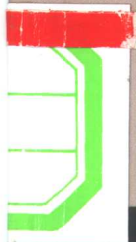


高等学校
信息管理类专业
信息核心课程教材

信息资源数据库

周宁 编著

武汉大学出版社



高等学校信息管理类专业核心课教材

信息资源数据库

周 宁 编著

武汉大学出版社

图书在版编目(CIP)数据

信息资源数据库/周 宁编著. —武汉:武汉大学出版社,2001.3
ISBN 7-307-03112-4

I. 信… II. 周… III. 信息资源—数据库系统 IV. TP311.135.1

中国版本图书馆 CIP 数据核字(2000)第 52637 号

责任编辑:曾建华

责任校对:程小宜

版式设计:支 笛

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.whu.edu.cn)

发行:新华书店湖北发行所

出版:武汉市新华印刷厂

开本:880×1230 1/32 印张:10.75 字数:305千字 插页:2

版次:2001年3月第1版 2001年3月第1次印刷

ISBN 7-307-03112-4/TP·93 定价:15.00元

版权所有,不得翻印;所购教材,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。

前 言

在当今知识经济时代，信息是物资、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、模糊数据库等新型数据库不断涌现，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——数据库系统原理，分五章（第一至第五章）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等方面进行了讨论。数据字典是设计和管理数据库的工具，也在本章进行了专门介绍。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——信息资源数据库，分四章（第六至第九章）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）分别进行了讨论，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、建立与应用等方面进行了讨论。第九章为 Web 数据库，它是因特网快速发展的产物。本章讨论了其特点、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——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。第十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，将其安排在附录中，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，作者得到了中国科学技术信息研究所、中国科学院等单位的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

目 录

| | | |
|-----|--------------------|----|
| 1 | 数据库系统引论 | 1 |
| 1.1 | 信息资源管理与数据库技术 | 1 |
| 1.2 | 信息模型 | 9 |
| 1.3 | 数据模型 | 15 |
| 1.4 | 三级模式 | 25 |
| 1.5 | 数据库系统体系结构 | 29 |
| 1.6 | 数据库管理系统 | 33 |
| | 习题一 | 40 |
| 2 | 关系数据库系统 | 42 |
| 2.1 | 关系与关系模式 | 42 |
| 2.2 | 关系数据库模型 | 47 |
| 2.3 | 关系代数 | 51 |
| 2.4 | 关系演算 | 60 |
| 2.5 | RDBMS 的级别 | 68 |
| 2.6 | RDBMS 实例 | 70 |
| | 习题二 | 86 |
| 3 | 结构式查询语言 SQL | 88 |
| 3.1 | SQL 简介 | 88 |
| 3.2 | SQL DDL | 89 |

| | | |
|-----|--------------|-----|
| 3.3 | SQL 查询 | 93 |
| 3.4 | SQL DML | 103 |
| 3.5 | SQL DCL | 109 |
| 3.6 | 嵌套 SQL | 110 |
| | 习题三 | 112 |
| 4 | 关系模型设计理论 | 114 |
| 4.1 | 函数相关 | 116 |
| 4.2 | 1NF 2NF 3NF | 118 |
| 4.3 | BCNF | 122 |
| 4.4 | 4NF | 123 |
| 4.5 | 5NF | 126 |
| 4.6 | 关系规范化的步骤 | 128 |
| | 习题四 | 129 |
| 5 | 数据库设计 | 130 |
| 5.1 | 数据库的设计条件 | 130 |
| 5.2 | 数据字典 | 132 |
| 5.3 | 数据库设计过程 | 134 |
| 5.4 | 数据完整性措施与恢复策略 | 143 |
| 5.5 | 信息资源数据库的分类 | 146 |
| | 习题五 | 147 |
| 6 | 文献资源数据库 | 149 |
| 6.1 | 文献资源的数字化 | 149 |
| 6.2 | 文献信息的标准化 | 150 |
| 6.3 | 联机文献数据库结构 | 167 |
| 6.4 | 联机文献库的设计与建立 | 170 |
| 6.5 | 联机文献库的维护 | 173 |
| 6.6 | 多语种文献数据库 | 176 |

