

*УЧЕБНИК
ДЛЯ ВУЗОВ*

ПИТЕР®

И. П. Карпова



Базы данных

**ДОПУЩЕНО УЧЕБНО-МЕТОДИЧЕСКИМ ОБЪЕДИНЕНИЕМ
ВУЗОВ В КАЧЕСТВЕ УЧЕБНОГО ПОСОБИЯ**

И. П. Карпова
Базы данных. Учебное пособие

Серия «Учебник для вузов»

Рецензенты:

д. ф.-м. н., профессор РГГУ О. М. Аншаков;
нач. лаб. инфокоммуникационных технологий ИТЭФ, к. ф.-м. н.,
зам. рук. кафедры информатики ФНБИК МФТИ Ю. А. Семенов

Руководитель проекта
Ведущий редактор
Научный редактор
Литературный редактор
Художник
Корректор
Верстка

А. Кривоц
Ю. Сергиенко
А. Пасечник
Е. Пасечник
М. Кольцов
В. Листова
Е. Волошина

ББК 32.973.233-018я7
УДК 004.65(075)

Карпова И. П.

K26 Базы данных. Учебное пособие. — СПб.: Питер, 2020. — 240 с.: ил. — (Серия «Учебник для вузов»).

ISBN 978-5-4461-1607-2

В учебном пособии излагается лекционный материал по курсу «Базы данных», а также материал для практических занятий (семинаров и лабораторных работ) и курсового проектирования, в рамках которого создается проект базы данных по произвольной предметной области. Изложение снабжено большим количеством примеров и конкретных рекомендаций по применению получаемых знаний в процессе создания и эксплуатации баз данных.

Содержание учебного пособия соответствует программе курса «Базы данных» по специальности 230100 «Информатика и вычислительная техника» и может быть рекомендовано в качестве учебного пособия для студентов технических факультетов, изучающих автоматизированные информационные системы и системы управления базами данных.

16+ (В соответствии с Федеральным законом от 29 декабря 2010 г. № 436-ФЗ.)

ISBN 978-5-4461-1607-2

© ООО Издательство «Питер», 2020

© Серия «Учебник для вузов», 2020

Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав.

Изготовлено в России. Изготовитель: ООО «Прогресс книга».
Место нахождения и фактический адрес: 194044, Россия, г. Санкт-Петербург,
Б. Сампсониевский пр., д. 29А, пом. 52. Тел.: +78127037373.

Дата изготовления: 11.2019. Наименование: книжная продукция. Срок годности: не ограничен.

Импортер в Беларусь: ООО «ПИТЕР М», РБ, 220020, г. Минск, ул. Тимирязева, д. 121/3, к. 214, тел./факс 208 80 01.

Налоговая льгота — общероссийский классификатор продукции ОК 034-2014,
58.11.11.000 — Учебники печатные общеобразовательного назначения.

Подписано в печать 20.11.19. Формат 70х100/16. Бумага офсетная. Усл. п. л. 19,350.

Заказ 0000

Отпечатано в цифровой типографии «Галерея печати «ИПК НП-Принт».
190005, Санкт-Петербург, Обводный канал, 199.

Оглавление

Список используемых сокращений	10
Введение	11
1. Основные сведения	15
1.1. Информация, данные, знания. Терминология	16
1.2. Автоматизированная информационная система	18
1.3. Предметная область информационной системы	20
1.4. Назначение и основные компоненты системы баз данных	24
1.5. Уровни представления данных	25
2. Основные модели данных	28
2.1. Понятие модели данных	28
2.1.1. Типы структур данных.....	29
2.1.2. Операции над данными	31
2.1.3. Ограничения целостности	32
2.2. Сетевая модель данных (СМД)	33
2.3. Иерархическая модель данных (ИМД)	35
2.4. Реляционная модель данных (РМД)	38
2.4.1. Понятие отношения.....	38
2.4.2. Свойства отношений	39
2.4.3. Достоинства и недостатки РМД.....	42
2.5. Объектно-реляционная модель данных	43
2.6. Объектно-ориентированная модель данных	44
3. Введение в язык SQL	46
3.1. Операции реляционной алгебры	46
3.1.1. Основные операции реляционной алгебры	47
3.1.2. Вспомогательные операции реляционной алгебры	48
3.2. Общие сведения о языке SQL	50
3.3. Создание таблиц	51

3.4. Команды модификации данных	54
3.5. Извлечение данных из таблиц	57
3.5.1. Синтаксис команды SELECT	57
3.5.2. Операторы и предикаты	61
3.5.3. Функции агрегирования	65
3.5.4. Запрос SELECT на нескольких таблицах	68
3.5.5. Подзапросы	70
3.5.6. Самосоединение	74
3.5.7. Замечания по использованию NULL-значений	75
3.5.8. Оператор CASE	76
3.5.9. Работа с представлениями	77
3.6. Удаление объектов базы данных	80
Задания для практических занятий	81
4. Системы управления базами данных	84
4.1. Классификация СУБД	85
4.2. Правила Кодда для реляционной СУБД (РСУБД)	86
4.3. Основные функции реляционной СУБД	88
4.4. Администрирование базы данных	89
4.5. Словарь-справочник данных	90
5. Физическая организация баз данных	91
5.1. Механизмы среды хранения и архитектура СУБД	91
5.2. Структура хранимых данных	93
5.3. Управление пространством памяти и размещением данных	94
5.4. Виды адресации хранимых записей	97
5.5. Способы размещения данных и доступа к данным в РБД	98
5.5.1. Способы доступа к данным	99
5.5.2. Индексирование данных	100
5.5.3. Хеширование	107
5.5.4. Кластеризация данных	110
6. Многопользовательский доступ к данным	114
6.1. Механизм транзакций	114
6.2. Взаимовлияние транзакций	118
6.3. Уровни изоляции транзакций	120
6.4. Блокировки	121
6.5. Временные отметки	123
6.6. Многовариантность	124
7. Защита данных в базах данных	125
7.1. Обеспечение целостности данных	125
7.2. Обеспечение безопасности данных	126
7.2.1. Виды сбоев	127

