



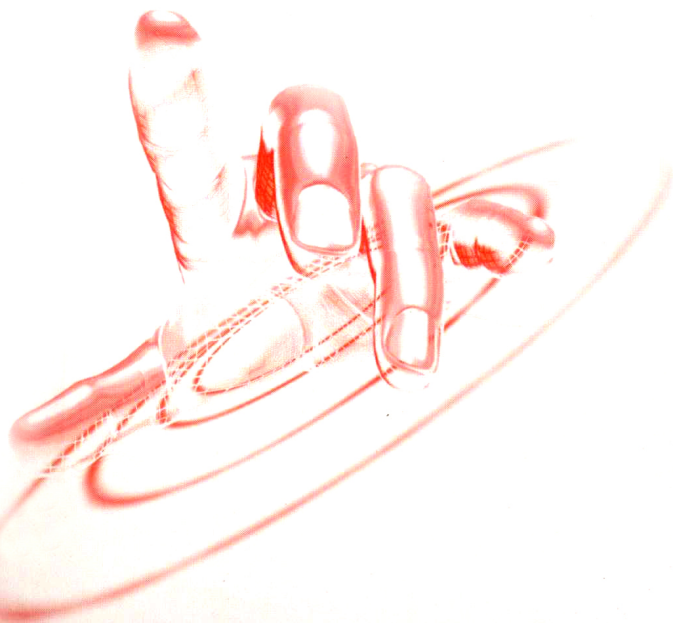
普通高等教育精品规划教材

高等学校信息管理专业系列教材

第三版

信息资源数据库

周 宁 吴佳鑫 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

INFORMATION MANAGEMENT SCIENCE



普通高等教育精品规划教材

高等学校信息管理专业系列教材

第三版

信息资源数据库

周 宁 吴佳鑫 编著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

信息资源数据库/周宁,吴佳鑫编著. —武汉:武汉大学出版社,
2010.9

普通高等教育精品规划教材

高等学校信息管理专业系列教材

ISBN 978-7-307-07983-0

I. 信… II. ①周… ②吴… III. ①信息系统—高等学校—教材
②数据库系统—高等学校—教材 IV. ①G202 ②TP311.13

中国版本图书馆CIP数据核字(2010)第131151号

责任编辑:林莉 责任校对:王建 版式设计:詹锦玲

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:cls22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北民政印刷厂

开本:720×1000 1/16 印张:26 字数:463千字 插页:1

版次:2000年10月第1版 2006年9月第2版

2010年9月第3版第1次印刷

ISBN 978-7-307-07983-0/G·1705 定价:30.00元

版权所有,不得翻印;凡购我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。

第 1 版前言

在当今知识经济时代,信息是物质、能源之后的“第三级资源”。人们每天都在生产信息、消费信息。信息资源的开发与利用水平已成为衡量一个国家发展状况和综合国力的重要尺度。信息是国家的战略资源和人类的宝贵财富。在因特网快速发展的今天,建立信息系统、发展信息产业已成为各国追求的目标。而信息系统的基础是数据库。

数据库技术诞生于 20 世纪 60 年代末,如今已过“而立”之年。它已从幼年走向了成熟期。数据库系统已经历了第一代(层次式数据库系统和网状数据库系统)、第二代(关系数据库系统),现正在开发和应用第三代数据库系统。数据库理论和方法不断出现新突破。面向对象的数据库系统、多媒体数据库系统、分布式数据库系统、数据仓库、Web 数据库、知识库与智能数据库系统、工程数据库、并行数据库、主动数据库、模糊数据库等新亮点不断涌现,使数据库大家族更加兴旺。

《信息资源数据库》是为信息管理专业编写的新教材。全书共分三个部分:

第一部分为数据库系统原理,分五章(第一章至第五章)讲述。第一章为数据库系统引论,从信息资源管理的需要引入了信息模型,数据模型和数据库技术,对数据库系统、数据库管理系统作了全面而简要的介绍。第二章讨论了关系数据库系统,从数学理论的高度定义了关系与关系模型、关系代数与关系演算;并对国内外流行的关系数据库管理系统进行了介绍。第三章论述了关系数据库的标准语言 SQL,从理论与实践上进行了深入讨论。第四章介绍了关系模型设计理论,对于函数相关性和关系的规范化形式的不同级别进行了讨论。第五章论述数据库设计专题,从数据库设计条件、设计过程和完整性措施进行了讨论。数据字典是设计和管理数据库的工具,也在本章进行专门介