| | |
|--|-----|
| 6.7 文献数据库实例 | 178 |
| 习题六 | 180 |
| 7 事实数据库与数值数据库 | 182 |
| 7.1 科技成果数据库 | 182 |
| 7.2 物价数据库 | 185 |
| 7.3 科学数据库 | 190 |
| 习题七 | 197 |
| 8 多媒体数据库 | 198 |
| 8.1 多媒体与多媒体计算机系统 | 198 |
| 8.2 多媒体数据库的关键技术 | 201 |
| 8.3 多媒体信息的数字化 | 202 |
| 8.4 多媒体数据库的设计与建立 | 206 |
| 8.5 网上多媒体数据库的广泛应用 | 210 |
| 习题八 | 216 |
| 9 Web 数据库 | 217 |
| 9.1 Internet 的飞速发展 with Web 数据库的诞生 | 217 |
| 9.2 Web 数据库的基本结构及基础技术 | 218 |
| 9.3 Web 数据库的特点 | 219 |
| 9.4 Web 服务器与数据库服务器的连接技术 | 220 |
| 9.5 虚拟数据库 | 231 |
| 习题九 | 232 |
| 10 数据压缩 | 233 |
| 10.1 为什么要进行数据压缩 | 233 |
| 10.2 逻辑压缩与物理压缩 | 235 |
| 10.3 文献数据库数据压缩的常用方法 | 253 |
| 10.4 大型联机数据库数据压缩的实例 | 260 |

| | |
|------------------------------|------------|
| 10.5 多媒体数据的压缩技术 | 265 |
| 习题十 | 272 |
| 11 数据库技术的新进展 | 273 |
| 11.1 分布式数据库 | 273 |
| 11.2 数据仓库与数据挖掘 | 277 |
| 11.3 知识库和智能数据库系统 | 280 |
| 11.4 工程数据库 | 284 |
| 11.5 并行数据库 | 286 |
| 11.6 主动数据库 (Active Database) | 288 |
| 11.7 模糊数据库 (Fuzzy Database) | 293 |
| 习题十一 | 297 |
| 附录 A IMS 系统简介 | 298 |
| 附录 B DBTG 系统简介 | 309 |
| 索引 | 324 |
| 主要参考文献 | 329 |

1

数据库系统引论

1.1 信息资源管理与数据库技术 ■

1.1.1 信息资源与信息社会 ●

当今世界，人类每时每刻都在生产信息，消费信息。信息（information）是物资、能源之后的“第三级资源”。在各种媒体上我们可以看到这样一些报道：一条信息救活了一个企业，一条信息挽救了一个人的生命，一条信息能赢得一场战争……人们在实践中认识到：迅速、方便地获取所需信息是正确决策、增强竞争力、取得事业成功的关键。因而，各国政府把信息当做战略资源和宝贵财富，竞相开发信息资源和网络技术。人类已进入了网络信息时代。据2000年1月的统计数据：因特网上可编索引的网页已超过10亿，网站493万多个，网民2亿。网上信息资源已汇成汪洋大海。它隐藏着无限商机，又提出了严峻挑战。人类正迎接这一挑战，积极创造条件，由“工业社会”向“信息社会”过渡。

所谓信息社会，就是信息成为比物资和能源更为重要的资源，以信息价值的生产为中心，使社会和经济发展起来的社会。

在信息社会中，人们认为只有数字化、上网的信息才是有效信息。信息资源的数字化，信息服务的网络化是发展的主要趋势，而信息资源的数字化、自动化管理的核心技术是数据库（database）。

数据库技术是 20 世纪 60 年代末发展起来的一门信息管理技术。它是信息资源开发、管理和服务的最有效的手段。目前，数据库的建设规模、信息量的大小和使用频率已成为衡量一个国家信息化水平和综合国力的重要尺度。如何设计、建立和应用数据库已成为当前热点课题之一。

在数字化的信息资源中，信息大体可分为两类：

一类信息能够用数据或统一的结构加以表示，我们称之为结构化数据，如：数字、字符；而另一类信息根本无法用数据或者统一的结构表示，如：文本、图像、声音乃至网页等，我们称之为非结构化的数据。结构化数据是非结构化数据的特例，而非结构化数据是结构化数据的一般情况。它包含结构化数据。

1.1.2 信息处理的进展 ●

信息即音信、消息和报道。它是人们对现实世界的了解，是帮助人们作出正确决策的知识。如交通民警观察到迎面开来的一辆汽车，凭借观测能知道该汽车的车型、颜色、外形、运行速度和噪音等信息。同时通过观察各个路口的交通状况，可立即发出正确的指挥信号。这些信息可以用文字、图形和声音的数字表示存入计算机。被存入计算机的文字、图形和声音统称数据（data）。数据是信息的载体，信息是数据的含义。

信息和数据是紧密相联的。在某些场合，对二者又不加严格区分。可以把信息处理系统称作数据处理系统。所谓数据处理系统，就是对信息的搜集、加工、存储、检索、传递和利用等一系列活动的总称。随着计算机技术的发展，数据处理发展很快。在短短的 50 多年中，其发展大体上已经历了三个阶段。

(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、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 1.1 所示）为：

