальником радиотехнической службы эскадренного миноносца на Черноморском флоте.

В 1961 г. с отличием окончил Военно-морскую ордена Ленина академию (г. Ленинград) по специальности "Военный инженер по радиоэлектронике" и с 1961 по 1974 г. служил в научно-исследовательском центре Министерства обороны, где в 1965 г. защитил диссертацию.

Информационной деятельностью занимается с 1976 г.:

1976–1978 гг. — внештатный референт редакции "Машиностроение" Всесоюзного института научной и технической информации (ВИНИТИ);

1978–1997 гг. — старший научный сотрудник, зав. сектором и зав. лабораторией комплектования и эксплуатации машинных массивов Интегрированной системы информационного обеспечения (ИСИО) Всесоюзного института межотраслевой информации (ВИМИ) — в то время являвшегося головным информационным органом оборонных отраслей промышленности СССР;

1988–1994 гг. — заведующий кафедрой информационного и библиотечного обслуживания, директор Научно-информационного и вычислительного центра Института повышения квалификации информационных работников (ИПКИР).

В начале 1993 г. разработал и приступил к реализации Программы автоматизации массовых (публичных) библиотек Москвы. Под его непосредственным руководством практически с нулевого уровня была заложена основа автоматизации Центральной городской публичной библиотеки Москвы им. Н. А. Некрасова и 16 центральных районных библиотек.

С октября 1994 г. работает в ГПНТБ России: заведует сектором анализа, исследований и экспертиз по проблемам автоматизации библиотечно-информационных процессов.

Ф. С. Воройский — профессор кафедры информационных технологий и электронных библиотек Московского государственного университета культуры и искусств (МГУКИ).

Начиная с 1985 г. и по настоящее время принимает участие в разработке ряда крупных проектов в области автоматизации информационных и библиотечных процессов; является ответственным исполнителем проектов создания Корпоративной сети публичных библиотек Москвы, Российского центра корпоративной каталогизации и Единой библиотеки г. Обнинска — единой университетской библиотеки и др.; член Совета и председатель НТС "Информационные ресурсы и обслуживание" Ассоциации российских библиотечных консорциумов (АРБИКОН); эксперт Российского фонда фундаментальных исследований (РФФИ); член постоянных оргкомитетов международных конференций "Крым" и "Либком"; автор более 240 научных работ, в том числе 5 монографий, а также 15 авторских свидетельств на изобретения и промышленные образцы изделий.

Доктор технических наук, профессор

Я. Л. Шрайберг

ОТ АВТОРА

Вычислительная техника и телекоммуникации глубоко вошли во все сферы человеческой деятельности и стали их неотъемлемой частью. Поэтому овладение немалым объемом знаний в области современных информационных технологий и инструментальных средств их обеспечения столь же необходимо каждому, как и элементарная грамотность. В первую очередь это касается терминологии. К сожалению, здесь возникает много проблем, которые связаны не только с тем, что Информатика является относительно новой наукой. Не менее существенно и то, что, будучи существенно более динамичной, чем любая другая область знаний, Информатика вводит как в профессиональный, так и бытовой язык большое количество новых понятий и терминов, многие из которых не отражены в отечественных стандартах и трактуются далеко не однозначно. Наряду с этим даже у профессионалов в области информационных технологий систематически возникает необходимость ознакомиться в краткой (реферативной) форме с характеристикой программных, технических и других средств, создаваемых и используемых в смежных для них областях деятельности, связанных с информационными и телекоммуникационными технологиями.

Определенную помощь в преодолении упомянутых проблем призваны оказать существующие учебные пособия по Информатике, специальная литература по различным программным и техническим средствам, а также терминологические словари. Однако пособия читать довольно трудно, поскольку они не ориентированы на терминологические проблемы и перегружены сведениями, которые конкретному пользователю не всегда необходимы. Словари не дают возможности получить систематизированные знания, так как служат лишь для толкования отдельных терминов, понимание которых может быть затруднено, поскольку требует определенных исходных знаний, включающих представление и о других взаимосвязанных понятиях и терминах, извлечь которые из существующих словарей — задача далеко не простая, более того, часто просто невыполнимая.

Так возникла идея создания тематически организованного словаря-справочника, который можно было бы читать "по рядку" или по разделам, а для обращения к отдельным терминам, понятиям и объектам описания, связанным с информационными технологиями, использовать алфавитные указатели.

В процессе работы над первыми изданиями словаря появилась необходимость расширения многих словарных статей. Кратких определений терминов (дефиниций) оказалось недостаточно. В начале это было связано с неодинаковым их толкованием рядом уважаемых авторов публикаций и специалистов, практикующих в разных областях Информатики. В дальнейшем представилось весьма актуальным дополнить словарные статьи наиболее важными справочными данными о характере объектов описания, в том числе разных перспективных технологий, программных и технических средств, стандартов, ведущих организаций и фирм, ученых и разработчиков. В свою очередь это привело к необходимости введения в словарные статьи ссылок на использованную литературу, чтобы предоставить читателям возможность обращения к первоисточникам для более полного ознакомления с интересующими их аспектами содержания статей словаря.

В результате появился "Систематизированный толковый словарь-справочник", отличный по жанру от традиционных толковых словарей. По отзывам ряда читателей, преимущественно из числа профессорско-преподавательского состава различных учебных заведений, отдельные разделы словаря они используют в качестве основы учебных курсов по дисциплинам, связанным с Информатикой, современными телекоммуникационными технологиями и инструментальными средствами их обеспечения. Автор также использует материалы данного словаря-справочника в курсах лекций, которые читает в Московском государственном университете культуры и искусств (МГУКИ) и в системе повышения квалификации библиотечных и информационных работников.

Помимо сказанного, о пользе такой организации словаря говорят многочисленные отзывы коллег, работающих в области проектирования и/или эксплуатации современных автоматизированных информационных и телекоммуникационных систем различного назначения.

Таким образом, словарь-справочник стал также "Введением в современные информационные и телекоммуникационные технологии в терминах и фактах".

Как уже отмечалось, изменения в предметной области, которую представляет словарь, происходят стремительно. Многое из того, что всего год назад относилось к "последнему слову", успело "устареть", заменено или отошло на второй план. В то же время для многих практических приложений вопросы, ставшие "историей" остаются актуальными. Поэтому в данном издании сохранены описания объектов, имеющих отношение к информационным технологиям, которые сыграли значительную роль в их развитии, хотя, казалось бы, они навсегда отошли в прошлое. Особое внимание уделено более полному охвату и подробному описанию телекоммуникационных стандартов и протоколов, новых направлений в развитии вычислительной техники и средств программного обеспечения. С учетом сказанного, данное издание автору было предложено назвать энциклопедическим. О правомерности этого судить Вам, уважаемый читатель!

Считаю своим приятным долгом искренне поблагодарить многих коллег, ценные замечания и предложения которых учтены в данном издании и, в первую очередь профессоров **А. Б. Антопольского**, **Н. И. Гендину**, **А. М. Елизарова**, **Ю. Н. Столярова**, **Я. Л. Шрайберга**.

Ф. С. Воройский

О ПОЛЬЗОВАНИИ СЛОВАРЕМ

В тексте статей часто встречаются слова **объект**, **тип объекта**, **понятие**, **термин** и **дефиниция**. Определим их значение:

- **объект [object]** — любой материальный предмет или явление, с которым связана познавательная, информационная или любая другая практическая деятельность человека;
- **тип объекта [object type]** — обобщенное имя какого-либо множества объектов, характеризующееся определенным набором признаков (в том числе свойств и характеристик), например "автомобиль", "гроза", "компьютерные игры" и т. д.
- **понятие [concept, notion]** — форма мысли, в которой отражаются общие и существенные признаки предмета или явления; целостная совокупность суждений о каком-либо объекте, отображающая его сущность и являющаяся результатом познания **объекта**;
- **термин [term]** — слово или словосочетание, обозначающее строго определенное **понятие**;
- **дефиниция [definition, concept definition]** — краткое научное определение какого-либо **понятия**; выделяет существенные черты определенного **объекта** таким образом, что он отличается от любых других объектов, и выражается законченным предложением, построенным по правилам логики.

Принципы построения словаря:

1. **Словарные статьи** вместо общепринятого алфавитного расположения представлены в тематических разделах в логической последовательности: от простого понятия к более сложному или от общего к частному.
2. **Пояснительная часть словарных статей** жестко не формализована, при необходимости расширена и содержит связанные с основным понятием термины¹, а также варианты толкования отдельных терминов, существующие в различных нормативных и других документальных источниках.
3. **Варианты терминов — синонимы**, обозначающие одно и то же понятие, отделены друг от друга запятой, точкой с запятой или косой чертой.
4. **Составные части сложных терминов**, которые используются как синонимы или носят факультативный характер (т. е. могут опускаться), заключены в круглые скобки.
5. **Русскоязычные термины** сопровождаются англоязычными эквивалентами, заключенными в квадратные скобки. В случае отсутствия устоявшегося англоязычного эквивалента термина, он не указывается.
6. **Переводы англоязычных терминов**, не имеющих устоявшихся русскоязычных эквивалентов, выделены кавычками и представлены с прописной буквы.

¹ Термины, выделенные **полужирным** шрифтом, как правило, могут быть найдены по алфавитным указателям.

7. **Англоязычные термины**, на которые русскоязычный эквивалент отсутствует и вводить его признано нецелесообразным, помечаются знаком * (звездочка).

8. **Сокращенные термины** представлены, как правило, через запятую вместе с их развернутой формой.

9. **Сокращенные англоязычные термины** и их развернутая форма разделяются круглыми скобками.

