

WUTP



普通高等学校计算机科学与技术专业新编系列教材

An Introduction to Database System

# 数据库系统概论



杨开英 主编

```
#include <stdio.h>
#include <stdio.h>
void main()
void main()
{
    void swap(int * ptr1,int * ptr2);
    void swap(int * ptr1,int * ptr2);
    int x,y, * ptr1, * ptr2;
    int x,y, * ptr1, * ptr2;
    printf("input x,y:");scanf("%d,%d",&x,&y);
    printf("input x,y:");scanf("%d,%d",&x,&y);
    printf("%d\t%d\n",x,y);ptr1=&x;ptr2=&y;
    printf("%d\t%d\n",x,y);ptr1=&x;ptr2=&y;
    if(x<y)
    if(x<y)
    swap(ptr1,ptr2);
    swap(ptr1,ptr2);
    printf("%d\t%d\n",x,y);
    printf("%d\t%d\n",x,y);
}

void swap(int * ptr1,int * ptr2)
void swap(int * ptr1,int * ptr2)
```

武汉理工大学出版社

Wuhan University of Technology Press



普通高等学校计算机科学与技术专业新编系列教材

# An Introduction to Database System

# 数据库系统概论

主 编 杨开英

副主编 袁 捷 江定汉

武汉理工大学出版社

Wuhan University of Technology Press

## 内 容 提 要

本书全面、系统地介绍了数据库的基本理论、技术和应用等方面的知识。主要内容有数据库的基本概念、关系数据库的相关知识(概念、语言和设计理论)、数据库保护、数据库设计、数据库技术的新进展、数据库应用系统和基于 Web 数据库的应用系统的开发技术,最后给出了以 SQL Server 为实验平台的实验指导。

本书既重视数据库技术的体系完整性,又突出了数据库技术面向应用的特性,取材新颖、实用性强。因此,本书既可作为大学本科、专科层次的计算机专业的教材,也可供工程技术人员阅读参考。

### 图书在版编目(CIP)数据

数据库系统概论/杨开英主编. —武汉:武汉理工大学出版社,2003. 1

普通高等学校计算机科学与技术专业新编系列教材

ISBN 7-5629-1907-0

I. 数… II. 杨… III. 数据库系统概论-高等学校-教材 IV. TP311.13

中国版本图书馆 CIP 数据核字(2002)第 106865 号

出版发行:武汉理工大学出版社(武汉市武昌珞狮路 122 号 邮政编码:430070)

HTTP://www.whut.edu.cn/chuban/

E-mail:wutp @ public.wh.hb.cn

经 销 者:各地新华书店

印 刷 者:湖北省荆州市翔羚印刷实业公司

开 本:787×960 1/16

印 张:21.5

字 数:410 千字

版 次:2003 年 1 月第 1 版

印 次:2003 年 1 月第 1 次印刷

印 数:1~5000 册

定 价:28.00 元

本书如有缺页、倒页、脱页等印装质量问题,请向出版社发行部调换。本社购书热线电话:(027)87397097 87394412

**普 通 高 等 学 校**  
**计算机科学与技术专业新编系列教材**  
**编 审 委 员 会**

**顾问:**

卢锡城 周祖德 何炎祥 卢正鼎 曾建潮  
熊前兴

**主任委员:**

严新平 钟 珞 雷绍锋

**副主任委员:**

李陶深 鞠时光 段隆振 王忠勇 胡学钢  
李仁发 张常年 郑玉美 程学先 张翠芳  
孙成林

**委员:**(以姓氏笔画为序)

|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 王 浩 | 王景中 | 刘任任 | 江定汉 | 朱 勇 |
| 宋中山 | 汤 惟 | 李长河 | 李临生 | 李跃新 |
| 李腊元 | 李朝纯 | 肖俊武 | 邱桃荣 | 张江陵 |
| 张继福 | 张端金 | 张增芳 | 陈和平 | 陈祖爵 |
| 邵平凡 | 金 聪 | 杨开英 | 赵文静 | 赵跃华 |
| 周双娥 | 周经野 | 钟 诚 | 姚振坚 | 徐东平 |
| 黄求根 | 郭庆平 | 郭 骏 | 袁 捷 | 龚自康 |
| 崔尚森 | 蒋天发 | 詹永照 | 蔡启先 | 蔡瑞英 |
| 谭同德 | 熊盛武 | 薛胜军 |     |     |