7.2.2. Средства физической защиты данных	128
7.2.3. Восстановление базы данных.....	128
7.3. Защита от несанкционированного доступа	130
8. Оптимизация реляционных запросов	134
8.1. Этапы оптимизации запросов в реляционных СУБД	134
8.2. Преобразования операций реляционной алгебры	137
8.3. Методы оптимизации	140
8.3.1. Метод оптимизации, основанный на синтаксисе	140
8.3.2. Метод оптимизации, основанный на стоимости	141
8.3.3. Другие возможности управления оптимизацией	144
8.3.4. Примеры использования методов оптимизации запросов	145
8.4. Настройка приложений	147
9. Проектирование баз данных	151
9.1. Требования к проекту базы данных	151
9.2. Этапы проектирования базы данных	152
9.3. Инфологическое проектирование	158
9.3.1. Метод «сущность–связь».....	160
9.3.2. Объединение локальных представлений	161
9.4. Определение требований к операционной обстановке	163
9.5. Выбор СУБД и инструментальных программных средств	163
9.6. Логическое проектирование БД	164
9.7. Физическое проектирование БД	165
9.8. Автоматизация проектирования БД	165
9.9. Особенности проектирования реляционных БД	166
9.9.1. Правила преобразования ER-диаграммы в схему БД	166
9.9.2. Выявление нереализуемых связей	168
9.9.3. Определение первичных ключей	169
9.9.4. Определение типов данных атрибутов	169
9.9.5. Описание ограничений целостности	170
9.9.6. Аномалии модификации данных.....	172
9.9.7. Нормализация отношений.....	172
9.9.8. Денормализация отношений.....	179
9.10. Пример проектирования реляционной базы данных	180
9.10.1. Инфологическое проектирование	180
9.10.2. Определение требований к операционной обстановке.....	183
9.10.3. Выбор СУБД и других программных средств	184
9.10.4. Логическое проектирование реляционной БД	184
9.10.5. Реализация проекта базы данных.....	199
10. Перспективы развития технологии баз данных	207

Приложение 1. Примеры заданий для лабораторных работ	211
Приложение 2. Варианты заданий на курсовое проектирование	227
Список литературы	233
Алфавитный указатель	235

Введение

Технология баз данных появилась почти полвека назад и с тех пор не только оказала огромное влияние на развитие информационных технологий, но и кардинально изменила методы работы многих организаций и предприятий. В наше время сложно найти такую компанию, в которой не использовались бы базы данных, точнее, информационные системы, основанные на базах данных. Широкое распространение этих систем, дружелюбность интерфейса и кажущаяся простота использования привели к тому, что за создание и эксплуатацию баз данных часто берутся люди, не имеющие достаточных специальных знаний в этой области, что влечет появление большого числа неэффективных, сложных в поддержке и сопровождении систем.

С другой стороны, существует большое количество учебных изданий, посвященных как отдельным вопросам теории и практики использования баз данных, так и претендующих на охват технологии БД в целом (или, по крайней мере, большей части основных аспектов этой технологии), отличающиеся большим объемом и подробным изложением теории — это литература, рассчитанная на профессионалов. Кроме того, есть целый ряд книг, которые условно можно отнести к серии «база данных за два дня». Такие издания не способствуют увеличению количества специалистов в области баз данных. При этом практически отсутствуют учебники, дающие достаточно полное введение в технологию баз данных, не опирающиеся на какую-то конкретную СУБД и сохраняющие компактность изложения. Данное учебное пособие призвано восполнить этот пробел.

Учебное пособие основано на материалах лекций и практических занятий, которые в течение многих лет проводятся автором на факультете информационных технологий и вычислительной техники Московского института электроники и математики и на факультете nano-, био-, информационных и когнитивных технологий Московского физико-технического института.

Предлагаемое вашему вниманию учебное пособие имеет ряд отличий от существующих изданий аналогичной тематики:

1. Содержание пособия соответствует Федеральному государственному образовательному стандарту по направлению подготовки 230100 «Информатика и вычислительная техника» (квалификация «бакалавр»).