10. **Составные термины** могут содержать англо- и русскоязычные части, если каждая из них в отдельности и их комбинации являются широкоупотребительными, например **Web-сервер²**, индекс **iCOMP**, **USB-шина**, **накопитель CD-RW** и т. п.

11. **В словарных статьях используются два близких по своему значению сокращения: ПК (рус.)** — по отношению ко всем персональным компьютерам и **РС (англ.)** — преимущественно по отношению к IBM-совместимым ПК.

12. **Поиск терминов** (в том числе аббревиатур) можно провести по алфавитным русско- и англоязычному указателям, помещенным в конце книги. В алфавитные указатели включены также фамилии разработчиков, наименования фирм, крупных международных и национальных организаций, наиболее распространенных программных и технических средств.

13. **Выделенные в пояснительной части статей полужирным шрифтом термины** обращают внимание читателей на возможность их поиска по алфавитным указателям.

14. **В специальные разделы выделены** компьютерный сленг (VII); пиктограммы, используемые в электронной почте (VIII); сокращения слов и выражений, используемых в Интернете (IX).

² В настоящее время в различных публикациях все чаще встречается русскоязычное написание — **Веб**. Однако, учитывая профессиональную специфику словаря-справочника, в нем преимущественно используется оригинальное международное англоязычное написание — **Web**.

І. ОСНОВЫ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ

1.1. Данные, информация, знания, логика

ДАННЫЕ [data, information]

Сведения, факты, показатели, выраженные как в числовой, так и любой другой форме.

Указанный в скобках англоязычный эквивалент термина (**information**) показывает, что он весьма часто рассматривается как синоним термина **информация**. Так, словосочетание **data system** и производные от него могут соответствовать одному из совсем не однозначных, хотя и связанных между собой понятий — **система данных** и **информационная система**. Аналогичное явление прослеживается при русскоязычном использовании этого термина. Например, в ГОСТ 15971-84 — как "Информация, представленная на материальных носителях" [2], а в ГОСТ 7.0-99 — "Информация, обработанная и представленная в формализованном виде для дальнейшей обработки" [22].

С целью смыслового разделения понятий "информация" и "данные" Ассоциация стандартов Франции (**АФНОР**) дает следующее определение: "Данные — факты, понятия или инструкции, представленные в условной форме, удобной для пересылки, интерпретации и обработки человеком или автоматизированными средствами".

Согласно другому важному для понимания этого термина определению: "Данные — некоторый факт, то на чем основан вывод или любая **интеллектуальная система**" [1]. Компонентами данных являются цифры и символы естественного языка или их кодированное представление в виде двоичных **битов**.

ИНФОРМАЦИЯ

[**Information** — от лат. *Informatio* — разъяснение, осведомление]

Данный термин и отражаемое им понятие являются сегодня одними из самых распространенных. Сказанное относится к их использованию, как на бытовом, так и на профессиональном уровнях.

Существует множество различных определений этого понятия, например:

- "Информация — содержание какого-либо сообщения, сведения о чем-либо, рассматриваемые в аспекте их передачи в пространстве и времени ..." [2];
- "Информация — сведения, подлежащие передаче";
- "Информация — это значение, вкладываемое человеком в данные на основании известных соглашений, используемых для их представления";
- "Информация — сведения, воспринимаемые человеком и/или специальными устройствами как отражение фактов материального или духовного мира в процессе коммуникации" [22];
- "Информация — содержание, значение данных, которое видят в них люди. Обычно данные состоят из фактов, которые ставятся информацией в определенном контексте и понятны людям" [35].

Следует упомянуть также классическое определение **К. Шеннона**, в соответствии с которым информация это то, что сокращает степень неопределенности (у Шеннона — **энтропии**¹) знаний у ее адресата о каком-либо объекте (в том числе явлении, передаваемом сигнале и т.п.). Другими словами, по Шеннону информация это то, что увеличивает степень знания ее адресатом интересующих его объектов и явлений окружающего мира. В указанном контексте количество информации можно даже рассчитать, в частности по увеличению вероятности успешного решения поставленной задачи.

С чем же связано различное представление понятия “информация” разными ее пользователями, включая профессионалов?

Во-первых, с его сложной и неоднозначной сущностью, которая к тому же имеет тенденцию достаточно быстро изменяться в ходе научно-технического прогресса. К примеру, в 1992 г. в журнале “Научно-техническая информация” в статье одного уважаемого автора достаточно убедительно доказывался тезис, согласно которому информация отнюдь не всегда повышает вероятность успешного решения некоторых прикладных задач.

Во-вторых, с тем, что цитируемые и другие определения этого понятия выделяют только те его признаки, которые служат достижению конкретных целей или соответствуют контексту документов, в которых они опубликованы. Так, наука “**Кибернетика**”, расширенно толкуя понятие “информация”, вывела его за пределы человеческой речи и других форм коммуникаций между людьми, связав его с целенаправленными системами любой природы — биологической (например, наследственности), технической (например, сигналы в электрических сетях), социальной (движение человеческих знаний в общественных системах) и т.п. Подробнее см. [4, 38].

Мы остановимся только на тех признаках понятия “информация”, которые необходимы большинству наших читателей, тем или иным образом преимущественно связанных с информационными сферами деятельности.

Если на бытовом уровне смешение понятий **данные** и **информация** вполне допустимо, то для профессионалов это может привести и приводит к серьезным последствиям.

Чтобы стать информацией, данные должны:

1. Правильно отражать объекты описания, в противном случае мы будем иметь дело с **дезинформацией** (ее англоязычные эквиваленты: **false information, misleading information**). Сама по себе “правильность отражения действительности” в соответствии с теорией познания всегда носит условный характер, поскольку связана с уровнем развития знаний на данном этапе развития общества или отдельных социальных групп и индивидуумов. Так, состав и точность данных, которыми владеют или которые необходимы различным организациям и лицам об одном и том же объекте, будут существенно различаться в зависимости от образовательного, возрастного, социального состояний субъекта информирования, а также целей их использования (например, для “повышения общей эрудиции” или для решения конкретных научных, технических, производственных, коммерческих или других задач). В указанном плане данные, которые для одного

¹ **Энтропия** — мера неопределенности. Измеряется вероятностью наступления одного из N возможных событий. Если вероятность одного из них становится равной 1, то неопределенность отсутствует (вероятность остальных равна 0). Максимальная неопределенность имеет место, если все события равновероятны [4].

субъекта будут представляться вполне точными, для другого — могут оказаться грубой дезинформацией;

2. Быть необходимы человеку для удовлетворения его **информационной потребности**;

3. Быть получены пользователем своевременно (*не раньше и не позже*) наступления в них **информационной потребности**. Всякое несвоевременное, а также повторное предоставление сведений, которыми адресат уже владеет, является **информационным шумом** (см. далее);

4. Быть представлены в форме, удобной для восприятия тем, кому они предназначены.

Перечисленные признаки информации можно кратко обобщить следующим образом:

"Информация — это данные, удовлетворяющие информационную потребность того, кому они передаются, соответствующие действительности и материализованные в форме, удобной для использования, передачи, хранения и/или обработки (преобразования) человеком или автоматизированными средствами".

И в заключение небольшое отступление, адресованное нетерпеливому читателю, раздраженному кажущейся ему пространностью этой статьи: "Скажите, с чем связаны бытовавшие и, увы, существующие еще способы оценки эффективности работы библиотек, информационных органов, а также отдельных их служб и работников по количеству выданных пользователям (или читателям) документов, справок, библиографических описаний и т.п. без учета соответствия содержащихся в них данных признакам, определяющим их информационную ценность?". О философских взглядах на сущность информации см. [1045].

Информационный шум [information noise] — данные, не соответствующие информационной потребности или не представляющие для субъекта информирования предмета новизны (другими словами, этими данными он уже владеет).

Понятие "информационный шум" может быть также распространено на данные, "не удобные для использования, передачи, хранения и/или обработки", поскольку и в этом случае они приводят к бесполезным, а, возможно, и вредным затратам материальных, временных и других ресурсов.

Дезинформация [misinformation] — передаваемые кому-либо и в любой форме данные, сведения, сообщения и т.п., неверно отражающие объекты описания реального мира или мыслительной деятельности человека.

Термин "информация" в сочетании с различными прилагательными широко употребляется в следующих случаях:

1. Сведения или данные, полученные в процессе какого-либо вида деятельности, отражают результаты этого вида деятельности или имеют отношение к ней и предназначены для **справочно-информационного обслуживания и/или информационного обеспечения** (см. далее) заинтересованных пользователей, например: **научно-техническая информация [scientific-technical information]**, **юридическая информация [juridical information]**, **патентная информация [patent information]** и т.п.

2. Сведения или данные имеют определенное назначение, например:

• **справочная информация [reference information]** — сведения или данные для выдачи справок о чем-либо;

• **сигнальная информация [alert information, current awareness information]** — информация, предназначенная для быстрого предварительного оповещения.

3. *Словарная статья, определяет характер, принадлежность, форму или вид данных, используемых в информационном процессе, например:*

- **априорная информация [aprior information]** и **апостериорная информация [aposterior information]** — соответственно — данные, имевшиеся до проведения какого-либо опыта или другого действия, и сведения, полученные после его выполнения;

- **коммерческая информация [commercial information]** — данные, сведения и содержащие их документы, являющиеся объектом продажи их собственником;

- **личная информация [private information]** — сведения (данные) о гражданах и организациях, затрагивающие их интересы и запрещенные для распространения без их согласия;

- **библиографическая информация [bibliographic information]** — библиографические данные, описания и их перечни;

- **графическая информация [graphical (pictorial, image, pattern) information]** — сведения или данные, представленные в виде схем, эскизов, изображений, графиков, диаграмм, символов;

- **ретроспективная информация [retrospective information]** — сведения, содержащиеся в накопленных за два и более лет массивах данных или полученных в результате поиска в этих массивах (так называемого **ретроспективного поиска**).