绍。根据有关的国际标准,对设计的信息资源数据库进行了分类,它是转入第二部分内容的自然过渡与衔接。

第二部分为信息资源数据库,分四章(第六章至第九章)论述。第六章为文献资源数据库,对文献型的源数据库(全文数据库)和咨询数据库(二次文献库)分别进行讨论,并对文献信息的数字化、标准化、网络化管理专题进行了具体探讨。本章还讨论了文献数据库的设计、建立与维护的理论和实践问题。第七章讨论了数值数据库与事实数据库。结合我国的应用实际,对科技成果数据库、物价数据库、科学数据库的利用进行了具体讨论。第八章为多媒体数据库,从理论方法、设计、建立与应用进行了讨论。第九章为 Web 数据库,它是因特网快速发展的产物。Web 数据库讨论了其特点、基本结构和实现技术,并对虚拟数据库进行了探讨。第二部分的四章论述了因特网环境下信息资源数据库的研究内容和最新成果,突出了专业特色。

第三部分为数据压缩及数据库技术的新进展。数据压缩是与数据库技术紧密相关的研究课题,在“信息爆炸”的今天研究它尤为重要。第八章具体讨论了数据压缩的理论与方法,对于传统数据库和多媒体数据库的数据压缩技术进行了探讨。第十一章讨论了数据库技术的新进展,就目前研究的主要领域分专题进行了具体论述。

由于 IMS 系统和 DBTG 系统是第一代数据库系统的代表,它的内容不可缺少。但考虑到这两个系统已先后逐步退出市场,加之课时有限,因而我们参考国外一些大学现行教材的处理方法,将其在附录中安排,作为教学参考或阅读材料。

第一部分为数据库原理教学大纲规定的内容,第二部分为专业建设需要的内容,这两部分为必选教学内容。第三部分,多数是“超纲”的新课题,为任选内容。根据教学需要,教师可自由取舍。

本书的编写出版得到了武汉大学各级领导的关心和支持,武汉大学出版社的同志为此付出了辛勤的劳动。多年来,得到了中国科学技术信息研究所、中国科学院科学数据库等单位大力支持和许多同行专家的关心、支持和帮助。撰稿过程中,参阅了大量文献和技术成果,在此一并致以衷心的感谢。由于作者水平有限,书中缺点和不足之处难免,敬请读者提出宝贵意见。

作者

2000年7月于珞珈山

第2版前言

科技进步日新月异,数据库领域内的新理论、新技术层出不穷。信息资源数据库第1版的出版转眼已过去5年;所以,第2版比第1版的内容有较大的更新。

数据模型是数据库的核心。第2版增写了第2章数据模型。在系统描述第一代数据库系统的数据模型(用层次模型图和网状模型图描述)、第二代数据库系统的数据模型(关系模型即用二维表的集合进行描述)和第三代数据库系统的数据模型(用对象类层次结构描述)的基础上,根据实际需要引入了实体联系(E-R)模型、鸭掌模型、对象关系(即扩展的RDM)模型等。数据库的模型方法是第2版的重要内容。

数据库管理系统(DBMS)是数据库系统(DBS)的大管家。一些流行的关系数据库管理系统(RDBMS)也独立成章(第12章)。在分别讨论了Access、VFP、DB2、Informix、INGRES、Oracle、Sybase、SQL Server之后,针对国内数据库系统的应用实际,对Oracle和SQL Server的应用作了说明,结合实例对系统安装与利用进行具体讨论。

在关系数据库系统中,补充了关系变量、超码、关联度与势等内容。在文献数据库中,对近几年发展较快的MARC21(ISO 2709/XML)进行了重点论述,结合应用实际作了实例说明。第2版重写了大部分章节。例如,随着科研与应用的深入,价格数据库、中国科学院系统的数据库工程发展很快,我们重写了数值数据库、科学数据库等。

