

高等学校文科教材

情报数据库系统

(第二版)

周 宁 编著

武汉大学出版社

高等学校文科教材

情报数据库系统

(第二版)

周 宁 编著

武汉大学出版社

(鄂)新登字 09 号

图书在版编目(CIP)数据

情报数据库系统/周宁编著·—2 版

——武汉:武汉大学出版社,1994.7

高等学校文科教材

ISBN 7-307-01797-0

I 情…

I 周…

Ⅲ 情报—数据库管理系统

NG356.1 TP311.13

武汉大学出版社出版

(430072 武昌 珞珈山)

武汉大学印刷厂印刷

新华书店湖北发行所发行

1988年8月第1版 1994年7月第2版

1994年7月第2版第1次印刷

开本:850×1168毫米 1/32 印张:14.5

字数:373千字 印数:1—3000

ISBN 7-307-01797-0/G·246 定价:8.15元

第二版前言

数据库系统从诞生至今已经历了第一代、第二代，现正在开发第三代数据库系统。数据库理论与方法有新的突破，数据库技术迅速发展。

面向对象的数据库系统、知识库与基于知识的系统、分布式数据库系统与数据库集成技术、时态与工程数据库、多媒体数据库等已成为数据库研究与应用领域的重要课题，特别是多媒体数据库的发展给数据库应用带来了革命性的变化。它正影响着我们的工作和生活，并将更加影响人们的未来。

《情报数据库系统》第二版的内容比第一版内容有较大的更新。具体增写了第十章以及第六章的第7、8两节、第七章第3节，重写了第九章和第五章的第6节。另外，对第一、二、三、四、八章的内容也进行了修改和补充。

这次再版得到了许多同行专家的关心和支持，许多同事和武汉大学出版社的同志们给予了具体支持和帮助，撰稿过程中参阅了大量文献，在此一并致以诚挚的谢忱。

作·者

1994年1月于珞珈山

第一版前言

在信息时代的今天，信息与物资、能源一样，已成为社会的宝贵资源。目前，信息量成指数函数的速度增长，处理这些信息的计算机系统，从以处理为中心，变成了以数据库为中心。数据库系统是信息处理系统的核心。

十多年来，数据库系统在各行各业、各条战线、政府部门、科研机构和军事部门等得到了广泛应用，成为人们的得力助手。

从70年代开始，一些国家的计算机专业和计算机应用专业已先后开设了数据库系统的课程。它对于普及数据库知识、研究和发挥数据库技术发挥了重要作用。

数据库系统课程的教学在我国从80年代初开始迅速发展，近几年在全国重点院校已得到普及。1986年8月，由国家教委高教一司委托召开了全国数据库概论教学研讨会。在这次会上有40多名代表交流了教学经验。会议提出并讨论通过了计算机专业本科生、研究生和非计算机专业本科生的数据库系统教学大纲。编著者根据非计算机专业用《数据库系统引论及应用》教学大纲的要求，总结了我院从1981年以来的教学经验，在前两次编印讲义的基础上编写了本教材。

本书分三部分：第一部分（第一章）为数据库系统引论。第二部分（第二——四章）为流行的数据库系统。第三部分（第五——九章）为情报数据库，这一部分分章讲述情报库的设计、生成与维护，情报库的利用、数据压缩和情报数据库系统的发展趋势等。

第三部分为本书的重点，其中许多思想与实例是编著者近几年科研实践的总结；同时也引述了国际上情报数据库系统的经验。

为了适应有关专业的使用需求，将超过教学大纲部分用星号（“*”）标识，以供教师在教学中取舍。

由于编著者水平有限，书中的缺点错误在所难免，诚恳希望读者批评指正，以便再版时修改。

在本书的编写过程中，作者得到了中国科学院科学数据库筹备处、中国科技情报所、国家科委情报局、国防科委情报所以及一些重点大学的许多专家、教授的支持和帮助。特别是我校计算机科学系郑振楣、石树刚副教授在百忙之中审阅了全稿，在文字和内容上都提出了许多建设性的意见。受国家教委委托，武汉大学于今年7月初专门召开了审稿会议。参加审稿会的有：冯玉才、许云涛、王业嘉、郑振楣、彭琪、石树刚、赖茂生、徐进鸿、萧作铭、陶梅生、彭斐章、王昌亚、何皓、王大可等同志，因故未能到会的曾民族、张凤楼、高崇谦同志提前寄来了书面意见。参加审稿的同志对本教材提出了一些宝贵意见。武汉大学出版社的同志为本书的出版付出了辛勤劳动。在此，对所有为本书的编写、审稿和出版给予支持和帮助的同志一并表示衷心的感谢。