4. *Словарная статья характеризует средства закрепления, отображения и/или передачи данных, например:*

- **документальная информация [documentary information]** — сведения, закрепленные на каком-либо материальном **носителе**; содержание документа или текста;

- **устная информация [oral information]** — содержание устного сообщения и т.п.

ЗНАНИЯ [knowledge]

Совокупность хранимых в памяти человека или **базах знаний** фактов (данных, сведений) о некоторой предметной области, их взаимосвязей и правил, которые могут быть использованы для получения новых фактов и/или решения каких-либо задач, связанных с различными видами интеллектуальной деятельности людей или их сообществ.

Знания отражают множество возможных ситуаций, характеризующихся состоянием и конкретной реализацией объектов определенного типа, способы перехода от одного описания объекта к другому. Для знаний характерна внутренняя интерпретируемость, структурированность, связанность и активность. Условно можно записать: "знания = факты + убеждения + правила" [265].

Часто понятие "знание" неверно отождествляют с "информацией". В указанном контексте "информация" может рассматриваться только как дополнение к "знанию", становящееся его частью после получения субъектом, которому она предназначена.

Различают следующие виды знаний [722]:

- **базовые (фундаментальные) знания [deep knowledge]** — систематизированные знания, основанные на модели, описывающей все значимые аспекты некоторой предметной области, которые описывают ее сущности, их свойства и различные связи между ними;

- **компилятивные знания [compiled knowledge]** — знания, полученные на основе уже известных знаний, путем их структурирования и/или систематизации в форме, необходимой для их использования в новых целях;

- **неполные знания [incomplete knowledge]** — знания, из которых не могут быть получены значимые факты в данной предметной области;
- **нечеткие знания [fuzzy knowledge]** — знания, основанные на нечеткой логике (см. далее);
- **поверхностные знания [surface knowledge]** — неполные, фрагментарные знания, часто связанные с недостаточной изученностью предметной области. Тем не менее, такие знания могут быть успешно использованы в случаях, не затрагивающих ситуаций, которые ими не охвачены;
- **предметные знания [domain knowledge]** — знания, относящиеся к конкретной узкой предметной области;
- **процедурные знания [procedural knowledge]** — знания, воплощенные в компьютерных программах для решения тех или иных задач.

ЛОГИКА [logic]

Наука о законах и формах мышления, методах познания и условиях определения истинности знаний и суждений.

Нечеткая логика [fuzzy logic] — в математике и вычислительных системах: форма представления знаний или данных, связанных с описанием различных объектов понятиями вида тяжелый, громкий, горячий и т. п., имеющими неточные или неопределенные значения. Более строгое определение значения указанных понятий возможно только с привлечением ряда дополнительных сведений или данных, входящих в **нечеткие множества** и составляющих перечни дополнительных данных или условий. Центральным понятием нечеткой логики является понятие **вероятность члена множества**, определяющее степень правомочности отнесения данного члена к указанному множеству. Например, для члена нечеткого множества, относящего предметы к тяжелым, вес 20 кг может иметь такое значение с вероятностью 90% — для миниатюрных женщин и, скажем, только для 20% — мужчин.

Нечеткая логика широко используется в различного рода **экспертных системах** для автоматизированного принятия решений, близких к человеческим, на основе адекватного реагирования на сигналы, поступающие от связанных с ними датчиков, а также команды с пульта управления.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ [artificial intelligence]

1. Искусственная система, имитирующая решение человеком сложных задач, связанных с различными видами его деятельности.

2. Научное направление, связанное с созданием на базе средств вычислительной техники средств обработки больших объемов данных и выработки на основе **моделирования** органов человека и/или заданных им **алгоритмов** решений определенных практических задач. Примерами использования искусственного интеллекта являются "**экспертные системы**", "**интеллектуальные системы**" и "**компьютерное зрение**" [4].

1.2. Информационные ресурсы, теория информации, информатика

ИНФОРМАЦИОННЫЕ РЕСУРСЫ, ИР [information resources]

1. Федеральный закон РФ "Об информации, информатизации и защите информации" трактует данный термин как "... отдельные документы и отдельные массивы документов, документы и массивы документов в информационных си-

стемах: библиотеках, архивах, фондах, банках данных, других видах информационных систем”.

2. В общем случае под ИР понимается вся совокупность сведений, получаемых и накапливаемых в процессе развития науки и практической деятельности людей, для их многоцелевого использования в общественном производстве и управлении. ИР отображают естественные процессы и явления, зафиксированные в результате научных исследований и разработок или других видов целенаправленной деятельности в различного рода документах (например в отчетах о НИР, патентах, проектно-конструкторской документации, массивах данных и т. п.), понятиях и суждениях, а также более сложных моделях действительности [4].

Этот термин начал широко использоваться в конце 1970-х — начале 1980-х гг. в результате осознания растущей зависимости промышленно развитых стран, отдельных организаций и фирм от источников информации (технической, политической, военной и т. д.), а также от уровня развития и использования средств передачи и переработки информации. С ним связаны термины: **национальные информационные ресурсы** (в том числе **государственные и негосударственные информационные ресурсы**), информационные ресурсы территориально-административных образований, фирм (организаций), их подразделений и т. п.

В современном обществе ИР относятся к материальным и наиболее важным видам ресурсов, определяющих экономическую, политическую и/или военную мощь их владельца. В подтверждение этого тезиса можно привести ставший классическим пример: Япония, страна, практически лишенная природных ресурсов и обладающая весьма скромными людскими ресурсами, является крупнейшим в мире производителем и экспортером не только изделий микроэлектроники, но и такой материалоёмкой продукции, как автомобили и супертанкеры [6].

Отличием ИР от других материальных видов ресурсов (например полезных ископаемых) является их воспроизводимость. Как и другие виды ресурсов, ИР являются объектами импорта-экспорта, а также конкуренции, политической и экономической экспансии. Следует отметить, что границы понятия ИР в настоящее время четко не установлены. Так, некоторые ученые включают в его толкование также степень профессиональной подготовки общества или его части, способность воспроизводить и использовать ИР. Другие ограничивают ИР только совокупностью зафиксированных в документах и данных сведений, “представляющих ценность для учреждения (предприятия)” или, добавим — другого владельца ИР [5]. Заметим, что в последнем случае в понятие ИР не включены средства передачи и переработки информации.

Не будем спорить ни с теми, ни с другими авторами, хотя признаемся, что мы считаем предпочтительней более широкий подход к определению этого очень важного и интересного понятия. Для тех, кто интересуется данной проблемой, рекомендуем монографию Г. Р. Громова [6]. Об общих проблемах ИР и ИР России см. монографию А. Б. Антопольского [1046].

Виртуальные (информационные) ресурсы [virtual (information) resources] — информационные ресурсы других организаций, предприятий, фирм и т. п., доступные пользователям в режиме **теледоступа** по каналам глобальной связи, например Интернета.

ТЕОРИЯ ИНФОРМАЦИИ [Information theory]

Раздел **Кибернетики**, изучающий общие стороны процессов передачи, хранения, извлечения и классификации **информации** различной природы (в том числе биологической, технической, социальной и др.) в независимости от ее

семантического (смыслового) содержания. Общим средством анализа, описания и количественной оценки исследуемых процессов теории информации является ее математический аппарат, представляющий собой основу разрабатываемых и используемых теорий и методов.

Важнейшей частью теории информации является **Теория передачи информации** [Theory of communication]. Ее основоположник — американский математик **К. Шеннон**. Основные понятия этого раздела теории: **энтропия** (количественная мера неопределенности ситуации) от греческого *entropia* — поворот, превращение и **количество информации**, измеряемое величиной изменения энтропии в условиях, связанных с получением информации. С использованием этих понятий выражается пропускная способность канала связи между источником информации и ее адресатом, равная максимально допустимой скорости передачи информации со сколь угодно малой вероятностью ошибки [4].

Составной частью теории информации является также **Теория кодирования** (или **Теория оптимального кодирования**), рассматривающая вероятностные аспекты проблем кодирования и декодирования информации. Большой вклад в разработку теории информации внесли отечественные ученые **А. Н. Колмогоров**, **А. Я. Хинчин**, **Р. Л. Добрушин**, **В. А. Котельников**, **А. А. Харкевич** и др. Возросшая необходимость не только количественного, но и содержательного анализа информационных процессов породила появление новой науки — Информатики.

ИНФОРМАТИКА [Informatics, Information science, Computer science]

Словарь по Кибернетике [4] содержит следующее определение: **"Информатика**: наука, изучающая информационные процессы и системы в социальной среде, их роль, методы построения, механизм воздействия на человеческую практику, усиление этого воздействия с помощью вычислительной техники. Информатика возникла как дополнение и конкретизация Теории информации из потребностей автоматизации социально-коммуникативных процессов, и начала формироваться в 1970-е гг., как научная база использования электронных вычислительных машин в управлении, науке, проектировании, образовании, сфере услуг и т. д."

Как всякая относительно новая и быстро развивающаяся отрасль знания, не только связанная с социальной сферой, но и широко используемая в ней, Информатика получила в последние годы множество толкований и не все они однозначны.