总之,第2版对第1版进行了增写、重写、改写与调整工作,使《信息资源数据库》(第2版)以全新的面貌奉献给读者。

第2版的内容可分为三大部分。

第一部分为数据库系统原理。具体论述了数据库的基础知识、基本理论

(第1章数据库与数据库系统、第2章数据模型、第3章关系数据库系统、第5章关系模型设计理论)和基本技能(第4章SQL、第6章数据库设计)。

第二部分(第7~9章)为信息资源数据库。具体讨论数据库在信息资源管理中的应用。第7章论述常规数据库(文献数据库、事实数据库、数值数据库与综合数据库)。第8章讨论多媒体数据库的关键技术与广泛应用。第9章讨论Web数据库。信息资源是国家的战略资源与宝贵财富。开发与利用信息资源数据库是我们面临的重要任务。

第三部分(第10~12章)论述了数据库的新技术与新进展。第10章是数据压缩。它是因特网环境下数据库应用的基础技术。无论是常规数据库还是多媒体数据库,一是要快速传送,二是要紧缩存储。这就必须研究数据压缩。若没有数据压缩的丰硕成果,多媒体数据库可能至今还不能达到实用化水平。DBS的大管家是DBMS。第12章专门讨论了一些流行的RDBMS。它们是第11章数据库新成果的支撑。无论是分布式数据库、并行数据库、主动数据库和工程数据库,还是数据仓库与数据挖掘、知识库与智能数据库系统、模糊数据库都需要大型RDBMS高版本的支持。根据目前国内数据库市场的分析,我们还对SQL Server和Oracle系统的具体应用进行了探讨。

在第2版书稿的撰写过程中,刘玮、张芳芳、吴佳鑫、张少龙分别参加了第5、7、9、12章初稿的撰写与插图绘制工作。

第2版的出版得到了国内同行许多专家的帮助,武汉大学各级领导给予了关心和支持,武汉大学出版社的同志付出了辛勤劳动,在此一并表示衷心感谢。由于作者水平有限,错漏之处难免,欢迎读者批评指正。

作者

2006年6月

第3版前言

光阴似箭,数据库技术从诞生、快速发展已到了“不惑之年”。由于信息资源管理的计算机化、数字化、网络化逐步普及,使其现代化、自动化水平提高到一个新阶段。数据库技术已成为信息资源管理的最佳技术。

本书第2版出版后,在数据库领域又出现了许多新理论、新技术、新成果。为此,第3版对第2版中的大部分章节作了改写、增写、重写和调整。特别是补充了一批实践方面的新知识和实例,使之更适合培养全面发展的人才。

第3版保留了第2版的体系结构,全书分三大部分共12章。

第一部分(第1~6章)为数据库系统应用基础。它系统论述了数据库系统的基本理论、基础知识和基本方法,以及数据库设计理论与方法等。

第二部分(第7~9章)具体讨论了信息资源数据库的分类方法,分别讲述了文献数据库、事实数据库、数值数据库和多媒体数据库的建立、标准化、维护和应用的理论与实践问题。对当前流行的Web数据库的系统结构、Web服务器与数据库服务器的连接技术进行了系统论述,力求理论与实际应用紧密结合。

第三部分(第10~12章)对数据库相关技术(数据压缩)和数据库系统新进展进行专题讨论,为了扩展知识面和提高学员的应用水平,对国内流行的8种RDBMS进行了专题介绍。

综上所述,第3版在第2版注重数据库理论与应用技术并重的基础上,加大了实际应用技术的比重。它有利于学员拓宽知识面,具有更强的实际工作能力。

参加本书第3版撰稿的有吴佳鑫、刘玮、张少龙、张芳芳,全书由周宁、吴佳鑫统稿。

本书的出版得到国内许多专家的支持和帮助,武汉大学出版社的同志为此付出了辛勤的劳动,在此一并致以谢忱。由于作者水平所限,书中错漏之处难免,恳请读者不吝指正。

作者

2010年4月11日

目 录

第 1 版前言	1
第 2 版前言	1
第 3 版前言	1
1 数据库与数据库系统	1
1.1 数据与信息资源	1
1.2 数据库的诞生与发展	3
1.3 为什么要研究数据库?	6
1.4 数据库系统	9
1.5 数据库系统体系结构	12
1.6 数据库管理系统	17
1.7 数据库应用系统开发	23
2 数据模型	25
2.1 数据建模	25
2.2 数据模型的进展	28
2.3 层次模型	31
2.4 网状模型	36
2.5 关系模型	43
2.6 实体—联系(E-R)模型	45
2.7 OODM	49
2.8 扩展的关系模型(对象/关系模型)	51
2.9 小结	51