2. Все определения, используемые в пособии, взяты из ГОСТа и международных стандартов (ANSI/ISO).
3. Теория баз данных излагается только в том объёме, который необходим как основа, база для понимания особенностей технологии БД. Изложение теоретических вопросов сопровождается большим количеством примеров, что облегчает восприятие материала.
4. Все аспекты технологии БД рассматриваются с точки зрения их применения на практике.
5. Отсутствует ориентация на какую-либо конкретную систему управления базами данных (СУБД), при этом даётся представление о возможностях СУБД в целом. Это позволит студенту (специалисту) в будущем оценивать различные системы с точки зрения их соответствия требованиям, предъявляемым к СУБД, и выбирать в каждой конкретной ситуации ту систему, которая в наибольшей степени соответствует условиям решаемой задачи.
6. Подробно излагается оптимизация реляционных запросов, в том числе приёмы написания оптимальных запросов. Под оптимальными в данном случае понимаются такие запросы, по которым СУБД сможет построить более эффективный план выполнения запроса. Рациональность этих приёмов подтверждается, в том числе, и личным опытом автора, полученным во время работы с реальными базами данных большого объёма.
7. Пособие ориентировано на учебный курс, предполагающий выполнение лабораторных работ и курсового проекта, и содержит примеры заданий. Это позволяет рекомендовать его преподавателям не только как основу лекционного курса, но и для использования на практических занятиях и в ходе курсового проектирования.

Материал книги разбит на главы, каждая из которых соответствует одной теме. Каждая глава представляет собой более или менее законченный фрагмент, который можно изучать отдельно от других глав.

В главе 1 приводятся определения основных терминов, а также общие сведения о базах данных, уровнях представления данных и автоматизированных информационных системах, которые основаны на данных.

Глава 2 посвящена описанию моделей данных. Основное внимание уделяется реляционной модели, но также приводятся сведения и о других моделях: *о сетевой, иерархической, объектно-реляционной и объектно-ориентированной*. Проанализированы достоинства и недостатки этих моделей.

Главу 3 занимает введение в язык SQL — структурированный язык запросов к базам данных. Синтаксис команд и примеры использования соответствует стандарту SQL-92, который поддерживается большинством реляционных СУБД. В этой главе рассматриваются команды языка определения данных (Data Definition Language, DDL) и языка манипулирования данными (Data Manipulation Language, DML). Большая часть главы посвящена возможностям команды SELECT: использованию операторов, предикатов,

функций агрегирования, запросам на нескольких таблицах, подзапросам, самосоединению таблиц и работе с представлениями.

Глава 4 касается общих вопросов организации систем управления базами данных (СУБД). Приводится классификация СУБД, требования к реляционным СУБД (по Кодду), описание основных функций реляционной СУБД и задач по администрированию баз данных.

Глава 5 посвящена физической организации баз данных, без знания которой невозможно организовать эффективно работающую БД. Рассматриваются механизмы среды хранения, структура хранимых данных и методы управления пространством памяти и размещением данных. Основное внимание уделяется изучению способов размещения данных и доступа к данным в реляционной БД, а именно: *индексированию*, *хешированию* и *кластеризации*. Для каждого из этих способов приводится не только назначение и описание принципов организации, но и правила использования.

В главе 6 представлены сведения об организации многопользовательского доступа к данным. Описывается механизм транзакций (ACID-транзакции), уровни изоляции транзакций (по стандарту ANSI/ISO), реализация механизма транзакций с помощью блокировок и временных отметок. Приводятся команды управления транзакциями и блокировками, входящие в состав языка SQL.

Глава 7 посвящена вопросам защиты данных в базах данных. В части обеспечения логической целостности данных приводится перечень ограничений целостности, которые входят в состав стандарта языка SQL. В части обеспечения безопасности данных рассматривается защита данных от сбоев (журналы транзакций, сегменты отката, резервное копирование, восстановление базы данных). В части защиты от несанкционированного доступа описываются системные и объектные привилегии (права доступа), команды языка SQL для управления доступом и правила назначения прав доступа.

Глава 8 касается практических аспектов оптимизации реляционных запросов. Описываются основные методы оптимизации (по синтаксису и по стоимости), а также другие возможности управления оптимизацией. Материал включает описание правил настройки приложений, приёмов написания эффективных запросов и примеры использования этих приёмов.

В главе 9 излагается методология проектирования баз данных, в первую очередь реляционных. Приводится описание последовательности этапов проектирования автоматизированных информационных систем, основанных на базах данных: предпроектная подготовка; собственно проектирование БД; реализация (создание БД и ППО). Основное внимание уделяется этапам инфологического, логического и физического проектирования баз данных. Описывается применение метода сущность—связь (entity—relation method, ER-метод) — это самый распространённый метод проектирования реляционных баз данных. Рассматриваются правила проектирования реляционных БД, в том числе правила преобразования ER-диаграммы в схему БД, определение первичных ключей и типов данных атрибутов, нормализация отношений (до

4-й нормальной формы) и денормализация. Описание сопровождается примером, иллюстрирующим полный цикл проектирования реляционной базы данных от описания предметной области до разработки стратегии резервного копирования.

В главе 10 рассматриваются те перспективные направления развития технологии баз данных, которые остались за рамками данного учебного пособия, но могут быть интересны читателю для дальнейшего изучения.

К главам 1–3, 8 и 9 прилагаются примеры заданий для практических занятий. В приложениях приведены примеры заданий для лабораторных работ и примеры заданий для курсового проекта.

Таким образом, материал данной книги охватывает наиболее важные вопросы проектирования и эксплуатации баз данных и ориентирован на практическое использование данной технологии.

От издательства

Ваши замечания, предложения, вопросы отправляйте по адресу электронной почты comp@piter.com (издательство «Питер», компьютерная редакция).

Мы будем рады узнать ваше мнение!

На веб-сайте издательства <http://www.piter.com> вы найдете подробную информацию о наших книгах.

