

情报数据库系统

高等学校文科教材

三
第
二
版
青
报
数
据
库
系
统

周 宁 编 著

武汉大学出版社

一九八八年·武昌

高等学校文科教材
情报数据库系统
周 宁 编著

*

武汉大学出版社出版
(武昌 珞珈山)
新华书店湖北发行所发行
武汉大学出版社印刷总厂印刷

*

850×1168mm 1/32 13.75印张 插页2 348千字

1988年8月第1版 1988年8月第1次印刷

印数: 1—4000

ISBN 7-307-00139-x/G·36

定价: 3.20元

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、研究和发发展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

内 容 提 要

本书属于国家教育委员会组织制定的《一九八五年至一九九〇年高等学校文科教材编选计划》的一种，供高等学校情报学专业和图书馆学专业使用，也可供有关专业选用。

本教材系统地论述了情报数据库系统的基本概念、基本理论和基本方法。主要内容有：数据库系统引论和流行的数据库系统，情报库的设计、建立、维护和利用，数据压缩以及情报数据库系统的发展趋势等。另外，在附录中还介绍了数据的物理组织以供读者参考。

本书也可供各类情报机构、图书馆、资料室、档案馆及有关单位的工作人员学习参考。

目 录

第一章 数据库系统引论	(1)
1.1 信息时代与数据库	(1)
1.1.1 信息与数据处理	(2)
1.1.2 联机情报系统	(5)
1.1.3 数据库的定义	(6)
1.1.4 数据库的目标	(7)
1.2 信息模型	(11)
1.2.1 现实世界	(11)
1.2.2 信息世界	(11)
1.2.3 计算机世界	(13)
1.2.4 信息模型	(14)
1.3 数据模型	(17)
1.3.1 层次模型	(20)
1.3.2 网状模型	(22)
1.3.3 关系模型	(23)
1.3.4 D I A M	(26)
1.4 数据库系统体系结构	(28)
1.4.1 数据库系统的定义	(28)
1.4.2 数据库系统的体系结构	(29)
1.4.3 数据库管理系统	(33)
1.4.4 数据库管理员 (DBA)	(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	(50)

2.3	关系演算	(59)
2.4	SYSTEM R 简介	(63)
2.5	SQL	(70)
2.6	QBE	(90)
2.7	关系的规模化理论	(96)
* 2.8	微机关系数据库系统的实例	(110)
2.8.1	INGRES	(111)
2.8.2	dBASE II	(127)
第三章 IMS 系统		(149)
3.1	概述	(149)
3.2	IMS数据模型	(150)
* 3.3	IMS的逻辑数据库	(157)
* 3.4	IMS存贮结构	(160)
* 3.5	数据操作语言DL/1	(176)
第四章 DBTG 系统		(182)
4.1	概述	(182)
4.2	DBTG数据模型	(185)
* 4.3	模式与子模式	(192)
* 4.4	运行单位与当前值	(206)
* 4.5	数据操作语言	(208)
第五章 情报数据库的设计		(224)
5.1	情报数据库的设计条件与过程	(224)
5.2	数据字典	(226)
5.3	系统分析	(229)
5.4	系统设计	(234)
5.5	系统实现	(236)

5.6	情报数据库的分类	(238)
第六章	文献数据库	(240)
6.1	文献数据库的产生与发展	(240)
6.2	文献数据的加工与存贮	(245)
6.3	文献数据的标准化	(251)
6.4	联机文献数据库结构	(264)
6.5	联机文献库的设计与建立	(268)
6.5.1	数据转换	(268)
6.5.2	MF的装入与MX的生成	(270)
6.5.3	抽词与排序	(271)
6.5.4	归并	(276)
6.5.5	IF文件的装入与IX文件的生成	(276)
6.6	联机文献库的维护	(277)
第七章	非文献型数据库	(286)
7.1	数值数据库	(287)
7.2	事实数据库	(304)
7.3	小结	(312)
第八章	数据压缩	(314)
8.1	数据压缩的意义	(314)
8.2	逻辑压缩与物理压缩	(316)
8.3	压缩技术与压缩策略	(333)
8.4	情报数据库数据压缩的常用方法	(334)
8.5	大型联机情报库数据压缩的实例	(341)
第九章	情报数据库系统的现状与发展趋势	(349)
9.1	概述	(349)

9.2	国际联机情报库	(350)
9.3	我国情报库的发展简况	(351)
9.4	发展趋势	(351)
9.4.1	分布式数据库	(359)
9.4.2	数据库机 (Data Base Machine)	(365)
9.4.3	知识库和智能数据库系统	(368)
附录 数据的物理组织		(380)
1.	概述	(380)
2.	存贮结构	(381)
2.1	层次模型的存贮结构	(381)
2.2	网状模型的存贮结构	(385)
2.3	关系模型的存贮结构	(389)
3.	文件组织	(390)
3.1	文件组织概述	(390)
3.2	顺序文件	(394)
3.3	倒排文件	(396)
3.4	随机文件	(399)
3.5	索引文件 (index file)	(404)
3.6	B树	(406)
3.7	IBM的ISAM与VSAM	(413)
主要参考文献		(421)
索引		(424)

第一章 数据库系统引论

1.1 信息时代与数据库

当今世界，人们时刻都在关注着信息，不断地搜集、加工、存贮信息，检索、传递和利用信息。信息已成为人们参加生产活动、科学研究、经济、贸易和各种社会活动并取得成就的关键。这一切都表明，我们正处在一个信息时代。