编 著 者

1987年9月1日

目 录

第一章 数据库系统引论	(1)
1.1 信息时代与数据库	(1)
1.2 信息模型	(11)
1.3 数据模型	(17)
1.4 数据库系统体系结构	(28)
练习一	(39)
第二章 关系数据库系统	(42)
2.1 关系与关系模型	(42)
2.2 关系代数	(49)
2.3 关系演算	(58)
2.4 SYSTEM R 简介	(61)
2.5 SQL	(69)
2.6 QBE	(88)
2.7 关系的规范化理论	(93)
练习二	(107)
第三章 IMS 系统	(111)
3.1 概述	(111)
3.2 IMS 数据模型	(112)
* 3.3 IMS 的逻辑数据库	(119)
* 3.4 IMS 存贮结构	(122)
* 3.5 数据操作语言 DL/1	(137)
练习三	(141)
第四章 DBTG 系统	(143)
4.1 概述	(143)

4.2	DBTG 数据模型	(146)
* 4.3	模式与子模式	(153)
* 4.4	运行单位与当前值	(167)
* 4.5	数据操作语言	(169)
	练习四	(183)
第五章	情报数据库的设计	(185)
5.1	情报数据库的设计条件与过程	(185)
5.2	数据字典	(187)
5.3	系统分析	(190)
5.4	系统设计	(195)
5.5	系统实现	(197)
5.6	情报数据库的分类	(199)
	练习五	(200)
第六章	文献数据库	(201)
6.1	文献数据库的产生与发展	(201)
6.2	文献目录数据的加工与存贮	(206)
6.3	文献目录数据的标准化	(211)
6.4	联机文献数据库结构	(225)
6.5	联机文献库的设计与建立	(228)
6.6	联机文献库的维护	(238)
6.7	全文数据库	(245)
6.8	多语种文献数据库	(246)
	练习六	(248)
第七章	非文献型数据库	(249)
7.1	数值数据库	(250)
7.2	事实数据库	(267)
7.3	多媒体数据库	(274)
7.4	小结	(276)
	练习七	(278)

第八章 数据压缩	(279)
8.1 数据压缩的意义	(279)
8.2 逻辑压缩与物理压缩	(281)
8.3 压缩技术与压缩策略	(298)
8.4 情报数据库数据压缩的常用方法	(299)
8.5 大型联机情报库数据压缩的实例	(307)
练习八.....	(313)
第九章 数据库技术的新进展	(314)
9.1 分布式数据库系统	(314)
9.2 数据库机	(321)
9.3 知识库和智能数据库系统	(324)
9.4 面向对象的数据库	(334)
练习九.....	(337)
第十章 几个具体的数据库系统	(338)
10.1 INGRES 系统	(338)
10.2 ORACLE 系统	(354)
10.3 dBASE 系统	(372)
10.4 Foxbase 与 FoxPro	(395)
练习十.....	(402)
附录 数据的物理组织	(403)
1. 概述.....	(403)
2. 存贮结构.....	(404)
3. 文件组织.....	(413)
主要参考文献	(443)
索引	(447)

第一章 数据库系统引论

1.1 信息时代与数据库

信息是国家的战略资源和巨大财富，是社会、经济、科技发展的重要基础。当今世界正进入信息经济时代，信息资源的开发利用水平已成为衡量一个国家综合国力的重要尺度。

当前，社会、经济、科技信息量剧增。人们惊呼“信息爆炸”和“信息危机”。实践使人们认识到：各种信息的电子化，用电子计算机来管理信息是解决“信息爆炸”的最佳方法。

为了适应信息时代的要求，各国政府都在积极采取对策。努力开发信息资源，强化采用现代化技术，积极开拓信息市场，高度重视信息产业。而采用现代化技术，主要是指用计算机管理信息，建立各种信息系统。计算机以惊人的效率出现在工农业生产、交通运输、商店、银行、机关、学校、科研机构、图书情报档案单位等各个领域。当许多台计算机通过通信网络联结起来时，就可能实现全部信息的电子化，并由计算机来管理整个系统中的信息交换。电子帐册和电子档案等将使企业的管理实现现代化。电子邮政和电子电话将使办公室工作实现自动化。电子报纸、书刊、电子图书馆等将使人们的文化生活现代化。电子订票、电子购物和家用电子信息系统等将使家庭生活走向自动化。所谓信息化社会，就是信息成为比物资或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。