**秘书长:**田道全

**总责任编辑:**段 超 徐秋林

## 出版说明

当今世界已经跨入了信息时代,计算机科学与技术正在迅猛发展。尤其是以计算机为核心的信息技术正在改变整个社会的生产方式、生活方式和学习方式,推动整个人类社会进入信息化社会。为了顺应时代潮流,适应计算机专业调整及深化教学改革的要求,充分考虑到不同层次高校的教学现状,满足广大高校的教学需求,武汉理工大学出版社经过广泛调研,与国内近 30 所高等院校的计算机专家进行探讨,决定组织编写“普通高等学校计算机科学与技术专业新编系列教材”。

我们在组织编写新编本套系列教材时,以培养现代化高级人才为重任,以提高学生综合素质、培养学生应用能力和创新能力为目的,以面向现代化、面向世界、面向未来为准绳,注重系列教材的特色和实用性,反映最新的教学与科研成果,体现本专业的时代特征。同时,面对教育改革的需要、人才的需要和社会的需要,在编写本教材时,借鉴、学习国外一流大学的先进教学体系,结合国内的实际需要,吸取具有先进性、实用性和权威性的国外教材的精华,以更好地促进国内教材改革顺利进行。从时代和国际竞争要求的高度来思考,为打造一套高起点、高水平、高质量的系列教材而努力。

本套教材具有以下特色:

与时俱进,内容科学先进——充分体现计算机学科知识更新快的特点,及时更新知识,确保教材处于学科前沿,以拓宽学生知识面,培养学生的创新能力。

紧跟教学改革步伐,体现教学改革的阶段性成果——符合全国高校计算机专业教学指导委员会、中国计算机学会教育委员会制订的“计算机学科教学计划 2000”的内容要求。

实现立体化出版,适应教育方式的变革——本套教材努力使用和推广现代化的教学手段,凡有条件的课程都准备组织编写、制作和出版配合教材使用的实验、习题、课件、电子教案及相应的程序设计素材库。

本套教材首批 25 种预定在 2003 年秋季全部出齐。我们的编审者、出版者决不敢稍有懈怠,一定高度重视,兢兢业业,按最高的质量标准工作。教材建设是我们共同的事业和追求,也是我们共同的责任和义务,我们诚恳地希望大家积极选用本套教材,并在使用过程中给我们多提意见和建议,以便我们不断修订、完善全套教材。

武汉理工大学出版社

2002 年 10 月

# 前 言

在计算机的三大主要应用领域(科学计算、数据处理和过程控制)中,数据处理迅速上升为计算机应用的主要方向,而数据库技术又是数据处理的最新技术。在当今信息化社会中,数据库是组织、管理和利用信息的最有效的方法。特别是以大量信息交流为目的的互联网技术的应用和普及,更使得数据库技术成为大众化技术。因此,数据库系统课程成为高等学校计算机教学中的主干课,它不仅是计算机专业的必修课,也是许多非计算机专业的必修课或选修课。

数据库技术是计算机科学中面向应用的一个重要分支。数据库技术由应用而催生,随着应用的普及和深入而蓬勃发展。因此,本书的宗旨是在系统介绍数据库技术的基本原理和方法的同时,突出其面向应用的特征。首先,突出了主流数据库——关系数据库的理论、操作和应用;其次,加强了对数据库设计这一实践性很强的内容的介绍;第三,紧跟当前的应用热点,介绍了数据库的新进展和最新的应用课题;最后,配合书中的习题,安排了上机实验指导。

关于教学安排的建议是:少学时课程可重点讲授第1~6章,对于各章理论性较强的内容可作适当精简,对于第7~9章可以有选择性地介绍。除完成SQL语言的上机之外,最好能安排课程设计,将数据库设计和应用系统开发有机地结合起来。

本书是集体劳动的成果,由杨开英同志担任主编,袁捷、江定汉同志担任副主编。具体参加编写工作的人员及分工如下:第1章由杨开英编写,第2章、第5章由江定汉编写,第3章、第8章由谭瑛编写,第4章、第9章由袁捷编写,第6章、第7章由刘洪星编写;由袁捷审阅第1~3章,杨开英审阅第4~6章,江定汉审阅第7~9章,最后由杨开英总纂定稿,并编写了附录。