3	关系数据库	57
3.1	关系与关系变量	58
3.2	关系数据库模型	62
3.3	关系代数	65
3.4	关系演算	72
3.5	RDBMS 的级别	79
4	结构式查询语言 SQL	82
4.1	SQL 的诞生与发展	82
4.2	SQL DDL	83
4.3	SQL DML	88
4.4	SQL 查询	89
4.5	SQL DCL	100
4.6	嵌套 SQL	101
4.7	查询优化	102
5	关系模型设计理论	109
5.1	函数相关	111
5.2	1NF 2NF 3NF	113
5.3	BCNF	117
5.4	4NF	118
5.5	5NF	120
5.6	关系规范化的步骤	122
5.7	关系规范化范例	123
6	数据库设计	132
6.1	数据库生命周期	132
6.2	数据字典	135
6.3	数据库设计过程	137
6.4	数据库的完整性	151
6.5	数据库安全性	155
6.6	数据库恢复策略	158

7	常规数据库	166
7.1	信息资源数据库的分类	166
7.2	文献资源的数字化与标准化	167
7.3	MARC21(ISO 2709/XML)	176
7.4	联机文献数据库结构	191
7.5	联机文献库的设计与建立	193
7.6	联机文献库的运行与维护	196
7.7	多语种文献数据库	198
7.8	事实数据库	202
7.9	数值数据库	205
7.10	科学数据库	210
8	多媒体数据库	215
8.1	多媒体与多媒体技术	215
8.2	多媒体数据库的关键技术	217
8.3	多媒体信息的数字化	219
8.4	多媒体数据库的设计与建立	222
8.5	Oracle 中多媒体数据库的建立方法	225
8.6	网上多媒体数据库的广泛应用	230
9	Web 数据库	237
9.1	Web 数据库的诞生	237
9.2	Web 数据库的基本结构及基础技术	238
9.3	Web 数据库的特点	239
9.4	Web 服务器与数据库服务器的连接技术	239
9.5	ASP 连接 SQL Server 进行数据库操作的示例	254
9.6	Jsp 连接 Oracle 的操作示例	267
9.7	虚拟数据库	275
10	数据压缩	279
10.1	数据压缩的意义	279
10.2	逻辑压缩与物理压缩	281
10.3	文献数据库数据压缩的常用方法	296
10.4	大型联机数据库数据压缩的实例	302

10.5	多媒体数据的压缩技术	307
11	数据库技术的新进展	314
11.1	分布式数据库	314
11.2	数据仓库与数据挖掘	317
11.3	知识库和智能数据库系统	319
11.4	工程数据库	322
11.5	并行数据库	324
11.6	主动数据库(active database)	326
11.7	模糊数据库(fuzzy database)	330
11.8	XML 数据库	333
12	RDBMS 实例	340
12.1	Access 系统	340
12.2	VFP 系统	341
12.3	DB2 系统	343
12.4	Informix 系统	348
12.5	Ingres 系统	351
12.6	ORACLE 系统	352
12.7	Sybase 系统	368
12.8	Microsoft SQL Server	370
	主要参考文献	389
	索引	393

1 数据库与数据库系统

当今世界,科学技术飞速发展,人类由工业社会向信息社会迈进。信息已成为继物质、能源之后的“第三级资源”,它是国家的战略资源和人类的宝贵财富。各国竞相开发信息技术、发展信息产业、鼓励信息生产、促进信息消费。信息生产与消费水平已成为衡量一个国家发展状况和综合国力的重要尺度。