目前，计算机的应用向纵深发展，日益普及。从儿童玩具到航天飞机，从设计制造到医疗诊断，从银行存取款项到飞机订票

系统，从排版印刷到文艺创作，从机器翻译、声像识别到用计算机朗读新闻，从图书情报工作的日常管理到国际联机情报检索系统都无处不用到计算机。使用国际联机检索终端可以跨越大洲大洋检索国外的信息库。据统计，平均用 10 分钟就可查阅 30 种文字出版的 2 000 种期刊中的上百万篇论文。在整个计算机应用领域，处理文字、图形、声音等数据的信息处理系统占着最大的比重，而信息处理系统建设的核心是数据库（Data Base）。

1.1.1 信息与数据处理

信息（information，或译为情报）即音信、消息和报导。它是人们对现实世界的了解，是帮助人们做出正确决策的知识。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、图形符号、数字表示存入计算机。被存入计算机的文字、符号和数字统称数据（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为数据处理系统。所谓数据处理系统，就是对信息的搜集、加工、存贮、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，数据处理进展很快。在短短的 40 多年中，大体上已经历了三个阶段。

1. 初级管理阶段（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贮设备，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。

（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。

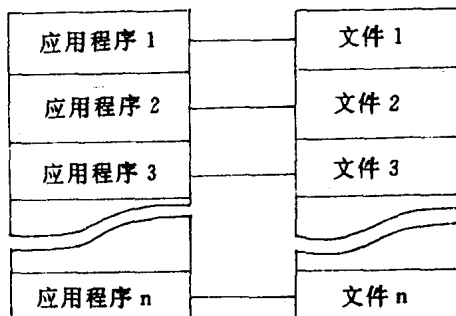


图 1.1 初级文件阶段

(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。

2. 文件系统阶段（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，又可进行实时处理。

这一阶段的数据管理特点（如图 1.2 所示）是：

(1) 数据可长期保留在外存。

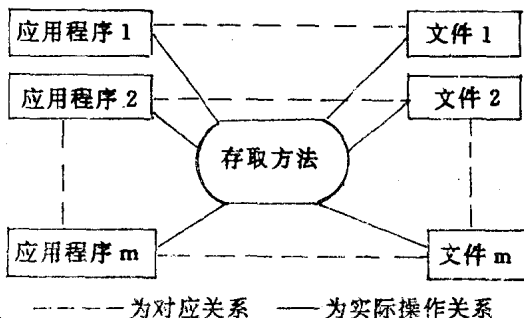


图 1.2 文件系统阶段

(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（60年代后期——）

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贮设备，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）。

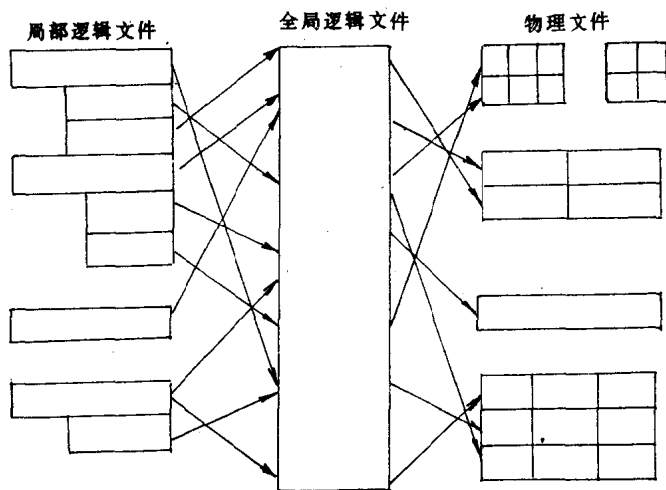


图 1.3 数据库系统阶段

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。

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。象矿产、水利资源那样,数据库中的各种数据已成为社会的宝贵财富。

1.1.2 联机情报系统

随着数据库系统的诞生和推广使用,国际联机情报系统得到极快的发展,且应用很广。联机情报系统是一种专用的数据库系统。如 Dialog 系统、ORBIT 系统和 ESA/IRS 系统等均为国际上大型的联机情报系统。其中,洛克希德公司的 Dialog 系统最大。据 1991 年统计,该系统有 430 多个数据库,3 亿个记录,10E+12 字节信息量。在这 430 多个数据库中,有 186 个文献目录数据库、86 个全文数据库(其中新闻库占 60 个)、82 个事实型(商业)数据库和一批数值型数据库。它在全世界 100 多个国家与地区有 14 万个终端用户。通过四通八达的通信网络昼夜不停地向全世界每个角落提供联机情报服务。根据 1992 年的统计,我国已在全国 50 多个城市设立了 110 多台国际终端。利用这些终端设备可随时检索 Dialog 系统、ORBIT 系统和 ESA/IRS 系统等 12 个大型信息系统中的 600 多个数据库几亿条信息记录。