限于编者的水平和时间仓促,书中欠妥之处在所难免,敬请读者不吝赐教。

编 者

2002年11月

# 目 录

|                          |      |
|--------------------------|------|
| <b>1 数据库系统概述</b> .....   | (1)  |
| 1.1 数据库系统基础知识 .....      | (1)  |
| 1.1.1 信息、数据和数据处理 .....   | (2)  |
| 1.1.2 数据处理的三个领域 .....    | (2)  |
| 1.1.3 数据库管理的发展过程 .....   | (5)  |
| 1.1.4 数据库的体系结构 .....     | (9)  |
| 1.1.5 数据库技术的特点 .....     | (11) |
| 1.2 数据模型 .....           | (12) |
| 1.2.1 数据模型的概念 .....      | (12) |
| 1.2.2 实体-联系(E-R)模型 ..... | (14) |
| 1.2.3 层次模型 .....         | (17) |
| 1.2.4 网络模型 .....         | (20) |
| 1.2.5 关系模型 .....         | (24) |
| 1.2.6 数据模型间的转化 .....     | (26) |
| 1.3 数据库管理系统 .....        | (27) |
| 1.3.1 DBMS 的功能 .....     | (27) |
| 1.3.2 DBMS 的组成 .....     | (28) |
| 1.3.3 数据字典 .....         | (28) |
| 1.3.4 流行的 DBMS 简介 .....  | (30) |
| 1.4 数据库系统组成 .....        | (32) |
| 1.4.1 计算机支持环境 .....      | (32) |
| 1.4.2 数据库管理人员 .....      | (33) |
| 1.4.3 数据库系统的数据访问过程 ..... | (33) |
| 1.5 数据库技术的研究内容和进展 .....  | (34) |
| 复习思考题与习题 .....           | (36) |
| <b>2 关系数据库</b> .....     | (38) |
| 2.1 关系数据模型的基本概念 .....    | (38) |

---

|       |                |      |
|-------|----------------|------|
| 2.1.1 | 关系模型及其描述       | (38) |
| 2.1.2 | 关系模型的特点        | (40) |
| 2.2   | 关系及关系数据库       | (41) |
| 2.2.1 | 关系的定义          | (41) |
| 2.2.2 | 关系模式           | (43) |
| 2.2.3 | 关系数据库          | (44) |
| 2.2.4 | 关系系统           | (44) |
| 2.3   | 关系的完整性约束       | (46) |
| 2.3.1 | 实体完整性约束        | (46) |
| 2.3.2 | 参照完整性约束        | (47) |
| 2.3.3 | 用户定义完整性约束      | (47) |
| 2.4   | 关系代数           | (48) |
| 2.4.1 | 传统集合运算         | (50) |
| 2.4.2 | 专门关系运算         | (51) |
| 2.5   | 关系演算           | (57) |
| 2.5.1 | 元组关系演算         | (57) |
| 2.5.2 | 域关系演算          | (63) |
| 2.6   | 关系运算的等价变换原则    | (68) |
| 2.6.1 | 关系运算的安全性       | (68) |
| 2.6.2 | 关系运算的等价性       | (71) |
| 2.7   | 查询优化           | (77) |
| 2.7.1 | 查询实例           | (77) |
| 2.7.2 | 查询优化的一般准则      | (78) |
| 2.7.3 | 关系代数的等价变换规则    | (79) |
| 2.7.4 | 关系代数表达式的优化算法   | (80) |
| 2.7.5 | 优化的一般步骤        | (81) |
|       | 复习思考题与习题       | (84) |
| 3     | 关系数据库的标准语言 SQL | (87) |
| 3.1   | SQL 概述         | (88) |
| 3.1.1 | SQL 的主要功能      | (88) |
| 3.1.2 | SQL 的特点        | (89) |
| 3.2   | 数据定义           | (90) |
| 3.2.1 | 创建与删除数据库       | (91) |
| 3.2.2 | 创建、删除、修改基本表    | (91) |
| 3.2.3 | 创建与删除索引        | (94) |