随着社会信息化、信息社会化程度的提高,管理和利用海量信息是我们面临的紧迫任务。人们在实践中认识到数据库技术是信息资源管理的最佳技术。

1.1 数据与信息资源

计算机每日都在处理着大量数据,人们时刻都在共享信息资源。在日常生活中,数据和信息是密不可分的。

1.1.1 数据资源

数据是未经加工处理的事实,即原始的、“生”的事实。在日常生活中,我们会接触到大量的数据。

例如,东方商城每天要接待 10 万名顾客,每天大约要开 68 000 张发票。而这 68 000 张发票的汇总、分析、统计可以向我们提供大量的信息,如:顾客信息、商品流通信息、营业额、热销商品信息、商城的业绩等。我们可以用各种报表、直方图、折线图或圆饼图准确而又形象地表示出来。

在日常生活中,我们每天可以从网上和电视上看到股市行情。如,一个股票交易日从早上开盘起,各只股票随着时间的推移,股票行情千变万化。我们看到的这些仅仅是一些数据,它是未经加工的事实。但只要我们调出某股票

的 K 线图(见图 1.1),就会立即获得这只股票的具体信息。通过价格波动曲线显示,我们不仅可以清楚地看到该只股票的价格变化信息,而且通过 K 线图分析可以预测其走势。及时获取准确信息是决定投资策略的依据,正确的决策是在全球市场中求生存、谋发展的关键。

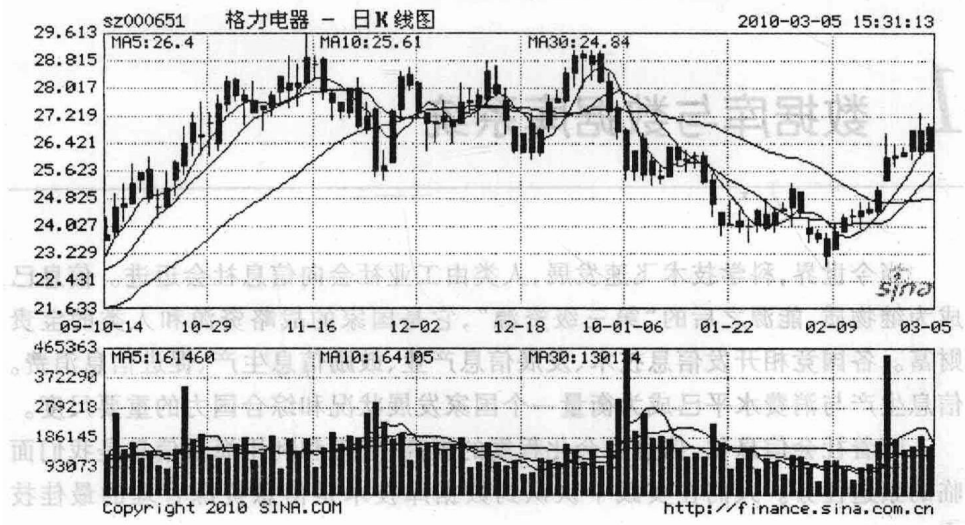


图 1.1 格力电器 K 线图

从以上实例可知:信息块是由数据组成的;只有处理过的数据才能产生信息;只有准确的数据才能加工、生产出正确的信息;信息像物资、能源一样是国家的战略资源;一个国家信息生产的规模和信息消费水平是衡量一个国家综合国力和发展水平的重要尺度。

1.1.2 数据、信息与知识的关系

数据(data)是事实、概念或指令的一种形式化的表示形式,适合于用人工或自然方式进行通信、解释或处理。它是离散的、互不关联的客观事实,孤立的文字、数值和符号,缺乏关联和目的性。

信息(information)是数据所表达的客观事实。人们对数据进行系统组织、整理和分析,使其具有相关性。数据是信息的载体,信息是数据的内容。信息和数据在有些情况下不严格区分。