1. Основные сведения

Цивилизация развивается за счёт расширения числа важных операций, которые можно выполнять, не думая о них.

*Альфред Норт Уайтхед,
британский математик, логик, философ*

Развитие средств вычислительной техники и информационных технологий обеспечило возможности для создания и широкого применения автоматизированных информационных систем (АИС) разнообразного назначения. Разрабатываются и внедряются информационные системы управления хозяйственными и техническими объектами, модельные комплексы для научных исследований, системы автоматизации проектирования и производства, всевозможные тренажеры и обучающие системы.

Различают АИС, основанные на знаниях, и АИС, основанные на данных. К первым можно отнести, например, экспертные системы (ЭС), интеллектуальные системы поддержки принятия решений (СППР) и т. п. Ко вторым — всевозможные прикладные системы, которые сейчас активно используются как на предприятиях, так и в учреждениях. Такие прикладные системы применяются очень широко, и в рамках данного курса наше внимание будет сосредоточено именно на системах, которые основаны на данных.

Существует две основные предпосылки создания таких систем:

1. Разработка методов конструирования и эксплуатации систем, предназначенных для коллективного использования.
2. Возможность собирать, хранить и обрабатывать большое количество данных о реальных объектах и явлениях, то есть оснащение этих систем «памятью».

Массив данных общего пользования в системах, основанных на данных, называется *базой данных*. База данных является моделью предметной области информационной системы.

На заре развития вычислительной техники обрабатываемые данные являлись частью программ: они располагались сразу за кодом программы в так называемом *сегменте данных* (рис. 1.1, а). Следующим шагом стало хранение данных в отдельных файлах (рис. 1.1, б). Недостатком этих подходов являлась зависимость программ от данных: сведения о структуре данных включались в код программы. При изменении структуры данных необходимо было вносить изменения в программу.

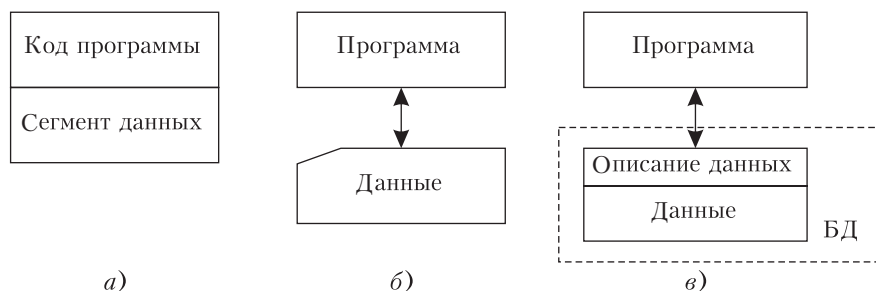


Рис. 1.1. Развитие принципов обработки данных

Логичным продолжением этой эволюции является перенос описания данных в массив данных (рис. 1.1, в). Это позволило обеспечить независимость данных от программ.

Основным принципом организации баз данных является совместное хранение данных и их описание.

Описание данных называют *метаданными*. Метаданные хранятся в части базы данных, которая называется *каталогом* или *словарём-справочником данных* (ССД). Зная формат метаданных, можно запрашивать и изменять данные без написания дополнительных программ.

Одна и та же база данных может быть использована для решения многих прикладных задач. Наличие метаданных и возможность информационной поддержки решения нескольких задач — это принципиальные отличия базы данных от любой другой совокупности данных, находящихся во внешней памяти ЭВМ.

1.1. Информация, данные, знания. Терминология

Информация — любые сведения о каком-либо событии, сущности, процессе и т. п., являющиеся объектом некоторых операций: восприятия, передачи, преобразования, хранения или использования.

Данные — это информация, зафиксированная в форме, пригодной для последующей обработки, передачи и хранения, например находящаяся в памяти ЭВМ или подготовленная для ввода в ЭВМ.

Подготовка информации состоит в её формализации, сборе и переносе на машинные носители.

Обработка данных — это совокупность задач, осуществляющих преобразование массивов данных. Обработка данных включает: ввод данных в ЭВМ, отбор данных по заданным критериям, преобразование структуры данных, перемещение данных на внешней памяти ЭВМ, вывод данных, являющихся результатом решения задач, в табличном или ином, удобном для пользователя виде.

Система обработки данных (СОД) — это набор аппаратных и программных средств, осуществляющих выполнение задач по управлению данными.

Управление данными — совокупность функций обеспечения требуемого представления данных, их накопления и хранения, обновления, удаления, поиска по заданному критерию и выдачи данных [5].

Предметная область (ПрО) — часть реального мира, подлежащая изучению с целью организации управления и, в конечном итоге, автоматизации.

База данных (БД) — совокупность данных, организованных по определённым правилам, предусматривающим общие принципы описания, хранения и манипулирования данными, независимая от прикладных программ [5]. Эти данные относятся к определённой предметной области и организованы так, что могут быть использованы для решения *многих задач многими пользователями*.

Ведение базы данных — деятельность по обновлению, восстановлению и изменению структуры базы данных с целью обеспечения её целостности, сохранности и эффективности использования [5].

Система управления базами данных (СУБД) — это совокупность программ и языковых средств, предназначенных для управления данными в базе данных, ведения базы данных и обеспечения взаимодействия её с прикладными программами [5].