---

|                              |       |
|------------------------------|-------|
| 3.3 查询                       | (96)  |
| 3.3.1 简单查询                   | (98)  |
| 3.3.2 连接查询                   | (101) |
| 3.3.3 嵌套查询                   | (104) |
| 3.3.4 组合查询                   | (107) |
| 3.4 数据更新                     | (109) |
| 3.4.1 插入数据                   | (109) |
| 3.4.2 修改数据                   | (111) |
| 3.4.3 删除数据                   | (112) |
| 3.5 视图                       | (113) |
| 3.5.1 视图的创建                  | (113) |
| 3.5.2 视图删除                   | (114) |
| 3.5.3 视图的查询                  | (114) |
| 3.5.4 视图的更新                  | (114) |
| 3.6 数据控制                     | (115) |
| 3.6.1 权限的类型                  | (115) |
| 3.6.2 授权                     | (115) |
| 3.6.3 收回授权                   | (116) |
| 3.7 嵌入式 SQL                  | (116) |
| 3.7.1 基本技术                   | (117) |
| 3.7.2 举例                     | (120) |
| 复习思考题与习题                     | (125) |
| <b>4 关系数据库设计理论</b>           | (128) |
| 4.1 关系模式的设计问题                | (128) |
| 4.1.1 关系模式的质量                | (128) |
| 4.1.2 泛关系假设                  | (129) |
| 4.1.3 本章的符号约定                | (130) |
| 4.2 函数依赖                     | (130) |
| 4.2.1 函数依赖的定义                | (130) |
| 4.2.2 两种闭包                   | (131) |
| 4.2.3 函数依赖推理规则               | (131) |
| * 4.2.4 Armstrong 公理的有效性和完备性 | (132) |
| 4.2.5 属性集闭包算法                | (133) |
| 4.2.6 关系模式的键                 | (135) |
| 4.2.7 最小函数依赖集                | (136) |

---

|                               |              |
|-------------------------------|--------------|
| 4.3 模式分解与等价性 .....            | (137)        |
| 4.3.1 模式的表示能力 .....           | (137)        |
| 4.3.2 无损的分解 .....             | (138)        |
| 4.3.3 分解无损性检验 .....           | (140)        |
| 4.3.4 保持函数依赖的分解 .....         | (142)        |
| 4.3.5 模式等价性标准 .....           | (143)        |
| 4.4 函数依赖与规范化 .....            | (143)        |
| 4.4.1 第一范式~第三范式 .....         | (143)        |
| 4.4.2 Boyce-Codd 范式 .....     | (144)        |
| 4.4.3 分解成 3NF 模式集的算法 .....    | (145)        |
| * 4.4.4 分解成 BCNF 模式集的算法 ..... | (145)        |
| 4.5 其他数据依赖与规范化 .....          | (147)        |
| 4.5.1 多值依赖和第四范式 .....         | (147)        |
| * 4.5.2 连接依赖和第五范式 .....       | (148)        |
| 复习思考题与习题 .....                | (151)        |
| <b>5 数据库保护 .....</b>          | <b>(154)</b> |
| 5.1 安全性 .....                 | (154)        |
| 5.1.1 安全控制的一般方法 .....         | (155)        |
| 5.1.2 存取控制 .....              | (159)        |
| 5.2 数据完整性 .....               | (160)        |
| 5.2.1 完整性约束 .....             | (160)        |
| 5.2.2 完整性控制 .....             | (162)        |
| 5.2.3 SQL Server 的完整性 .....   | (165)        |
| 5.3 事务处理 .....                | (167)        |
| 5.3.1 事务的基本概念 .....           | (167)        |
| 5.3.2 事务的两段提交 .....           | (168)        |
| 5.4 并发控制 .....                | (170)        |
| 5.4.1 并发控制概述 .....            | (170)        |
| 5.4.2 封锁 .....                | (172)        |
| 5.4.3 并发控制的正确性 .....          | (173)        |
| 5.4.4 两段锁协议 .....             | (174)        |
| 5.5 数据库故障的恢复 .....            | (175)        |
| 5.5.1 故障种类 .....              | (175)        |
| 5.5.2 恢复的实现技术 .....           | (176)        |
| 5.5.3 恢复策略 .....              | (181)        |