数据、信息和知识的关联性十分密切(如图 1.2 所示)。知识(knowledge)是由信息加工和提炼而成的结晶。信息是加工知识的原材料,它是现代社会

的宝贵资源。在知识经济时代的今天,知识尤为重要。知识是人们对客观规律性的认识。

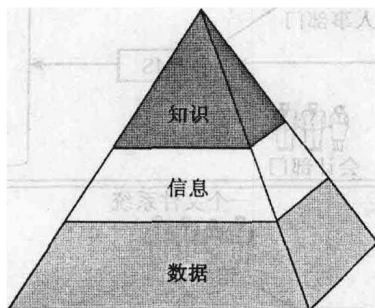


图 1.2 数据、信息与知识的关系

1.2 数据库的诞生与发展

从 1946 年 ENIAC(Electronic Numerical Integrator and Calculator)问世,计算机的应用日益广泛。先以科学计算为主,后来是数据处理唱主角。在数据处理阶段,人们先用文件系统管理信息。由于文件系统的先天不足,实践迫切要求人们采用一种优于文件系统的新技术,数据库就应运而生了。

数据库(data base)是关于某企(事)业单位业务管理的、集成的有穷数据集合。数据库里不仅存储数据,而且还存储数据之间的联系。

首先,一个单位的信息管理往往用多个实体描述,数字化之后对应着多个记录类型。数据库是结构化的,它定义了这些记录类型的数据结构并进行集成。

其次,描述数据和数据之间的联系,这就是元数据(metadata)。因而,数据库里不仅存储用户数据,而且还存储描述这些数据的元数据。

最后,需要说明的是数据库与文件系统之间的关系:数据库是由文件系统发展而来的,但它与文件系统有着本质的差别(如图 1.3 所示)。计算机用于数据管理经历了三个发展阶段。

1.2.1 初级管理阶段(20 世纪 50 年代中期以前)

数据处理初级管理阶段的管理水平不高,主要是受计算机硬件和软件的限制。当时没有直接存取存储设备,也没有管理信息资源的系统软件,因而数据处理停留在批处理阶段。这个阶段的数据管理特点(如图 1.4 所示)为:

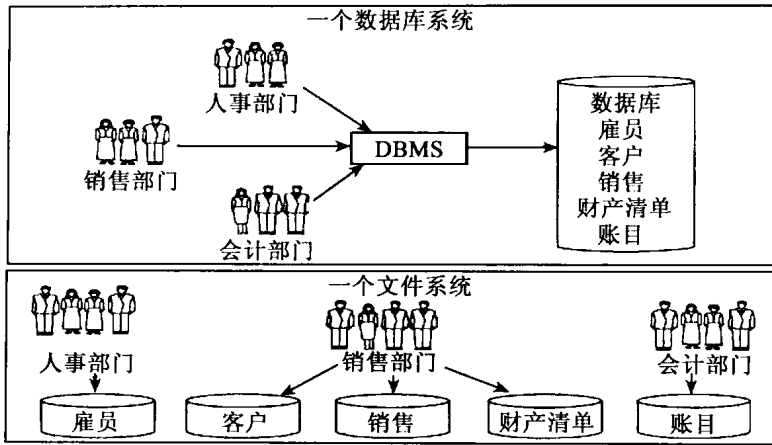


图 1.3 数据库和文件系统的比较

(1) 数据不长期保留在外存。

(2) 由于系统没有管理数据的软件, 所以在使用时由用户自己管理数据。数据与程序之间的独立性比较差。

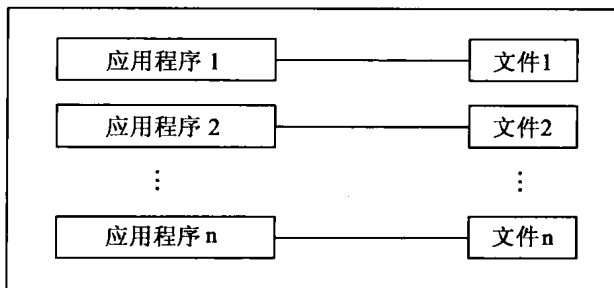


图 1.4 初级文件阶段

(3) 基本上没有文件的概念(只有顺序文件)。

(4) 数据文件与程序一一对应。由于数据冗余度大, 因而数据更新容易出错。

1.2.2 文件系统阶段(20 世纪 50 年代后期至 60 年代中期)

在这一阶段, 由于计算机有了较大的发展, 因而为数据文件(顺序文件)