① 数据不长期保留在外存。

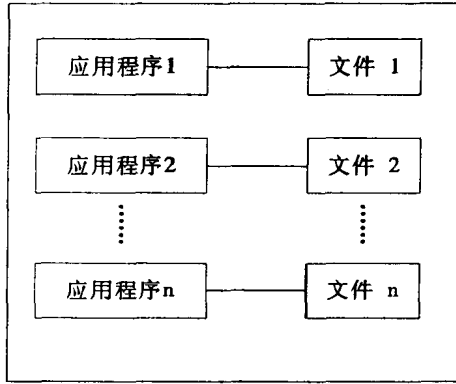


图1.1 初级文件阶段

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。

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。

(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）向文件系统过渡提供了较好的物质基础。硬件提供了直接存取设备（磁盘），软件有了操作系统，因而既可以批处理，又可以进行实时处理。

这一阶段的数据管理特点（如图 1.2 所示）为：

① 数据可长期保留在外存。

② 由于有文件管理或数据管理系统软件，勿须用户直接管理外存上的数据，提高了数据与程序的独立性。

③ 除顺序文件外，还提供了多种文件组织（如索引文件、随机文件、链接文件、倒排文件等）。

④ 程序与数据的逻辑关系仍基本是一一对应的。由于数据的

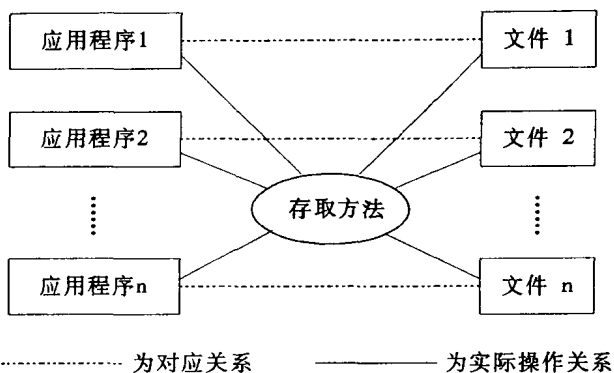


图1.2 文件系统阶段

冗余度大，因而文件不易更新，没有反映出信息之间的自然联系。

(3) 数据库系统阶段（20 世纪 60 年代后期至今）

60 年代末计算机已大量用于数据处理。计算机管理的数据范围越来越大，数据量成倍地增长，而其冗余量也大大增加。这样，人们很自然地希望多个用户共同使用公共数据，并希望用多种语言来操作数据。由于硬件技术的迅速发展不仅已能提供百兆字节到千兆字节的外存储设备，且硬件价格也不断下跌，但软件成本却不断上升，故要求人们设法在降低软件研制和维护费用的情况下，更有效地来利用数据。这样，就出现了将面向单个应用的文件系统集成成为共享的面向企业整体的数据库系统（如图 1.3）。

这是客观的需要，也是文件系统向数据库系统发展的必然趋势。同时，现阶段的科学技术条件为数据库技术的开发提供了可能性。随着整个科学技术的发展，人们已充分认识到信息（数据）的重要性。信息、能源与材料已被视为现代科学技术的三大支柱。像物资、能源那样，数据库中的各种数据已成为社会的宝贵财富。

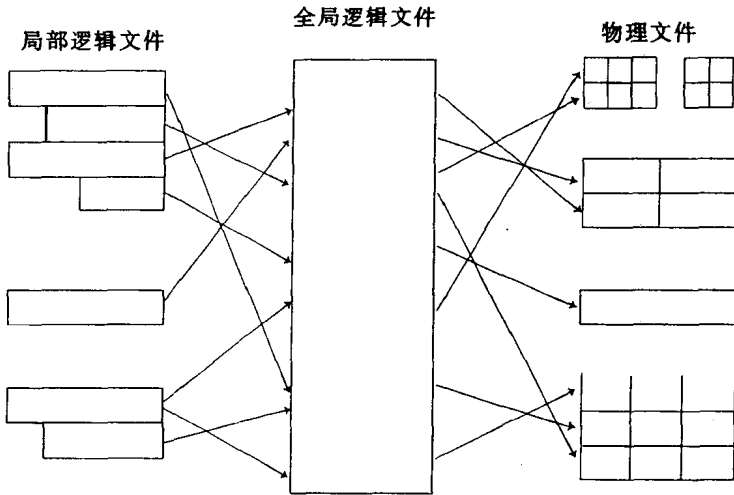


图1.3 数据库系统阶段

1.1.3 数据库的定义 ●

数据库是现代生产管理和科学技术发展的产物。它是人类从工业社会向信息社会过渡的重要标志之一。

什么是数据库？国内外的专家、学者们给出了不少的定义，下面列举一些有代表性的描述。戴特先生在《数据库系统导论》一书中对数据库这样定义：“数据库是供某些特定企（事）业的应用系统使用的存储操作数据的集合。”马丁先生在《计算机数据库组织》一书中这样描述：“数据库是多种记录类型出现的集合。它包含记录、数据聚合和数据项之间的联系。”日本学者中原啓一在《情报检索》一书中说：“数据库是一个某一时候是某一资料档、别的时候是别的资料档、对不同的利用者是不同的资料档、并可按不同的利用者和不同的利用目的组成若干逻辑资料档的数据集合。”