社会的发展，科学技术的进步使人类已步入了知识信息的海洋，不时地发出“信息爆炸”、“信息危机”的呼声。计算机的诞生与迅速发展，开创了“信息革命”的新时代。

为了适应时代的要求，各国政府都在积极采取对策，高度重视新的技术革命，竞相采用电子计算机武装各个部门。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、图书情报单位等各个领域。当许多计算机通过通讯网络建立起分布式网络系统时，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。电子电话和电子公文将使办公室工作实现自动化。电子报纸、期刊、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，计算机的应用日益向纵深发展。从儿童玩具到航天飞机，从设计制造到医疗诊断，从银行支付款项到飞机订票系统，从新闻排版到文艺创作，从机器翻译、声象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情报库。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10万篇论文。在整个计算机应用领域中,数据处理占着最大的比重,而数据处理的核⼼问题是数据库(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

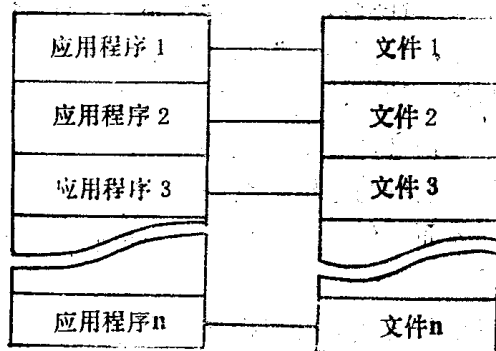


图1.1 初级文件阶段

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(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。

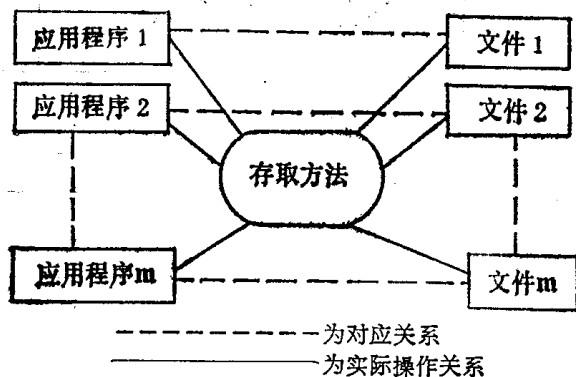


图1.2 文件系统阶段

(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——)

六十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)。

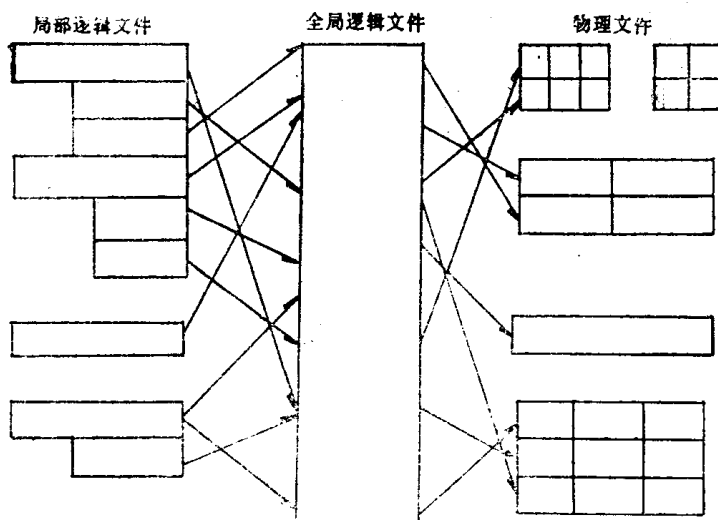


图1.3 数据库系统阶段

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。仅以Dialog系统为例，它拥有数据库220多个，约1亿条记录。该系统在世界上各大洲近百个国家和地区的2000多个城市里安装了6万多台联机检索终端。这些终端通过国际通讯网络和通讯卫星与设在美国西部的Dialog系统相连，提供各种服务活动。目前，我国大部分省市均建立了国际联机检索终端，利用本地的终端设备可以随时检索Dialog系统、ORBIT系统和ESA系统的数据库。

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

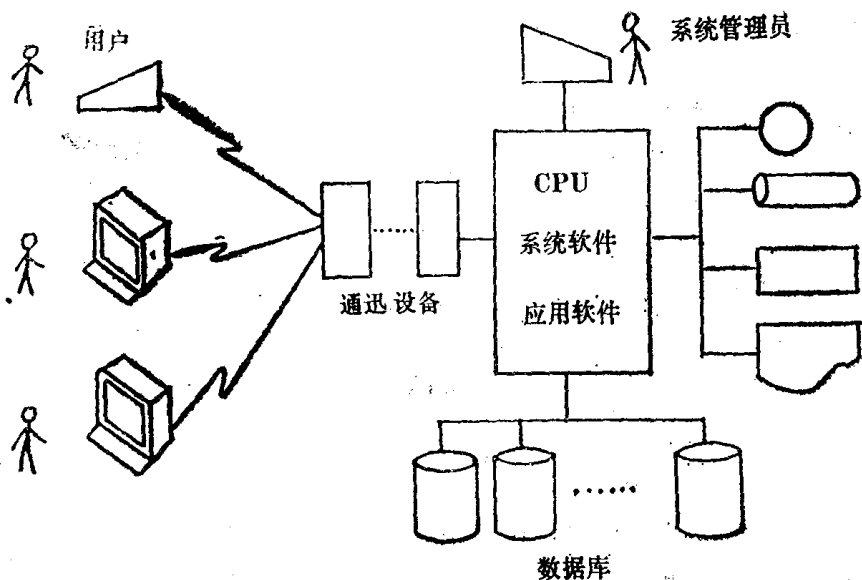


图1.4 联机情报系统

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。

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

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

合和数据项之间的联系。”^①

“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”。^②

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。

① J. Martin: Computer Data Base Organization.

②中原啓一：情报检索。