Наибольшие противоречия связаны с той частью этого понятия, которая определяет его семантические границы распространения. В качестве примера приведем другое определение: **"Информатика**: отрасль знания, изучающая закономерности сбора, преобразования, хранения, поиска и распространения документальной информации и определяющая оптимальную организацию информационной работы на базе современных технических средств" [5].

Видимые отличия цитируемых определений заключаются, в частности, в том, что второе ограничивает понятие **"Информатика"** технологическими процессами, входящими в функции информационных органов, а также документальной информацией. Следует отметить, что, несмотря на давность этого определения (1971 г.), оно используется и в настоящее время в среде работников информационных органов и служб, в недрах которых изначально и было порождено.

Еще один подход связан с организациями, которые ранее были подведомственными Комитету по информатизации России. Основное внимание этот под-

ход акцентирует на инструментальных (программных и технических) средствах "Информатики" и "информатизации": "Информатика: ... группа дисциплин, занимающихся различными аспектами применения и разработки ЭВМ: Прикладная математика, Программирование, Программное обеспечение, Искусственный интеллект, Архитектура ЭВМ, Вычислительные сети" [265].

Анализируя сказанное, мы склонны предпочесть вариант "кибернетиков" как более объективный и полный.

С целью более глубокого понимания указанного термина продолжим выборочное цитирование соответствующей статьи Словаря по Кибернетике: "... Важнейшими категориями Информатики являются понятия информационных сред (социальных подсистем, в которых осуществляются информационные процессы и куда внедряются ЭВМ как усилители человеческого интеллекта), полного информационного цикла (включающего зарождение информации, ее переработку, передачу, использование для снижения энтропии рассматриваемой социальной системы), полезной работы (отдачи) ЭВМ. Отдача ЭВМ, коэффициент полезного действия зависят от уровня функционирования социальной среды, в которой они задействованы, ее упорядоченности, системности, условий для творческой деятельности людей, сложности и важности задач, решаемых с помощью машин. Информатика не заменяет собой Кибернетику, теорию информации, электронику, системотехнику, а взаимодействует с ними, имея ряд общих проблем. Интегральный характер Информатики заключается также в ее взаимодействии с такими дисциплинами, как теория познания, семиотика, лингвистика, документалистика, библиотековедение" [4].

1.3. Носители информации, документы, документация, издания

НОСИТЕЛЬ [media]

Обобщающее наименование материала, на который можно записывать данные. Носители подразделяются на **человекочитаемые (твердые)** и **машиночитаемые** (см. далее).

- **Машиночитаемый носитель [machine-readable media]** — носитель, пригодный для непосредственной записи и считывания данных программно-техническими средствами (ЭВМ). Термин обычно применяется к устройствам **внешней памяти ЭВМ**, например **магнитным** и **оптическим** дискам, **дискетам** и т.п. Однако он может использоваться и по отношению к определенной части **твердых носителей**, если они допускают использование специальных считывающих устройств, например **сканеров**. Подробнее о видах носителей см. раздел 3.5.2. "Память ЭВМ — понятия и термины", а об их характеристиках — раздел 3.6.1. "Внешняя память ЭВМ, накопители и связанные с ними термины".

- **Человекочитаемый (твердый) носитель [human readable media, hard media]** — носитель, пригодный или используемый для записи данных непосредственно считываемых человеком, например — бумага.

- **Микроформа, микроноситель [microform]** — общее наименование носителей, на которых тексты или графические изображения представлены в уменьшенном фотографическом способе виде. Микроформы широко используются в библиотеках для хранения в человекочитаемом виде (чтение их производится с использованием специальных читальных аппаратов) больших объемов копий

документов и данных. Создание микроформ производится при помощи специальной (микрофильмирующей) фотоаппаратуры, а также компьютерной записи (см. также «СОМ»). Основными типами микроносителей являются:

1. **Микрофильм [microfilm]** — пленка обычного фотоаппарата содержащая уменьшенные копии текстов и графических изображений;

2. **Микрофиша [microfiche]** — стандартный (105 × 148 мм) прямоугольный лист фотопленки, на котором располагается 420 кадров уменьшенных изображений страниц текста или графики.

ДОКУМЕНТ

[document — от лат. *documentum* — свидетельство, доказательство]

Материальный носитель информации, зафиксированной вне памяти человека. В соответствии с ГОСТ 16487-70 [8] «документ» является средством «закрепления различным способом на специальном материале сведений о фактах, событиях явлениях объективной действительности и мыслительной деятельности человека». Документы могут содержать текстовую, цифровую, графическую и аудиоинформацию. Они могут подвергаться процессам записи (преобразования), хранения, поиска, передачи, получения, сбора и чтения. В свою очередь, ГОСТ 7.60-2003 [11] понимает под документом «зафиксированную на материальном носителе информацию с реквизитами, позволяющими ее идентифицировать». В зависимости от характера обработки содержащихся в них данных документы принято разделять на **первичные и вторичные**:

• **Первичный документ [primary document, source]** — документ, содержащий исходную запись сведений, полученных в процессе исследований, разработок, наблюдений, анализа или других видов человеческой деятельности независимо от ее характера или тематики. Деление документов на первичные и вторичные в значительной степени условно, поскольку один и тот же документ может содержать сведения, относящиеся и к первой, и ко второй группе. Так, считающиеся безусловно первичными видами документов отчеты по НИР, проектные документы, монографии и т.д., могут содержать наряду с оригинальными сведениями или данными ранее включенные в другие первичные, или даже вторичные документы.

• **Вторичный документ [secondary document]** — документ, полученный в результате аналитико-синтетической и логической переработки сведений или данных, содержащихся в первичных документах. Примерами вторичных документов являются справочные и энциклопедические издания, рефераты и реферативные издания, библиографические издания, указатели и списки, обзоры (за исключением аналитических обзоров) и т.п.

Юридический документ [juridical (legal) document] — документ, оформленный в соответствии с действующим юридическим законодательством и имеющий правовое значение.

В зависимости от характера связи документов с технологическими процессами в автоматизированных системах и вида физического носителя информации различают:

• **машинно-ориентированный документ [machine-oriented document]** — документ, предназначенный для обработки части содержащейся в нем информации средствами вычислительной техники [9]. Примерами машинно-ориентированных документов могут служить заполненные специальные формы бланков библиографической записи, различного рода анкет и т.п., предназначенные для

последующего считывания в ЭВМ записанных в них данных с использованием клавиатурных операций;

- **машиночитаемый документ [machine-readable document]** — документ, пригодный для автоматического считывания содержащейся в нем информации [9]. Средства автоматического считывания — **сканеры** предъявляют определенные требования к характеру оформления соответствующих текстовых, графических и других видов записей, включая их **формат**, виды шрифтов, наличие специальных служебных знаков, контраст и т. п.;

- **документ на машиночитаемом носителе, электронный документ [electronic document]** — документ, созданный средствами вычислительной техники, записанный на **машиночитаемый носитель**: магнитную ленту (МЛ), магнитный диск (МД), дискету, оптический диск и т. п. и оформленный в установленном порядке³. В зависимости от того, на каком носителе записан машиночитаемый (электронный) документ или документы, принято указывать его вид, например, “документ(ы) на магнитном (оптическом) диске”, “документы на магнитной ленте” и т. п.;

- **документ-машинограмма, распечатка [hard copy document]** — документ на бумажном носителе, созданный средствами вычислительной техники и оформленный в установленном порядке;

- **документ на экране дисплея** — документ, созданный средствами вычислительной техники, отображенный на экране дисплея (монитора) и оформленный в установленном порядке;

- **title — электронный документ**, который идентифицируется и передается как единое целое;

- **служебный документ [internal document, in-house document]** — документ, содержание которого отражает способ или результаты решения какой-либо функциональной задачи автоматизированной системы;

- **входной документ [input document]** — документ, составленный по определенной форме и содержащий данные, предназначенные для ввода в ЭВМ. Входные документы могут быть условно разделены на две категории: **информационные документы и запросы**;

- **информационный документ [information document]** — документ, основное назначение которого — пополнение массивов или баз данных (БД) ЭВМ;

- **выходной документ [output document]** — документ, являющийся носителем результатов обработки данных ЭВМ и/или формируемый автоматизированной системой и выданный системными средствами вывода.

ДОКУМЕНТАЦИЯ

[documentation, collection of documents, file of documents]

Совокупность документов, объединенных по определенным признакам (например, по назначению, содержанию и т. п.) и оформленных по единым правилам. Применительно к программным продуктам и средствам вычислительной техники термином “документация” обозначают:

1. Руководства по использованию;
2. Совокупность текстов, описывающих строение и применение соответствующих средств или изделий.

³ *Примечание:* Фраза “оформленный в установленном порядке” в цитируемом выше ГОСТе [9], относится к форме документа, а не к его юридическому статусу. Однако при выполнении определенных условий “оформления” эти дефиниции можно распространить и на юридические документы.

Документация предназначена для облегчения использования программных и технических средств и включает руководства, справочники, учебники, краткие справочники, обучающие программы, а также средства диалоговой документации и подсказки [9].

СИСТЕМА ДОКУМЕНТАЦИИ [documentation system]

Совокупность документов, состав, содержание, структура и правила оформления которых определены государственными стандартами.

Часто понятие "система документации" подменяется другим — **вид документации**, который, строго говоря, не является его синонимом. Так, различают следующие виды документации: научно-техническая (научная и техническая), директивная, юридическая, нормативная (нормативно-техническая), технологическая, проектно-конструкторская (проектная и конструкторская), эксплуатационная, технико-экономическая и др. Данная классификация достаточно условна в том смысле, что состав и правила оформления большинства используемых в реальной практике видов документации, нормативно не определены (см. ГОСТ 6.10.1-88) [9].