以上这些定义，虽然形式各有特色，描述方法不同，但其实质是一样的。这就是：（1）数据库是一个有穷的集合体。（2）数据

之间有较复杂的逻辑关系。数据库不仅存储数据，而且连同其间的逻辑关系一起存储。(3) 数据库具有数据共享和数据独立等特征。由此，我们给数据库下这样一个定义：数据库是多个记录型数据集成的集合体。它不仅存储数据，而且还存储数据之间的联系。

1.1.4 数据库的目标 ●

人们为什么要开发数据库呢？这是由于数据库具有文件系统所无法拥有的优点，它与文件系统有着本质的区别。

(1) 数据共享性

数据共享 (data sharing) 是建立数据库的主要目标之一。传统的计算机文件通常是为一应用目的而设计的，一般为一种应用专用。而数据库是为多个用户、多种应用目的而建立的，可以同时为多个用户服务。这样，数据库能减轻用户负担，降低系统成本，促进应用的发展。社会发展要求社会信息化，信息社会化，信息资源共享是开发信息资源数据库的动力。

(2) 数据独立性

数据独立性 (data independence) 是指数据与应用之间可以互相独立，包括物理数据独立性和逻辑数据独立性。所谓逻辑数据独立性，是指局部逻辑数据结构与总体逻辑数据结构间的独立。当总体逻辑数据结构变化时，局部逻辑数据结构可以不变。这样，当添加新的数据项扩充整体逻辑数据结构时，不必改写应用程序。所谓物理数据独立性是指应用程序对存储结构与存取方法的独立性。当存储结构与存取方法改变时，不会影响逻辑数据结构，应用程序也不必重写。数据的物理独立性与逻辑独立性是由两个映象来实现的，如图 1.4 所示。当数据库整体结构和存储结构改变时，只须调整这两个映象就可以实现。

(3) 最小冗余性

图 1.5 为一个数据库系统的简图。数据库可为多个用户共享。这就是说，数据库的数据是“集成的”，即数据库是多种性质不同的数据文件的统一体。这些文件中的各种冗余 (redundancy) 可以

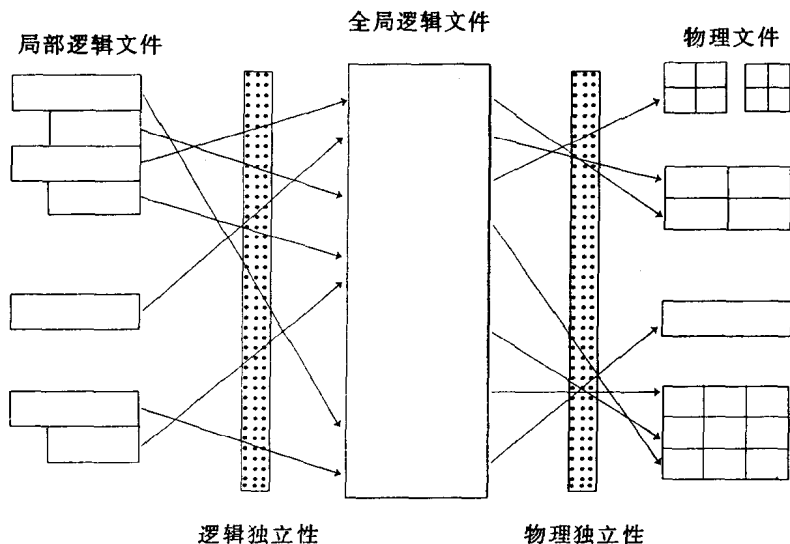


图 1.4 数据独立性

部分地或整体地删掉，从而实现了数据的最小冗余。为了提高响应速度，有时也有意保留部分冗余数据。但数据库总是把数据冗余限制在尽可能小的范围内。

(4) 安全性

数据库是为多个用户和多种应用目的服务的，因而实现了数据共享，但在共享中有安全和保密的问题。有些数据要严格控制共享的范围，防止对数据的不合法的使用。数据的保密和安全是彼此相关的两个问题。数据的保密涉及由个人或团体控制他们自己的信息的收集、存储、传播和使用权。安全性涉及对数据的存取控制、修改和传播的技术手段，即实现对数据库中数据的合法权检验，防止不合法的查询、存入、更改和破坏数据库数据的事件发生。

系统在保证数据库数据的安全性方面可采用多种方法。常用的有：

① 识别与鉴定