|                       |       |
|-----------------------|-------|
| 复习思考题与习题              | (182) |
| <b>6 数据库设计</b>        | (183) |
| 6.1 数据库设计概述           | (183) |
| 6.1.1 数据库设计的任务和特点     | (183) |
| 6.1.2 数据库设计的基本步骤      | (185) |
| 6.1.3 数据库设计方法         | (187) |
| 6.1.4 数据库设计工具         | (188) |
| 6.2 数据库概念设计           | (189) |
| 6.2.1 概念设计方法概述        | (189) |
| 6.2.2 数据抽象            | (191) |
| 6.2.3 扩展的 E-R 模型      | (193) |
| 6.2.4 设计局部概念模式        | (197) |
| 6.2.5 概念模式汇总          | (203) |
| 6.3 数据库逻辑结构设计         | (205) |
| 6.3.1 概念模式到关系数据库模式的转换 | (205) |
| 6.3.2 关系模式的优化         | (209) |
| 6.3.3 逻辑结构的图示和自动生成    | (211) |
| 6.4 数据库物理设计           | (213) |
| 6.4.1 物理设计的内容和特点      | (213) |
| 6.4.2 索引设计            | (214) |
| 6.4.3 聚簇设计            | (215) |
| 6.5 数据库的实施与维护         | (217) |
| 6.5.1 数据库实施           | (217) |
| 6.5.2 数据库运行维护         | (219) |
| 复习思考题与习题              | (220) |
| <b>7 数据库技术的新进展</b>    | (222) |
| 7.1 面向对象数据库技术         | (223) |
| 7.1.1 面向对象数据库系统       | (223) |
| 7.1.2 对象-关系数据库系统      | (226) |
| 7.2 分布式数据库系统          | (232) |
| 7.2.1 分布式数据库系统概述      | (232) |
| 7.2.2 分布式数据库系统的管理     | (234) |
| 7.3 数据仓库技术            | (238) |
| 7.3.1 数据仓库的概念         | (238) |
| 7.3.2 数据仓库系统的组成模型     | (240) |

|                                     |       |
|-------------------------------------|-------|
| 复习思考题与习题                            | (243) |
| <b>8 数据库应用系统</b>                    | (244) |
| 8.1 数据库应用系统的结构                      | (244) |
| 8.1.1 多用户 DBS                       | (245) |
| 8.1.2 客户机/服务器式 DBS                  | (246) |
| 8.1.3 Web DBS                       | (247) |
| 8.1.4 并行式 DBS 和分布式 DBS              | (249) |
| 8.2 编程接口                            | (249) |
| 8.2.1 应用编程中的几个问题                    | (252) |
| 8.2.2 存储过程及触发器                      | (254) |
| 8.3 典型的数据库应用系统开发环境                  | (255) |
| 8.3.1 Power Builder                 | (255) |
| 8.3.2 Delphi                        | (259) |
| 8.3.3 Visual Basic                  | (260) |
| 8.3.4 Developer/2000                | (261) |
| 复习思考题与习题                            | (263) |
| <b>9 Web 数据库技术基础</b>                | (264) |
| 9.1 Internet 和 Web                  | (264) |
| 9.1.1 有关的 Internet 概念               | (265) |
| 9.1.2 有关的 Web 概念                    | (267) |
| 9.2 Web 数据库应用体系结构                   | (270) |
| 9.2.1 分布计算和客户机/服务器结构                | (271) |
| 9.2.2 Web 数据库解决方案                   | (273) |
| 9.3 Web 数据库应用编程接口                   | (279) |
| 9.3.1 ODBC                          | (279) |
| 9.3.2 JDBC                          | (284) |
| 9.3.3 ADO                           | (290) |
| 9.4 可扩展标记语言 XML                     | (297) |
| 9.4.1 XML 及其应用类型                    | (297) |
| 9.4.2 XML 数据访问机制                    | (301) |
| 9.4.3 XML 与 DBS                     | (303) |
| 复习思考题与习题                            | (308) |
| <b>附录 A 基于 SQL SEVER 2000 的实验指导</b> | (310) |
| A.1 SQL Server 2000 使用基础            | (310) |
| A.1.1 SQL Server 2000 的版本和安装设置简介    | (310) |

---

|                                                   |       |
|---------------------------------------------------|-------|
| A. 1. 2 用户管理 .....                                | (313) |
| A. 2 系统基本工具 .....                                 | (314) |
| A. 2. 1 SQL Server Enterprise Manager 的使用简介 ..... | (314) |
| A. 2. 2 SQL Query Analyzer 简介 .....               | (319) |
| A. 2. 3 其他主要工具 .....                              | (320) |
| A. 3 实验安排 .....                                   | (321) |
| 实验 1 数据库的定义实验 .....                               | (321) |
| 实验 2 数据库的建立和维护实验 .....                            | (321) |
| 实验 3 数据库的简单查询和连接查询实验 .....                        | (322) |
| 实验 4 数据库的嵌套查询和组合查询 .....                          | (322) |
| 实验 5 数据库的更新、视图的定义及使用 .....                        | (323) |
| 实验 6 数据的导入/导出、数据库备份和恢复 .....                      | (323) |
| 实验 7 数据完整性和数据安全性 .....                            | (324) |
| 实验 8 综合实验 .....                                   | (324) |
| 附录 B 术语 .....                                     | (326) |
| 参考文献 .....                                        | (330) |