ДОКУМЕНТАЛИСТИКА [Documentation]

Направление в **Кибернетике**, занимающееся изучением и оптимизацией документальных систем независимо от их назначения (в первую очередь — документов).

ИЗДАНИЕ [publication, edition]

ГОСТ 7.60-2003 трактует издание, как "документ, предназначенный для распространения содержащейся в нем информации, прошедший редакционно-издательскую обработку, самостоятельно оформленный, имеющий выходные сведения". В зависимости от способа исполнения издания делятся на **печатные и электронные** (см. далее) [11].

Печатное издание [publication, edition] — издание, полученное печатанием или теснением и имеющее самостоятельное полиграфическое оформление. Печатные издания различаются по многим признакам, включая периодичность выпуска, содержание, знаковую природу информации, ее целевое назначение, вид аналитико-синтетической переработки содержащегося в них материала и т. д. Подробнее см. [11].

В зависимости от периодичности выхода различают [11]:

- **непериодическое издание [non-periodical edition]** — издание, выходящее однократно, не имеющее продолжения;
- **сериальное издание [serial edition]** — издание, выходящее в течение времени, продолжительность которого заранее не установлена, как правило, нумерованными и/или датированными выпусками (томами), имеющими одинаковое заглавие;
- **периодическое издание [periodical edition]** — сериальное издание (см. ранее), выходящее через определенные промежутки времени, как правило, с постоянным для каждого года числом номеров (выпусков), неповторяющимися по содержанию, однотипно оформленными, нумерованными и/или датированными выпусками, имеющими одинаковое заглавие. Периодические издания могут быть ежедневными, еженедельными, ежемесячными, ежеквартальными или ежегодными;

• **продолжающееся издание [continued edition]** — сериальное издание (см. ранее), выходящее через неопределенные промежутки времени по мере накопления материала, не повторяющимися по содержанию, однотипно оформленными, нумерованными и/или датированными выпусками, имеющими общее заглавие.

Виды научных изданий [11]:

• **монография [monograph]** — научное или научно-популярное издание, содержащее полное и всестороннее рассмотрение одной проблемы или темы и принадлежащее одному или нескольким авторам⁴;

• **сборник научных трудов [collection of scientific papers]** — сборник (т. е. издание, содержащее ряд произведений одного или разных авторов), содержащий исследовательские материалы научных учреждений, учебных заведений или обществ;

• **материалы конференции (съезда, симпозиума) [proceedings]** — неперидический сборник, содержащий итоги конференции в виде докладов, рекомендаций, решений;

• **препринт [preprint]** — научное издание, содержащее материалы предварительного характера, опубликованные до выхода в свет издания, в котором они могут быть помещены;

• **пролегомены, введение [prolegomena]** — научное или учебное издание, содержащее первичные сведения и основные принципы какой-либо науки;

• **тезисы докладов (научной конференции, съезда, симпозиума) [scientific conference abstracts]** — научный неперидический сборник, содержащий опубликованные до начала конференции материалы предварительного характера (аннотации, рефераты докладов и/или сообщений);

• **автореферат диссертации [authors abstract, synopsis of thesis]** — научное издание в виде брошюры, содержащее составленный автором реферат проведенного им исследования, представляемого на соискание ученой степени;

• **стандарт [standard]** — нормативное производственно-практическое издание, содержащее комплекс норм, правил, требований к объекту стандартизации, которые устанавливаются на основе достижений науки, техники и передового опыта, и утверждаются в соответствии с действующим законодательством. В зависимости от уровня организации, утвердившей и выпустившей стандарт, стандарты подразделяются на международные, государственные, отраслевые и стандарты предприятий. Требования стандартов вышестоящих уровней обязательно распространяются на все стандарты нижестоящих уровней;

• **электронное издание, ЭИ [electronic publication]** — существуют два определения ЭИ, которые хорошо дополняют друг друга и поэтому могут рассматриваться в непосредственной взаимосвязи:

1. ЭИ — электронный документ (документ или группа документов на машиночитаемом носителе), прошедший редакционно-издательскую обработку, предназначенный для распространения в неизменном виде, имеющий выходные сведения. По наличию печатного эквивалента ЭИ подразделяются на электронные аналоги печатного издания (полностью их воспроизводящие) и самостоятельные электронные издания. Подробнее см. [946, 947];

⁴ Монография в соответствии с ГОСТ 7.60-2003 принадлежит к категории моноизданий, поскольку содержит одно произведение.

2. ЭИ — самостоятельный законченный продукт, содержащий информацию, представленную в машиночитаемой (электронной) форме, все копии (экземпляры) которого соответствуют оригиналу, и предназначенный для длительного хранения, широкого распространения и многократного использования неопределенным кругом пользователей.

Электронные издания представляют собой основную часть **контента электронных библиотек**. Подробнее см. [687; 878, С. 29, 30].

По характеру содержащейся в ЭИ информации различают [946]:

- **текстовое (символьное) ЭИ [textual (symbol) electronic publication]** — ЭИ, которое содержит преимущественно текстовую информацию, представленную в форме, допускающей посимвольную обработку;
- **изобразительное ЭИ [graphic(al) electronic publication]** — ЭИ, содержащее преимущественно электронные образы объектов (в том числе текстов), рассматриваемых как целостные графические сущности, которые представлены в форме, допускающей просмотр и печатное воспроизведение, но не допускают посимвольной обработки;
- **звуковое ЭИ [audio (sound) electronic publication]** — ЭИ, содержащее цифровое представление звуковой (аудио-) информации в форме, допускающей ее прослушивание, но не предназначенной для печатного воспроизведения;
- **программный продукт [software product]** — самостоятельное отчуждаемое произведение, представляющее собой публикацию текста программы (или программ) на языке программирования или в виде исполняемого кода;
- **мультимедийное ЭИ [multimedia electronic publication]** — ЭИ, в котором информация различной природы присутствует равноправно и взаимосвязано для решения определенных разработчиком задач, причем эта взаимосвязь обеспечена программными средствами.

По технологии распространения различают [946]:

- **локальное ЭИ [local electronic publication]** — ЭИ, предназначенное для локального использования и выпускающееся в виде определенного количества идентичных экземпляров (тиража) на переносных машиночитаемых носителях;
- **сетевое ЭИ [network electronic publication]** — ЭИ, доступное потенциально неограниченному кругу пользователей через телекоммуникационные сети;
- **ЭИ комбинированного распространения [local and wide-spread propagation electronic publication]** — ЭИ, которое может использоваться как в качестве локального, так и сетевого.

1.4. Принципы структурированного представления документов и данных

1.4.1. Информационные элементы и их виды

ИНФОРМАЦИОННЫЙ ЭЛЕМЕНТ, ИЭ [information element, item]

Единица информации, подлежащей обработке, хранению и передаче пользователям системы или предназначенная для обеспечения ее работы,

ИЭ является обобщенным наименованием структурной единицы информации, не зависящей от ее назначения, состава данных, характера или материала

носителя и т.п. Например, ИЭ может быть названо содержание книги, статьи, библиографического описания или его части. С учетом сказанного ИЭ могут быть делимыми на части или не делимыми.

Виды информационных элементов

● **Составной ИЭ [composite data item]** — информационный элемент, который может быть разделен на части средствами системы (без участия человека в выполнении логических или интеллектуальных операций, связанных с анализом и разделением содержащихся в этом информационном элементе данных), причем каждая из его частей также является информационным элементом системы.

Исходя из этого определения, составной ИЭ характеризуется видом и организацией содержащихся в нем данных. Так, неструктурированные вторичные документы — библиографическое описание, реферат и аннотация — не могут считаться составными ИЭ, если **автоматизированная система** не позволяет на физическом уровне своими средствами выделить содержащиеся в них различные виды данных (например, в библиографическом описании: сведения об ответственности, фамилии авторов, продолжение названия и т.п.) или аспекты описания.

В общем случае ИЭ, которые не могут быть разделены на части автоматизированными средствами, принято называть **данными**.

В указанном контексте уместно упомянуть еще одно определение понятия “данные”: это “информация, представленная в виде пригодном для обработки автоматическими средствами при возможном участии человека” [265].

● **Элемент данных, ЭД [data element, data item]** — неделимый информационный элемент, ЭД являются минимальной структурной единицей информации, поскольку части ЭД теряют признаки информационного элемента. Примерами ЭД в библиографическом описании могут служить: фамилия автора, год издания, название издания и т.п. Деление на части **лексических единиц**, образующих соответствующие понятия, не приводит к образованию новых ИЭ.

Из сказанного следует, что вид ЭД определяется как характером содержащихся в нем сведений, так и особенностью его организации или записи (см. далее статьи “Логическая запись” и “Физическая запись”).

В зависимости от характера содержащихся в ИЭ данных они могут подразделяться на следующие виды:

● **документографический (документальный) ИЭ [documentary data item]** — информационный элемент, содержащий сведения о документах (например: библиографическое описание, реферат, аннотация, поисковый образ документа);

● **полнотекстовый ИЭ [full text(ual) data item]** — информационный элемент, содержащий полные тексты документов или их частей;

● **фактографический ИЭ [fact(ual) data item]** — информационный элемент, содержащий описание отдельных фактов или некоторой ограниченной совокупности фактов, объединенных по каким-либо признакам (например: сведения о курсе валют, расписания движения транспорта и т.п.);

● **объектографический ИЭ [object(ual) data item]** — информационный элемент, содержащий структурированное и логически связанное описание различных объектов науки, техники или управления. Объектографические информационные элементы являются более сложным вариантом **фактографических ИЭ**, поэтому многие специалисты и организации, работающие в области Информатики, используют в рассматриваемых случаях термин фактографический ИЭ или его синонимы.