Автоматизированная информационная система (АИС) представляет собой совокупность данных, экономико-математических методов и моделей, технических, программных средств и обслуживающих систему специалистов, предназначенную для обработки информации и принятия управленческих решений.

Банк данных (БнД) — это автоматизированная информационная система, включающая комплекс специальных методов и средств (математических, информационных, программных, языковых, организационных и технических) для поддержания динамической информационной модели предметной обла-

сти с целью обеспечения информационных запросов пользователей. В задачи *банка данных* входит:

- 1) обеспечение информационных потребностей внешних пользователей;
- 2) обеспечение возможности хранения и модификации больших объёмов многоаспектных данных;
- 3) обеспечение заданного уровня достоверности хранимых данных и их непротиворечивости;
- 4) ограничение доступа к данным только кругом пользователей с соответствующими полномочиями;
- 5) обеспечение поиска данных по произвольной группе признаков;
- 6) удовлетворение заданных требований по производительности при обработке запросов;
- 7) обеспечение выдачи пользователям данных в различной форме;
- 8) обеспечение простоты и удобства обращения внешних пользователей к данным.

Помимо этого банк данных должен допускать возможность реорганизации при изменении границ ПрО.

1.2. Автоматизированная информационная система

Под *автоматизированной информационной системой* (АИС) понимается совокупность программно-аппаратных средств, предназначенных для автоматизации деятельности, связанной с хранением, передачей и обработкой информации.

АИС, основанная на базе данных, служит для сбора, накопления, хранения информации, а также её эффективного использования для различных целей. Информация представляется в виде данных, хранимых в памяти ЭВМ. При проектировании АИС, с одной стороны, решается вопрос о том, какие сведения и для каких целей будут содержаться в системе, с другой — как соответствующие данные будут организованы в памяти ЭВМ и как они будут обрабатываться при эксплуатации АИС.

По сферам применения и правилам организации различают два основных класса АИС, основанных на базе данных: *информационно-поисковые* (ИПС) и *системы обработки данных* (СОД). ИПС ориентированы, как правило, на извлечение подмножества хранимых данных, удовлетворяющих некоторому поисковому критерию. Пользователя ИПС интересуют в основном сами извлекаемые из базы данных сведения, а не результаты их обработки. Примером ИПС является любая справочная служба: к ней обращаются с запросом и получают в результате те данные, которые удовлетворяют этому запросу.

Обращения пользователя к СОД чаще всего приводят к обновлению данных. Вывод данных может вовсе отсутствовать или представлять собой результат программной обработки хранимых сведений. Пример СОД — банковские системы, осуществляющие открытие/закрытие счетов, пересчёт вкладов в зависимости от процентов, приём/снятие сумм и т. п.

В зависимости от характера информационных ресурсов, с которыми имеют дело АИС, их подразделяют на *документальные* и *фактографические* системы. На практике используются также системы комбинированного типа.

Фактографические АИС хранят сведения об объектах предметной области, их свойствах и взаимосвязях. Сведения о каждом объекте могут поступать в систему из множества различных источников. Кроме поиска и модификации данных, фактографические системы поддерживают статистические функции (нахождение суммы, минимума, максимума и т. п.). Фактографические АИС обычно принадлежат к классу систем обработки данных.

В документальной системе объект хранения — документ, который содержит информацию, относящуюся к определённой предметной области. Это могут быть графические изображения (например, географические карты); информация на естественном языке (монографии, тексты законодательных актов, научные отчёты и т. п.); звуковая информация (например, мелодии — для системы, хранящей фонотеку) и т. д. Для обработки данных не важно, какие сведения хранятся в документах. Обычно (но не всегда) документальные АИС реализуются в виде информационно-поисковых систем (ИПС).

Основные компоненты документальной ИПС:

- ✦ программные средства;
- ✦ поисковый массив документов;
- ✦ средства поддержки информационного языка системы.

Программные средства ИПС служат для организации управления данными (ввода, хранения, защиты, поиска и выдачи). Поисковый массив документов в ИПС обычно называется базой данных. Он представляет собой набор ссылок на документы (или их описаний), хранящий основную информацию о документах и организованный так, чтобы обеспечить быстрый поиск документов. Описание документа зависит от предметной области и состоит из значений атрибутов, характеризующих содержание документа. Например, для БД географических карт это могут быть координаты и масштаб, а для БД законодательных актов — тип документа (закон, постановление и др.), дата принятия, область действия и т. п.

Информационный язык ИПС предназначен для того, чтобы пользователь мог запросить данные у системы. Системными средствами пользовательский запрос преобразуется в формальный запрос, понятный системе. Информационный язык ИПС может быть основан на подмножестве естественного языка, которое относится к обслуживаемой ПрО. Но чаще поиск документа осуществляется с помощью шаблонов — экранных форм, включаю-

щих поля описания документа. В эти поля вносятся конкретные значения, которые и определяют условия поиска документов. На рис. 1.2 приведён пример поиска через экранную форму в справочнике библиотек. В форме указаны два условия — округ и фрагмент названия библиотеки, и в результате поиска система выдала три записи, удовлетворяющие этим условиям.

Id	Номер	Название	Тип
259	43	им. Горького А.М.	
258	43	им. Горького А.М.	
248	23	им. Горького А.М.	Детская

Рис. 1.2. Пример поиска данных через экранную форму

Мы будем основное внимание уделять фактографическим АИС, имея в виду, что ИПС и документальные АИС создаются с помощью тех же программных средств и на тех же принципах, что и СОД, а специфические моменты обработки данных реализуются через приложения (программы, внешние по отношению к ядру СОД).