# **1 数据库系统概述**

## **本章提要**

本章主要介绍数据库系统最基本、最重要的概念,例如什么是数据、数据管理、数据库、数据模型、数据库管理系统和数据库系统。数据库是最重要的概念,在理解数据库体系结构的基础上掌握什么是数据库及数据库技术的特点;数据模型是数据库的组织基础,要求掌握数据模型的结构特点及不同模型的区别,重点是E-R模型和关系模型;数据库管理系统是数据库系统的核心,要求重点掌握数据库管理系统的组成与功能,并借助数据字典理解数据库管理系统的复杂功能;数据库系统是数据库技术的实现系统,要求重点掌握数据库系统中各个部分的作用和相互关系;最后,要求了解数据库技术的研究内容和进展。

数据库技术是现代社会的**重要基础技术之一**,是计算机科学领域中发展最为迅速的重要分支。随着计算机应用的普及和深入,数据库技术越来越重要,了解、掌握数据库系统的基本概念和基本技术成为应用数据库技术的前提。本章首先介绍数据库系统的产生、发展和涉及的基本概念,包括数据模型、数据库管理系统、数据库系统的组成,最后简述数据库技术的研究内容和新发展。

## **1.1 数据库系统基础知识**

数据库系统是由于数据处理而产生和发展的,因此信息、数据和数据处理是最基本的概念。

### 1.1.1 信息、数据和数据处理

一般地,人们经常将术语“信息”与“数据”互换使用,但是在数据库系统中,区别它们是必要的。

信息是现实世界中事物的状态、运动方式和相互关系的表现形式,是自然界、人类社会和人类思维活动中普遍存在的一切物质和事物的属性。这里的事物不仅指那些看得见、摸得着的具体物体,如人、树木、房子,也包含那些不可触及的抽象概念,如产量、质量等。因此,信息可以被看成是现实世界在人脑中的抽象反映,是通过人的感官(眼、耳、鼻、舌、身)感知出来并经过人脑的加工而形成的反映现实世界中事物的概念。

在我们的学习、生活和工作中,经常接触到各式各样的信息,如文字、数字、声音、图形、图像等,可以说现实世界到处充满了信息。信息可以用人工或自动化装置进行记录、解释和处理。在记录、传播和利用信息时,必须将现实世界的信息转换为能够识别的符号,符号就是通常所说的数据。

数据是一种物理符号序列。数据有数据类型和数据值之分,不同类型的数据记录事物的性质是不一样的。例如,数值型数据 1,2,3……可以用来描述事物的多少,而文字型数据红、黄、蓝、绿……可以用来描述事物的颜色。如果说数据是反映客观的记录符号,信息则是潜在于数据中的意义。

数据和信息是两个互相联系、互相依赖但又互相区别的概念。所有的信息都是数据,数据只有经过提炼和抽象之后,具有使用价值时才能成为信息。数据是信息存在的形式,是信息的载体,随着介质不同而改变。信息以数据为载体而表现,同一信息可以有不同的数据表示方式,但其意义为不变。为了处理方便,经常用数字表示信息。用数值表示信息有时是自然的,如国家的人口数量、钢铁的产量等;而有时是人为的,如令 1 表示红,2 表示黄……这就是人为的。经过这样的变换后,世界上几乎所有的信息都可以用数字表示,这就是计算机能够处理各种信息的原因。

### 1.1.2 数据处理的三个领域

数据处理是指对信息进行收集、整理、存储、加工及传播等一系列活动的总和,其基本目的是从大量的、杂乱无章的甚至是难于理解的数据中,提炼、抽取人们所需要的有价值、有意义的数据(信息),借以作为决策的依据。数据的组织、存储、检查和维护等工作是数据处理的基本环节,这些工作一般统称为数据管理。