Развитие вычислительной техники может привести к появлению терминов, расширяющих приведенный ряд, например **графический ИЭ**, **аудио ИЭ** и т. п., хотя, следует признать, что в литературе мы с такими терминами или их синонимами пока не встречались.

ТИП ДАННЫХ [data type]

Множество допустимых значений данных, объединенных общим содержанием и именем (например: *"Библиографическое описание"*, *"Автор"*, *"Год издания"*, *"Стоимость"* и т. п.), а также совокупностью допустимых операций, которые можно выполнять над этими данными, включая способ их хранения в памяти ЭВМ. Понятие *"тип данных"* делает манипулирование данными с использованием средств вычислительной техники абстрактным процессом и скрывает лежащее в основе обращения с ними представление их в виде двоичного кода [35].

Виды типов данных:

- **аналоговые данные [analog data]** — данные, принимающие произвольные значения из заданного диапазона, и представляемые в виде непрерывно изменяющихся физических величин, например напряжения, длительности сигнала и т. п.;
- **дискретные (цифровые) данные [digital data]** — данные, представленные в дискретном коде в определенной, например, двоичной системе счисления;
- **аналого-цифровые данные [analog-digital data]** — аналоговые данные, преобразуемые для обработки в цифровой код;
- **двоичные данные [binary data]** — данные, представленные в двоичном коде;
- **десятичные данные [decimal data]** — данные, представленные в десятичном коде;
- **алфавитно-цифровые (текстовые) данные [alphanumeric data]** — данные, значения которых составлены из любых знаков алфавита;
- **числовые (арифметические) данные [arithmetic data]** — данные, над которыми можно выполнять арифметические операции.

МНОЖЕСТВО [set]

Совокупность каких-либо **объектов**, представляемых как единое целое. Множество может включать в себя как однородные объекты, объединенные каким-либо общим признаком, так и неоднородные. Объединение последних в конкретное *"множество"* определяется на основе ассоциативных и других связей между его элементами (см. далее *"Теория множеств"*).

Виды множеств:

- **конечное множество [finite set]** — множество, содержащее определенное (конечное) число элементов;
- **нечеткое множество [fuzzy set]** — множество, принадлежность объекта к которому определяется функцией, принимающей значения в области $[0, 1]$;
- **ограниченное множество** — упорядоченное множество, имеющее нижнюю и верхнюю границы;
- **пустое множество [empty set]** — множество, не имеющее ни одного элемента;
- **подмножество [subset]** — множество, являющееся частью другого множества;

• **эквивалентное множество [equivalent set]** — множество, имеющее взаимно однозначное соответствие с другим множеством, при котором каждому элементу одного из них соответствует один единственный элемент другого и наоборот.

ТЕОРИЯ МНОЖЕСТВ [set theory]

Раздел математики, исследующий общие свойства **множеств** и операций над ними. Понятия и методы теории множеств широко используются в **Информатике**, в частности, в теории и практике построения **информационно-поисковых систем**⁵.

1.4.2. Запись, файл, массив, ключ

ЗАПИСЬ [record, writing]

1. Единица обмена данными между программой и внешней памятью ЭВМ (**record**).

2. Процесс или результат закрепления (фиксирования) данных на носителе информации (**writing**). Запись может содержать один **информационный элемент**, включая и составной.

Логическая запись [logical record] — совокупность записей взаимосвязанных ИЭ (в том числе элементов данных, данных и составных ИЭ), рассматриваемая в логическом плане как единое целое. Одна логическая запись может состоять из нескольких физических или быть частью одной физической записи.

Физическая запись [physical record] — порция данных, пересылаемых как единое целое между основной и внешней памятью ЭВМ.

На машиночитаемом носителе (например магнитном или оптическом диске, дискете и т.п.) физическая запись реализуется в форме **поля данных, файла** (см. далее) и др. Физическая запись может содержать одну логическую запись, ее часть или несколько логических записей (см. также "**Блок данных**").

ТИП ЗАПИСИ [record type]

1. Обобщенное имя записи типа данных.

2. В **базах данных**: тип, к которому относится данная запись⁶.

3. Характеристика, определяющая возможность записи менять свою длину (по этому признаку различают записи: **фиксированной, переменной и неопределенной длины**). См. также "**Формат записи**" и "**Переменный формат**".

БЛОК ДАННЫХ, БЛОК [data block, block]

1. Несколько последовательных логических записей, объединенных в одну физическую.

2. Единица доступа к внешней памяти ЭВМ.

3. Выделенный фрагмент текстового материала, который можно удалить, переместить или выполнить над ним некоторые преобразования.

4. В **вычислительных сетях**: последовательность передаваемых данных, рассматриваемая как единое целое. Обычно имеет **адрес и контрольную сумму**, позволяющую обнаруживать ошибки, и другие служебные элементы, требуемые соответствующим **протоколом** канала связи.

⁵ Подробнее см.: Словарь по кибернетике [4].

⁶ Подробнее о типах записей в базах данных см. [722].

ПАКЕТ [packet]

Единица информации в сети передачи данных. Пакет имеет строго определенную структуру. Она предусматривает наличие заголовка, который содержит адреса получателя и отправителя, данные для контроля ошибок, а также самого сообщения или его части, поскольку передаваемое сообщение (например текстовый файл) может быть разделено и пересылаться в виде последовательности пакетов (см. также **“Фрейм”**).

ФАЙЛ [file]

1. Идентифицированная последовательность или множество записей одно-типных **информационных элементов**.

2. Поименованная целостная совокупность данных на машиночитаемом носителе;

3. Поименованная область **внешней памяти** ЭВМ.

4. В английском языке: картотека, архив, комплект, подшивка и т. п. (см. также **“Массив”**).

Файл — основной структурный элемент хранения данных в ЭВМ, обеспечивающий возможность машине и человеку отличать один набор данных от другого при их поиске, изменении, удалении или выполнении с ними других операций.

Содержанием файла может стать одна поименованная логическая запись или несколько логических записей, объединенных общим именем. Например, файл может содержать записи некоторого множества библиографических описаний, каждое из которых по своей сути является логической записью, объединенной с другими каким-либо общим признаком (вид документа, тематика и т. п.). В других случаях содержанием записи в файле могут быть отдельные документы, управляющие работой ПЭВМ, программы или их части. **“Идентификация”** или **“поименование”** указанных записей или групп записей выражается присвоением каждому из них **“уникального”** или отличного от других имени, позволяющего автоматизированным средствам находить их в массивах других записей, в том числе файлов.

Наименования файлов состоят из двух частей:

1) **основного имени файла [basic file name (filename)]**, отражающего в краткой форме содержание и/или назначение находящихся в нем данных;

2) **расширения имени файла [file name broadening]**, которое характеризует вид записанных данных и их организацию (см. также **“Формат файла”**). Расширение отделяется от основного имени файла точкой и записывается в форме кода, принятого для каждого вида формата файлов. Код расширения, как правило, содержит от двух до четырех буквенных или буквенно-цифровых символов, например: **“*.aiff”, “*.au”, “*.exe”, “*.doc”, “*.txt”, “*.tif”, “*.jpg”, “*.wav”, “*.3ds”** и т. п.

На физическом уровне содержимое файла может быть не структурировано и представлять собой единственную физическую запись, или структурировано, например **полями данных**. В последнем случае оно будет включать соответствующее множество физических записей.

Некоторые типы файлов⁷

• **Файл данных [data file]** — файл, содержащий данные в отличие от **“программных”** файлов, содержащих записи программ или их частей.

⁷ О других типах файлов см. Толковый словарь по информатике В.И. Першикова и В.М. Савинкова [265].

- **Файл регистрации [log file]** — файл, в котором хранятся записи о других файлах, например, если какие-либо файлы были архивированы (см. “Архивация”), список имен этих файлов может храниться в файле регистрации.

- **Файл с произвольным доступом [random-access file]** — файл, в котором любая запись может быть считана, записана или изменена без необходимости считывания других записей. Произвольный доступ должен быть обеспечен **операционной системой** и реализован специальной машинной программой.

- **Основной файл [master file]** — файл, являющийся основным источником данных для решения определенного класса задач, для какой-либо цели или назначения. Поддержка этого файла в актуальном состоянии, при котором содержащиеся в нем данные отвечают требованиям новизны и точности, производится с использованием операций обновления файла.

- **Упорядоченный файл [sequential file]** — файл, в котором записи упорядочены по ключевому полю для ускорения доступа к определенной записи.

- **Последовательный файл [serial file]** — файл, в котором записи не упорядочены; поэтому чтобы прочитать нужную запись, требуется прочитать все предыдущие.

- **Программный файл [program file]** — файл, содержащий программу или ее часть, написанную на каком-либо из языков программирования.

- **Исполняемый файл [executable file]** — **программный файл** (см. ранее), предназначенный для запуска **операционной системой**. К исполняемым файлам относятся: командный файл, имеющий расширение “*.bat”, и машинный файл, имеющий **расширение имени** (см. ранее) “*.com” или “*.exe”.

- **Скрытый файл [hidden file]** — файл, который не отображается на экране **монитора** при просмотре каталога файлов (директории). Скрытым можно сделать любой файл с целью затруднения его нахождения, чтения и/или порчи другими лицами, которые могут иметь **доступ** к ЭВМ и др. [27, 265, 369, 722].

МАССИВ [array]

1. Упорядоченная структура множества документов или данных одного типа.

2. Поименованная совокупность однотипных (логически однородных), упорядоченных по индексам записей ИЭ.