Разработка любой АИС начинается с определения предметной области.

1.3. Предметная область информационной системы

Предметная область (ПрО) информационной системы рассматривается как совокупность реальных процессов и объектов (*сущностей*), представляющих интерес для её пользователей [6]. Каждая из сущностей ПрО обладает определённым набором свойств (атрибутов).

Для упрощения процедуры описания ПрО в большинстве случаев прибегают к определению *типов сущностей*. Тип позволяет выделить из всего множества сущностей ПрО группу сущностей, однородных по структуре и поведению (относительно рамок рассматриваемой ПрО). Например, для ПрО «Институт» в качестве типов сущностей могут рассматриваться студенты, преподаватели, дисциплины и т. п.

Данные предметной области представляются *экземплярами сущностей* (студент Иванов, преподаватель Сидоров, дисциплина «Базы данных»). Экземпляры сущностей одного типа обладают одинаковыми наборами атри-

бутов, но должны отличаться значением хотя бы одного атрибута для того, чтобы быть узнаваемыми (например, студенты могут иметь одинаковые ФИО, но должны иметь разные номера зачётных книжек).

Среди атрибутов сущности можно выделить существенные и малозначительные. Признание какого-либо свойства существенным носит относительный характер. Например, атрибут *Должность* для сотрудника является существенным, а для читателя библиотеки — малозначительным.

Атрибуты можно условно классифицировать следующим образом:

1. *Идентифицирующие и описательные атрибуты*. Идентифицирующие атрибуты имеют уникальное значение для сущностей данного типа, описательные атрибуты включают в себе интересующие свойства сущности.
2. *Составные и простые атрибуты*. Простой атрибут состоит из одного компонента, его значение неделимо; составной атрибут является комбинацией нескольких компонентов, возможно, принадлежащих разным типам данных.
3. *Однозначные и многозначные атрибуты* (могут иметь соответственно одно или много значений для каждого экземпляра сущности).
4. *Основные и производные атрибуты*. Значение основного атрибута не зависит от других атрибутов. Значение производного атрибута вычисляется на основе значений других атрибутов.
5. *Обязательные и необязательные*. Значение обязательного атрибута всегда устанавливается при помещении данных в БД; значение необязательного атрибута может быть пропущено.

Спецификация атрибута состоит из его названия, типа данных, размера и описания ограничений целостности — множества значений, которые может принимать данный атрибут.

ПРИМЕЧАНИЕ

В данном учебном пособии наименования сущностей, атрибутов и связей выделяются курсивом и подчёркиванием. Кроме того:

1. Сущность записывается прописными буквами (*ОТДЕЛ*).
 2. Атрибут сущности начинается с прописной буквы (*Название*). Ключевой атрибут выделяется полужирным шрифтом (*Табельный номер*).
 3. Связь между сущностями определяется глаголом (*работает*).
-

Между сущностями ПрО могут существовать *связи*, имеющие различный содержательный смысл (семантику). Например, студент *учится* в группе, врач *лечит* пациента, клиент *имеет* вклад в банке. Связи могут быть *факультативными* или *обязательными*. Если вновь порождённая сущность одного из типов оказывается по необходимости связанной с сущностью другого типа, то между этими типами сущностей есть обязательная связь. В противном

случае связь является факультативной. Примеры обязательной и факультативной связей приведены на рис. 1.3. Здесь связь *замещает* является обязательной (изображается двойной линией), потому что каждый сотрудник *должен* работать на определённой должности, а связь *замещается* является факультативной, поскольку должность *может быть* вакантна.

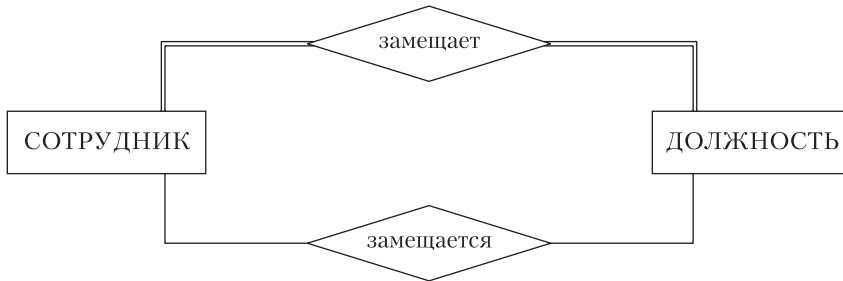


Рис. 1.3. Примеры обязательной и факультативной связей

Для удобства каждую связь между сущностями можно изображать одним ромбом (рис. 1.4). Выделяют также показатель *кардинальности связи*: «один к одному» (1:1), «один ко многим» (1:n) и «многие ко многим» (m:n) (рис. 1.4).

Связи, приведённые на рис. 1.4, с учётом семантики означают следующее:

- ✦ пациент—койка (1:1) — каждый пациент занимает одну койку, каждая койка в каждый момент времени может быть занята только одним пациентом;
- ✦ палата—пациент (1:n) — каждый пациент находится в одной палате, в каждой палате могут находиться несколько пациентов;
- ✦ пациент—врач (n:m) — каждый пациент может лечиться у нескольких врачей, каждый врач может лечить несколько пациентов.