数据表示信息,信息反应事物的客观状态,事物、信息、数据三者之间互为联系。从事物的状态到表示该状态的数据,经历了三个领域,这就是现实世界、信

息世界和计算机世界。

三个领域的联系如图 1.1 所示。

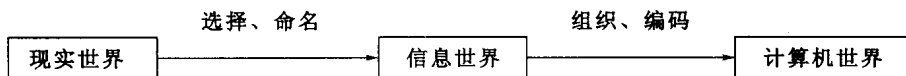


图 1.1 三个领域的关系

以下介绍三个领域中几个基本概念。

#### (1) 现实世界

现实世界是存在于人们头脑之外的客观世界,由客观事物及其联系组成。

① 事物 现实世界的事物可以是实际存在的人和物,例如张三、李四、人民大会堂、黄山迎客松等;也可以是抽象的事物和事物间的联系,例如操作系统、数据库系统概论等课程,或者借阅图书、购物等。

② 事物类 每一事物类由隶属于该类的具体事物组成,是同类事物的集合。例如,张三、李四等组成“人”这一事物类,而操作系统、数据库系统概论等组成“课程”事物类。

③ 特性 每一个事物具有它自己的特性,例如,对张三来说,有姓名/张三、性别/男、身高/1.72m、体重/60kg 等。特性有名和值之分。具有相同特性的事物属于一个事物类。

#### (2) 信息世界

信息世界是对现实世界的抽象和描述。与现实世界对应的概念是:

① 实体(Entity) 现实世界中客观存在并且可以相互区分的事物在信息世界称为实体。

② 实体集(Entity Set) 现实世界中的事物类,在信息世界中就叫实体集,是同类实体的集合。

③ 属性(Attribute) 现实世界中事物的特性就是实体的属性。属性也有名和值的区分,属性名用来划分实体所属的实体集,而属性值则是某个实体在该属性下的具体表现。属性值的集合称为属性的域(Domain)。

④ 实体标识符(Identification) 用于和同类实体相互区分的属性集合(不含多余的属性)称为实体标识符。

#### (3) 数据世界

在数据世界(计算机世界)中,与信息世界几个概念相对应的分别是:

① 记录实例值(Record Occurrence) 简称记录,表示实体。

② 文件(File) 是记录的集合。

③ 字段(Field)或数据项(Data Item) 它对应信息世界的属性。

④ 关键字(Key) 它对应信息世界的实体标识符。

三个世界的类比关系如表 1.1 所示。

表 1.1 三个世界的类比关系

| 现实世界 | 信息世界  | 计算机世界 |
|------|-------|-------|
| 事物   | 实体    | 记录    |
| 事物类  | 实体集   | 文件    |
| 特性   | 属性    | 字段    |
| 区分特征 | 实体标识符 | 关键字   |

如学生实体集合在计算机世界可用文件表示,见图 1.2。

|    |        |     |     |     |      |
|----|--------|-----|-----|-----|------|
| 文件 | 数据项名   |     |     |     |      |
|    | 学号     | 姓名  | 性别  | 年龄  | 籍贯   |
|    | 010301 | 李明  | 男   | 18  | 湖北武汉 |
|    | 010302 | 王平  | 男   | 19  | 河南郑州 |
|    | 010303 | 林丽萍 | 女   | 18  | 湖南衡阳 |
|    | ...    | ... | ... | ... | ...  |
|    | 011010 | 赵兵  | 男   | 18  | 广东广州 |

图 1.2 数据文件的平面表示

所以文件中的一条记录对应实体集中的一个实体,而一个字段对应于一个属性。

#### (4) 记录的型与值

正如在信息世界的实体有集合与个体之别,与之相对应的数据世界就有了记录的型与值的区别。文件中的一条记录实例是对应于一个特定实体的数据项值的有序集合,称为记录实例值,简称记录。而组成记录的各个数据项的名称和排列顺序称为记录的型,记录的型确定了记录的格式,同一个文件下的各条记录都对应同一个格式,所以记录的型描述实体集的整体特征,一个记录型对应于一个实体集。而记录值也就是记录型的实例值。记录型和记录值的表示如图 1.3 所示。

|           |     |    |    |      |
|-----------|-----|----|----|------|
| <u>学号</u> | 姓名  | 性别 | 年龄 | 籍贯   |
| (a)       |     |    |    |      |
| 010303    | 林丽萍 | 女  | 18 | 湖南衡阳 |
| (b)       |     |    |    |      |

图 1.3 记录型和记录值的表示

(a) 记录型; (b) 记录值

记录型中加下划线的数据项是记录的关键字。