3. Упорядоченное множество элементов одного типа. Каждый элемент массива должен иметь имя (идентификатор, индекс), обеспечивающее возможность его нахождения. Элементы массива могут быть одномерными и многомерными.

Применительно ко многим задачам автоматизированной обработки данных термины **массив** и **файл** могут использоваться как синонимы. Однако соответствующие понятия имеют и отличия: так, понятие **массив** не обязательно связано с записями информационных элементов (включая полнотекстовые документы) на машиночитаемых носителях. Понятие же **файл** в русскоязычной практике, как правило, предполагает указанное условие. О видах массивов см. [722].

Информационно-поисковый массив, поисковый массив [information collection, file] — массив документов или данных (соответствующие англ. эквиваленты — **document collection, data collection**), в котором производится **информационный поиск** [27, 265, 369].

АРХИВ [archives]

Организованная совокупность **массивов** данных или программ, длительно хранимых на внешних машиночитаемых носителях (например на гибких магнитных дисках, магнитных лентах и/или CD-ROM) с целью обеспечения возможности их дальнейшего использования.

Целями создания архивов являются:

1. Создание страховочных копий информационных и программных продуктов на случай их утраты или порчи в ходе эксплуатации вычислительных средств;

2. Освобождение внешней памяти ЭВМ (например **накопителя на жестком магнитном диске**) от программ и данных, потребность в оперативном использовании которых частично, полностью или временно отпала.

КЛЮЧ [key]

1. **Информационный элемент**, однозначно идентифицирующий запись или указывающий ее местоположение. Ключ может также служить средством для идентификации некоторого множества (в том числе массива) записей и располагаться в одном или нескольких из его полей.

В качестве ключа может служить **элемент данных** (в том числе **индекс, адрес, код** и др.) или группа элементов данных. Так, в **индексно-последовательном файле** ключ является обязательным элементом записи в каждом поле данных. Набор значений некоторой совокупности **атрибутов** в **реляционной модели данных** служит ключом, который однозначно идентифицирует **кортеж** или группу кортежей конкретного **отношения**.

2. Параметр шифрования, представляющий один из возможных вариантов шифра. Для дешифрования помимо ключа необходимо знать последовательность операций или правило его использования (см. также **“Алгоритм”**).

3. Значение, используемое для подтверждения полномочий на **доступ** к некоторой информации (в том числе к базам данных, отдельным файлам и т.п.), программным и/или техническим средствам. См. также значение **“Электронный ключ”** в разделе 1.6.3.

Единицы измерения количественных показателей записей данных в ЭВМ

БИТ [bit – от англ. binary digit]

Простое двоичное число (цифра или символ), принимающее значения 1 или 0 и служащее для записи и хранения данных в ЭВМ. Бит является минимальной двоичной единицей измерения **энтропии** и **количества информации** в ЭВМ, соответствующей одному **двоичному разряду**. Энтропия сообщения, выраженная в битах, определяется средним числом символов, необходимых для записи этого сообщения. Определенное количество бит составляет размер других единиц — **двоичных слов**, в том числе — **байта, килобайта, мегабайта** (см. далее) и т.д.

Байт [byte] — двоичное слово, способное записывать и хранить в памяти ЭВМ один буквенно-цифровой или другой символ данных. Каждый символ записывается в виде набора двоичных цифр (битов) при помощи определенного кода, например **ASCII**. Количество бит в байте определяет его разрядность, которая может составлять 8, 16, 32 и т.д. Соответственно байт называют 8-разрядным, 16-разрядным и т.д. Один 8-разрядный байт может определять 256 разных значений, например десятичных чисел от 0 до 256. Увеличение разрядности ведет к соответствующему увеличению числа возможных вариантов комбинаций, кодируемых одним байтом. Например, 16-разрядным — до 65536 или 2^{16} , 32-разрядным — до 2^{32} и т.д.

Килобайт, Кбайт [kilobyte] — единица измерения емкости памяти или длины записи, равная 1024 байтам. Часто под килобайтом понимается также величина, равная 10^3 байт.

Мегабайт, Мбайт [megabyte] — единица измерения емкости памяти или длины записи, равная 1024 Кбайт. Часто под мегабайтом понимается также величина, равная 10^3 килобайт или 10^6 байт.

Гигабайт, Гбайт [gigabyte] — единица измерения емкости памяти или длины записи, равная 1024 Мбайт. Часто под гигабайтом понимается также величина, равная 10^3 мегабайт, 10^6 килобайт или 10^9 байт.

Терабайт, Тбайт [terabyte] — единица измерения емкости памяти или длины записи, равная 1024 Гбайт. Часто под терабайтом понимается также величина, равная 10^3 гигабайт, 10^6 мегабайт, 10^9 килобайт или 10^{12} байт.

Кубит [quantum bit, qubit] — “Квантовый бит”: мера и измерения объема памяти в теоретически возможном виде компьютере, использующем квантовые носители, например — спины⁸ электронов. Кубит может принимать не два различных значения (“0” и “1”), а несколько, соответствующих нормированным комбинациям двух основных состояний спина, что дает большое число возможных сочетаний. Так, 32 кубита могут образовать около 4 млрд состояний [1027].

ЗНАК [charter]

Один символ, который может быть представлен и воспринят ЭВМ. К знакам относятся буквы, цифры, пробелы, знаки препинания, специальные символы (например математические, кодовые и т. п.).

1.4.3. Структуры, модели данных и связанные с ними термины

СТРУКТУРА [structure]

Фиксированное упорядоченное множество объектов и связей между ними.

С понятием “структура” связаны следующие термины:

- **структура данных [data structure]** — множество элементов данных, объединенных и упорядоченных определенным образом;

- **структура информационной базы [information support structure]** — упорядоченная по определенным правилам совокупность подмножеств записей информационных элементов, образующих информационную базу, и необходимых для реализации функций автоматизированной системы;

- **структура базы данных [DB structure]** — принцип или порядок организации записей в базе данных и связей между ними. Структуру БД принято рассматривать на разных уровнях абстракции (представления) и, в частности 1) концептуальном (с позиции администратора предприятия), 2) реализации или внешнем (с позиций конечного пользователя и прикладного программиста), 3) физическом или внутреннем (с позиций системного аналитика и системного программиста). Соответственно этим уровням различают концептуальную, внешнюю и физическую модели и/или схемы организации данных;

- **абстракция [abstraction]** — использование для описания или представления общих свойств объекта без конкретной его реализации (например, типов объектов — “читатель”, “фирма”, “автомобиль”, но не конкретно названных читателей, фирм, марок автомобилей и т. п.);

⁸ Спин (от англ. *spin* — вращаться, вертеться) — собственный момент количества движения элементарных частиц (например электрона, протона, нейтрона, нейтрино и т. д.), имеющих квантовую природу и не связанный с перемещением частицы как целого.

• **абстрактная структура данных [abstract logic design]** — структура данных, определенная функционально посредством выполняемых на ней операций. Такая структура не связана с поименованными типами объектов,

СПИСОК [list]

1. Структура данных, представляющая собой логически связанную последовательность записей — элементов списка;
2. Перечень различных объектов.

Связанные со “структурой” и “списком” понятия:

- **цепной список, связанный список [chained (linked) list]** — список данных, в котором порядок элементов списка задан посредством указателей, включенных в их запись;
- **подсписок [sublist]** — ветвь списковой **структуры**, представляющая собой отдельный **цепной список**, на который имеется указатель от одного из элементов цепного списка вышестоящего уровня иерархии в данной структуре;
- **цепная структура [chain structure]** — структура цепного списка;
- **ассоциативная (ассоциативно-адресная) структура [associative structure]** — совокупность **цепного списка** и всех, связанных с ним **подсписков**. Различают объектные и признаковые списковые структуры;
- **однородная структура** — структура, состоящая из однотипных элементов;
- **комбинированная структура [combined structure]** — структура данных, полученная путем объединения (композиции) нескольких исходных структур.

МОДЕЛЬ [model]

1. Результат корректного воспроизведения каким-либо способом или средствами различных **объектов** (в том числе процессов и явлений реального мира или мыслительной деятельности человека). Модели являются, с одной стороны, продуктом изучения свойств соответствующих объектов, процессов и явлений предметной области, с другой — служат инструментом для углубления **знаний** о них, а также решения разнородных прикладных задач (см. далее также **“Моделирование”**). В зависимости от характера средств, используемых для построения (создания) **“моделей”** последние подразделяются на описательные, математические, физические и комбинированные (например, физико-математические модели). Различают также статические и динамические модели (в том числе кибернетические модели) и др.

2. Тип, марка, образец конструкции (например, модель автомобиля — “ВАЗ 21099”).

3. Образец для подражания, образцовый экземпляр какого-либо изделия.

4. Оригинал, который служит для снятия копии, изображения или создания другого произведения, имеющего признаки сходства с ним (примерами могут служить неодушевленные предметы и люди, в том числе — **“фотомодели”**, **“натюршики”** в живописи и т. п.).

МОДЕЛЬ ДАННЫХ [data model]

1. Представление данных и их взаимосвязей (отношений), описывающих понятия проблемной среды. Модели данных используются для представлений структур данных на концептуальном и внешнем уровнях, но не физическом (см. монографию Т. Тиори. Дж. Фрая [28]). Понятие **модель данных** связано с их **логической структурой**.

2. Совокупность правил порождения **структур данных** в БД и выполнения операций над ними.

3. Формализованное описание структур данных и операций над ними.