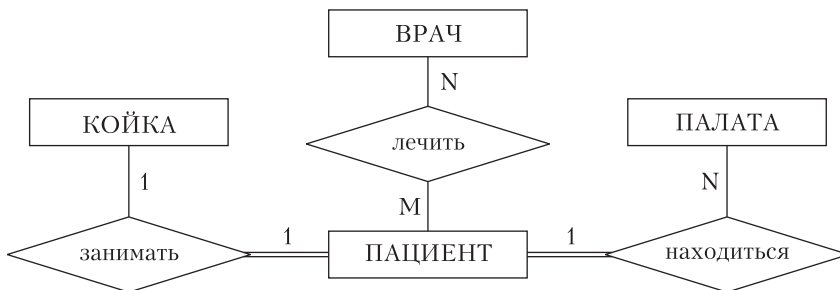


Рис. 1.4. Примеры различной кардинальности связей

Обратите внимание: необязательная связь имеет модификатор «может», а у обязательной связи его нет.

Степень связи — это количество типов сущностей, которые входят в связь. Различают *унарные* (рис. 1.5, а), *бинарные* (рис. 1.5, б) и *тернарные* (рис. 1.5, в) связи. (На практике связи с большей степенью редко используются). Унарная связь означает, что одни экземпляры сущности связаны с другими экземплярами этой же сущности (например, одни сотрудники руководят другими, а деталь может являться частью механизма).

Различают *тип связи* и *экземпляр связи*. Тип связи определяется её именем, обязательностью, степенью и кардинальностью, например бинарная связь *учится* между сущностями *ГРУППА* и *СТУДЕНТ*, обязательная для студента, кардинальностью 1:n. А экземпляр связи — это конкретная связь между студентом Сидоровым и группой Н-11, в которой он учится.

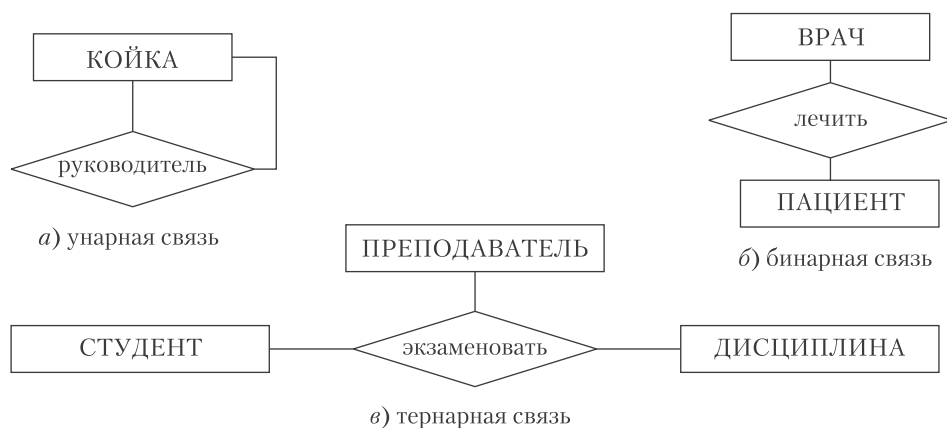


Рис. 1.5. Примеры связей различной степени

Совокупность типов сущностей и типов связей между сущностями характеризует структуру предметной области. Собственно данные представлены экземплярами сущностей и связей между ними. Данные экземпляров сущностей и связей хранятся в базе данных информационной системы, а описание типов сущностей и связей является метаданными.

Множества экземпляров сущностей, значения атрибутов сущностей и экземпляры связей между ними могут изменяться во времени. Поэтому каждому моменту времени можно сопоставить некоторое *состояние предметной области*. Состояния ПрО должны подчиняться совокупности правил, которые характеризуют семантику предметной области. В базе данных эти правила могут быть заданы с помощью так называемых *ограничений целостности*, которые накладываются на атрибуты сущностей, типы сущностей, типы связей и/или их экземпляры. Фактически ограничения целостности — это правила, которым должны удовлетворять значения данных в БД. Например, для библиотеки можно привести такие ограничения целостности: количество экземпляров книги не может быть отрицательным; номер паспорта читателя должен быть уникальным; каждая книга относится к определённому раз-

делу рубрикатора ББК — библиотечно-библиографической классификации и т. д.

Для того чтобы обеспечить соответствие базы данных текущему состоянию предметной области, база данных *динамически обновляется* (периодически или в режиме реального времени). Это обновление называется *актуализацией данных*. Актуализация может проводиться:

- ✦ вручную, если изменения в данные вносит пользователь (например, запись сведений о выдаче абоненту книги в библиотеке);
- ✦ автоматизировано, если изменения инициируются пользователем, но выполняются программно (например, обновление списка должников в библиотеке — читателей, которые просрочили дату возврата книг);
- ✦ автоматически, если данные поступают в электронном виде и обрабатываются программой без участия человека (это касается, например, автоматизированных систем управления производством).

Правильность обновлений может контролироваться программно, но правильнее контролировать их автоматически с помощью ограничений целостности БД.

База данных является информационной моделью внешнего мира, некоторой предметной области. Во внешнем мире сущности ПрО взаимосвязаны, поэтому в БД эти связи должны быть отражены. Если связи между данными в БД отсутствуют, то имеет смысл говорить о нескольких независимых БД и хранить их раздельно.