一个联机情报系统的结构如图 1.4 所示。情报系统由硬件(计算机系统和通讯网络)、软件(系统软件与应用软件)、数据库(包括文献数据库、事实数据库、数值数据库等)、系统管理员和用户五个要素组成。为了满足信息时代的需要,采用数据处理的技术与方法,把情报的生产、搜集、加工、存贮、检索、传递与利用联在一起,构成了一个有机的整体。

联机情报系统有大批的终端用户,他们是情报的吸收源。系统有自己的通讯网络。小型情报系统一般有专用网和地区通讯网;

大型情报系统有国际通讯网，通过通讯卫星来实现洲际之间的通信连接。计算机是系统的核心部分。它把各用户提出的情报需求逐个对相关的数据库进行查找，把检索结果立即回送给终端用户。这些功能是靠计算机软件来实现的。系统管理员负责建立整个情报系统并监控整个系统的运行和负责系统的维护。数据库在联机情报系统中是情报源，占有极其重要的地位。

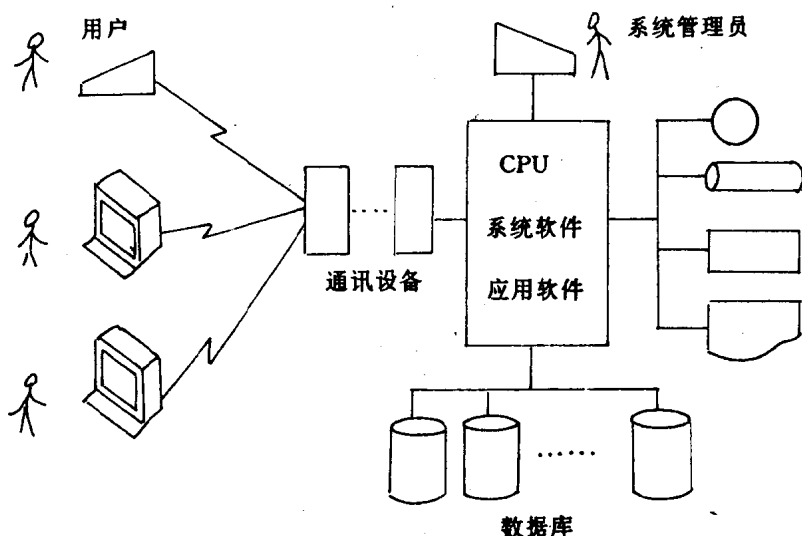


图 1.4 联机情报系统

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化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。

“数据库是供某些特定企（事）业的应用系统使用的存贮操作

数据的集合。”^①

“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。”^②

“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”。^③

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。数据库不仅存贮数据，而且连同其间的逻辑关系一起存贮。(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。它不仅存贮数据，而且还存贮数据之间的联系。

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达到的优点，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，能减轻用户负担，降低系统成本，促进应用的发展。

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

① C. J. DATE: An Introduction to Database Systems.

② J. Martin: Computer Data Base Organization.

③ 中原啓一：情报检索。

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贮结构与存取方法的独立性。当存贮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 1.5 所示。当数据库整体结构和存贮结构改变时，只须调整这两个映象就可以实现。

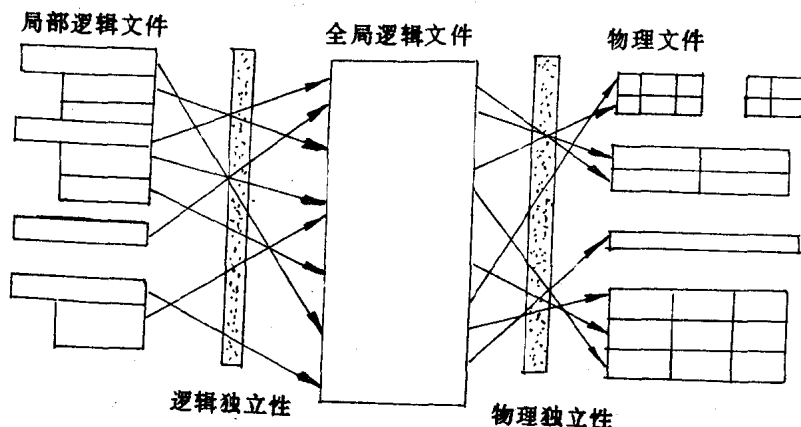


图 1.5 数据独立性

3. 最小冗余性

图 1.6 为一个数据库系统的简图。数据库可为多个用户共享。这是说，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 (security)