МОДЕЛИРОВАНИЕ [modeling]

Методология выполнения экспериментальных работ путем исследования свойств различных **объектов** на их **моделях** (см. ранее). Виды моделирования различаются в зависимости от целей его выполнения, характера исследуемых объектов и выбранных для исследования средств (о некоторых видах моделирования см. в Словаре по **Кибернетике** [4]). В связи с развитием вычислительной техники наиболее широкое применение в различных областях человеческой деятельности получило математическое моделирование.

Математическое моделирование [mathematical modeling] — процесс построения и исследования в динамике поведения математических моделей различных процессов, явлений и физических объектов с использованием средств вычислительной техники. В основе математического моделирования лежит использование фундаментальных законов естествознания и конкретных наук, связанных с целями и предметами моделирования. На основании этих законов разрабатывается математический аппарат, описывающий исследуемые явления и объекты, который преобразуется в соответствующий алгоритм и программу для реализации его на ЭВМ. В зависимости от общих целей математического моделирования полученные в ходе «испытания» математической модели данные могут использоваться для принятия определенных решений, в частности, для выбора альтернативных вариантов поведения специалистов или для уточнения исходной математической модели (например, в виде поправок, вводимых в математический аппарат, и таким образом использоваться в качестве средства ее совершенствования). Методы математического моделирования нашли широчайшее применение при построении так называемых **интеллектуальных (экспертных) систем, систем автоматизированного проектирования** и др. Об использовании математического моделирования в области трехмерной компьютерной графики и анимации, а также реализующих его средствах программного обеспечения см. [503, 504].

Некоторые общие термины, связанные с моделированием и моделями данных:

- **логическая структура [logical structure]** — представление логической организации данных в виде множества типов записей данных и связей между ними;
- **структурная модель данных [structured data model]** — модель данных, представленная в виде структуры — множества типов данных и связей между ними. Различают три основных вида структурной модели (логической структуры) организации данных: иерархической, сетевой и реляционной;
- **связь (между данными) [binding, link, relationship]** — установленный характер взаимозависимости данных в различных информационных моделях и структурах данных. Связи между данными идентифицируются видом связи и направлением. Характер вида и направления связи могут отражаться ее именем — **указателем связи** (поименованная связь, характеристика отношения). Различают **иерархическую** или **вертикальную, горизонтальную, ассоциативную, логическую, двунаправленную связи** и др.;
- **иерархическая (вертикальная) связь [hierarchical binding, vertical binding]** — вид связи, устанавливаемый между данными, находящимися на разных

уровнях иерархической структуры (например, связи: “род–вид”, “вид–подвид”, “отец–сын” и т. п.);

- **горизонтальная связь [horizontal binding]** — вид связи, устанавливаемый между данными, находящимися на одном уровне **иерархической структуры**;
- **логическая связь [logical relationship]** — вид связи, устанавливаемый между типами данных в иерархической и сетевой моделях данных (см. далее), в отличие от связи между конкретными экземплярами данных в **базах данных**;
- **ассоциативная связь [associative link]** — вид связи, устанавливаемый исходя из заданного сочетания признаков данных или **информационных элементов**, которые образуют таким путем упорядоченные последовательные цепочки. Указатели на связанные данные (**адреса связи**) могут размещаться в самих данных или программных средствах управления базой данных (см. также “СУБД” и “Гипертекст”);
- **двунаправленная связь [bidirectional binding]** — совокупность связей в прямом и обратном направлении.

ИЕРАРХИЧЕСКАЯ МОДЕЛЬ (СТРУКТУРА) ДАННЫХ [hierarchical data model]

Модель организации данных, представляющая собой древовидный **граф**, состоящий из ряда типов записей (**типов данных**) и связей между ними (**отношений** или **характеристик отношений**), причем один из типов записей определяется как **корневой** или **входной**, а остальные связаны с ним или друг с другом отношениями “**один-ко-многим**” или (реже) “**один-к-одному**”. При этом запись, идентифицируемая элементом “один”, рассматривается как исходная, а соответствующая элементу “многие”, как порожденная. Каждая запись может быть порожденной только в одной связи, следовательно ей соответствует только одна исходная запись. Однако каждая запись может быть исходной во многих связях. Корневая запись может быть только исходной. Пример иерархической модели приведен на рис. 1.1.

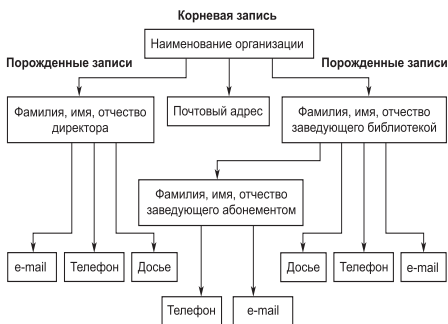


Рис. 1.1. Принцип построения иерархической модели организации БД

Граф [graph от греч. *grapho* — пишу, изображаю] — графическое представление математической модели системы связей между объектами любой природы. Объекты задаются в графе точками — **вершинами**, связи — линиями, соединяющими вершины, которые называются **ребрами** или **дугами** графа. Каждое ребро может быть ориентированным (т.е. иметь определенное направление от одной вершины к другой) либо неориентированным (двунаправленным). Ребро, соединяющее вершину с нею самой, называется **петлей**. Вершины, которым не соответствует ни одно ребро, являются **изолированными**. Число ребер, соединяющих две фиксированные вершины, может быть произвольным, поскольку оно определяется количеством и характером связей между соответствующими этим вершинам объектами. В то же время каждому ребру соответствует не более двух вершин. Раздел математики, изучающий свойства различных геометрических схем, образованных множеством точек и соединяющих их линий (графов), называется **Теорией графов**. Подробнее см. [4, С. 142–145].

СЕТЕВАЯ МОДЕЛЬ (СТРУКТУРА) ДАННЫХ [network data model]

Модель организации данных, подобная иерархической, но отличающаяся от нее тем, что каждая запись может вступать в любое количество поименованных связей с другими записями как исходная или порожденная, или как то и другое (см. “**Двунаправленная связь**”). Примеры (1 и 2) сетевой модели приведены на рис. 1.2.

РЕЛЯЦИОННАЯ МОДЕЛЬ (СТРУКТУРА) ДАННЫХ [relation data model]

Реляционная модель данных была предложена в 1969 г. сотрудником фирмы IBM **Е. Ф. Коддом**. Она представляет собой набор **плоских файлов** — таблиц, называемых **отношениями**, к которым применимы операции **реляционной алгебры** для реализации автоматизированного ответа на запросы пользователей системы⁹.

Потенциально в реляционной модели может быть организовано очень большое количество связей между данными, значительная часть которых является избыточными (т.е. не используемыми). Поэтому разработаны формы (варианты) нормализации отношений: первая (**1НФ**), вторая (**2НФ**), третья (**3НФ**) и четвертая (**4НФ**).

Примеры реализации реляционной модели приведены на рис. 1.3 и 1.4. В настоящее время существует достаточно большое число различных вариантов построения реляционных моделей. Одной из них является **постреляционная модель**. Подробнее см. [28, 29, 552].

Термины, связанные с реляционной моделью данных:

• Отношение [relation]

1. Форма связи между объектами (в рассматриваемом случае — разными типами данных или атрибутами), отражающая то общее, что их объединяет.

2. Два математических выражения, связанных знаком операции сравнения (“=”, “>”, “<” и т.п.).

3. Таблица реляционной модели данных.

4. Заданное подмножество *n*-ой декартовой степени некоторого множества.

См. также “**Отношения**” в разделе 2.4.2.

⁹ Весьма распространенной ошибкой является употребление термина **реляционная БД** по отношению к любым массивам данных, имеющих табличную форму организации, однако не обеспечивающим выполнение указанного условия.

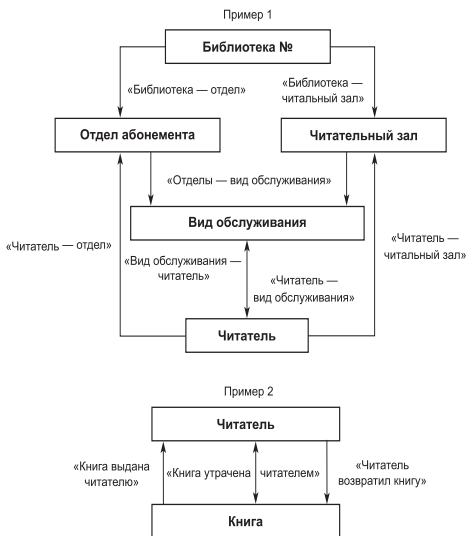


Рис. 1.2. Принцип построения сетевой модели организации БД

• **Отношения в реляционной модели данных** включают **атрибуты** и **кортежи** (см. далее), составляющие соответственно столбцы и строки таблицы. Количество атрибутов в отношении соответствует количеству содержащихся в нем элементов данных, количество кортежей — числу экземпляров (реализаций) записей. Порядок следования атрибутов и кортежей может быть произвольным. Значения атрибутов определяются из доменов. Наличие доменов обеспечивает связи между разными отношениями.

• **Реляционная алгебра [relational algebra]** — язык для описания операций над отношениями. Основные операции реляционной алгебры: проекция, соединение, пересечение и объединение.

• **Сущность [entity]** — то же, что **тип объекта**: обобщенное наименование множества однотипных объектов, называемых **экземплярами**. Каждый экземпляр обладает набором свойств — **атрибутов** (см. далее), отличающих его от всех остальных. Примеры сущности: автомобиль, самолет, врач и т.д. Примеры экземпляров: "автомобиль ВАЗ-2106", "самолет Ту-104", "врач Петров Иван Федорович" и т.д.