1.4. Назначение и основные компоненты системы баз данных

Система БД включает два основных компонента: собственно базу данных и систему управления базами данных — СУБД (рис. 1.6). Большинство СОД содержат также программы обработки данных (прикладное программное обеспечение, ППО), которые обращаются к данным через СУБД.

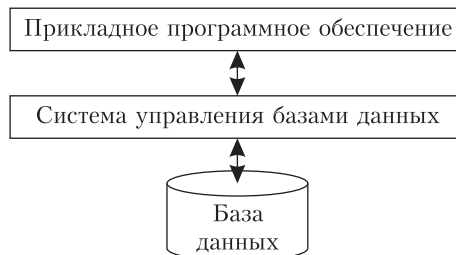


Рис. 1.6. Компоненты системы баз данных

В соответствии с рис. 1.6 СУБД обеспечивает выполнение двух групп функций:

- ✦ предоставление доступа к базе данных прикладному программному обеспечению (или квалифицированным пользователям);
- ✦ управление хранением и обработкой данных в БД.

Таким образом, обращение к базе данных возможно только через СУБД.

База данных предназначена для хранения данных информационной системы. Пользователи обращаются к БД обычно не напрямую через средства СУБД, а с помощью внешнего интерфейса — приложения, входящего в состав АИС. Если пользователей можно разделить на группы по характеру решаемых задач, то приложений может быть несколько (по количеству задач или групп пользователей). Например, для библиотеки можно выделить три группы пользователей: читатели, которым нужно осуществлять поиск книг по различным признакам; сотрудники, выдающие и принимающие у читателей книги (библиотекари); сотрудники отдела комплектации, осуществляющие приём новых книг и списание старых.

1.5. Уровни представления данных

Современная технология баз данных основана на концепции многоуровневой архитектуры СУБД. Эти идеи впервые были сформулированы в отчёте рабочей группы по базам данных Комитета по планированию стандартов Американского национального института стандартов (ANSI/X3/SPARC). Этот отчёт был опубликован в 1975 г. В нём предлагалась обобщенная трёхуровневая модель архитектуры СУБД, включающая концептуальный, внешний и внутренний уровни (рис. 1.7).

Концептуальный уровень архитектуры ANSI/SPARC служит для поддержки единого взгляда на базу данных, общего для всех её приложений и независимого от них и от среды хранения [5]. Концептуальный уровень представляет собой формализованную информационно-логическую модель ПрО. Описание этого представления называется *концептуальной схемой* или *схемой БД*.

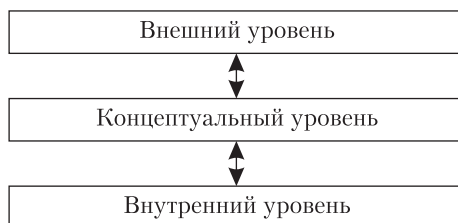


Рис. 1.7. Уровни представления данных

Схема базы данных — это описание базы данных в терминах конкретной модели данных.

Внутренний уровень архитектуры поддерживает представление данных в среде хранения и пути доступа к ним [5]. На этом архитектурном уровне БД представлена в полностью «материализованном» виде, тогда как на других уровнях идёт работа на уровне отдельных экземпляров или множества экземпляров данных. Описание БД на внутреннем уровне называется *внутренней схемой* или *схемой хранения*.

Внешний уровень архитектуры БД предназначен для групп пользователей. Описание представления данных для группы пользователей называется *внешней схемой*. Наличие внешнего уровня позволяет поддерживать разное представление одних и тех же данных для различных групп пользователей или задач [5].

Каждый из этих уровней может считаться управляемым, если он обладает внешним интерфейсом, обеспечивающим возможности определения данных. В этом случае становятся возможными формирование и системная поддержка независимого взгляда на БД для какой-либо группы персонала или пользователей, взаимодействующих с БД через интерфейс данного уровня.

В архитектурной модели ANSI/SPARC предполагается наличие в СУБД механизмов, обеспечивающих междууровневое отображение данных «внешний—концептуальный» и «концептуальный—внутренний». Функциональные возможности этих механизмов определяют степень независимости данных на всех уровнях. На переходе «внешний—концептуальный» обеспечивается *логическая независимость данных*, на переходе «концептуальный—внутренний» — *физическая независимость*. Под логической независимостью подразумевается возможность вносить изменения в концептуальный уровень, не меняя представление БД для пользователей, или изменять представление данных для пользователей без изменения концептуальной схемы. *Физическая независимость данных* подразумевает возможность вносить изменения в схему хранения, не меняя концептуальную схему БД.

Основной характеристикой баз данных является совместное использование данных многими пользователями АИС. Должно существовать какое-то общее понимание информации, представленной данными. Общее понимание должно относиться к чему-либо внешнему по отношению к пользователям, и оно должно быть зафиксировано. Для этого необходима некоторая предварительно определённая грамматика, которую принято называть моделью данных.

Задания для практических занятий

Тема: предметная область информационной системы

1. Выделить базовые и зависимые сущности для различных ПрО («Отдел кадров», «Магазин», «Институт», «Проектная организация»).
2. Определить набор атрибутов для различных сущностей («Студент», «Сотрудник», «Проект»).
3. Определить связи между сущностями в различных ПрО («Отдел кадров», «Магазин», «Институт», «Проектная организация»).