

数据库基础

北京工业大学
计算机科学系软件教研室

一九八一年九月

7221

早期的电子数据处理系统只是一些彼此无关的子系统。如工资子系统, 库存子系统等。每个子系统所带有的数据文件也只包含部分信息。六十年代中期磁盘存贮器问世以来, 产生了管理系统的概念, 即用电子计算机管理整个信息系统。然而, 对大单位管理信息系统的数据文件过多, 造成数据维护上的困难, 因而出现了数据管理的最新技术——数据库。各种不同的数据库系统按所用的数据模型而异。当前最盛行的有三种模型, 它们是层次模型, 网状模型和关系模型。

本书对数据库的基本概念, 三种模型及有关的问题都作了比较系统的介绍。并且每种模型都以一种或几种最有代表性的数据库系统作为典范, 进行较全面的介绍。

本书既可作为软件专业学生的教科书或参考书, 又可以供专门从事数据管理工作的科技人员使用。

本书在编写过程中主要参考如下文献:

〔1〕 C. J. Date: An Introduction to Database Systems (1977)

〔2〕 数据库系统讲义 (中国人民大学信息系)

〔3〕 J. Martin: Computer Database Organization

〔4〕 数据库方法概述 (国外电子技术1979年11期)

〔5〕 Ullman: Principles of Database Systems (1980)

由于编者水平有限, 书中有妥之处, 请批评指正。

北京工业大学计算机系 王连瑞

一九八一年九月一日

序 言.....	1
第一章：数据库的概念.....	8
1.1 什么是数据库.....	8
1.2 为什么需要数据库.....	11
1.3 数据库系统结构.....	11
1.4 基本术语.....	14
第二章 信息系统.....	15
2.1 信息系统.....	15
2.2 数据处理系统.....	16
2.2.1 批处理系统.....	17
2.2.2 联机系统.....	19
2.2.3 分布数据处理系统.....	19
2.3 数据管理的进展.....	20
2.3.1 无管理阶段.....	21
2.3.2 文件系统.....	21
2.3.3 数据库系统.....	23
第三章 数据模型与存贮结构.....	30
3.1 数据间的联系.....	30
3.2 格式化模型.....	31
3.2.1 格式化模型与图.....	31
3.2.2 层次模型.....	34
3.2.3 网状模型.....	34
3.3 关系模型.....	36
3.4 数据独立存取模型.....	38
3.5 数据模式的级别.....	38
3.6 数据库的存贮结构.....	40
3.6.1 树形数据的存贮结构.....	41
3.6.2 网状数据的存贮结构.....	42

第四章 数据库语言和数据处置	44
4.1 数据库语言	44
4.2 存取机制	47
4.2.1 入口点存取方法	47
4.2.2 导航的存取方法	48
4.3 数据的选取	49
4.3.1 内容地址法和数据关连法	49
4.3.2 查找	52
第五章 DBTG 系统	53
5.1 DBTG 简介	53
5.2 DBTG 系统结构	53
5.3 数据结构的分解	54
5.3.1 系型	55
5.3.2 系值	56
5.3.3 奇异系	58
5.3.4 系的例	59
5.4 系序	62
5.4.1 排序的系	62
5.4.2 时间先后的系	63
5.5 系的实现	65
5.5.1 键的方法	65
5.5.2 指引元陈列	70
5.5.3 实现系结构的其它方法	71
5.6 数据的存放	73
5.6.1 域	73
5.6.2 数据库码	74
5.7 运行单位与当前值	75
5.8 记录存放方式	76
5.8.1 LOCATION MODE IS SYSTEM(DEF- AULT)	77

5.8.2	LOCATON MODE IS CALC Procedure USING data-item, data-item.....	77
5.8.3	LOCATION MODE IS THRU Set SET...	77
5.8.4	LOCATION MODE IS DIRECT {db-id- 1 name-1 }	77
5.9	属籍类别.....	78
5.9.1	加入类别.....	78
5.9.2	移去类别.....	78
5.10	系值选择.....	78
5.10.1	SET SELECTION IS THRU Set-name-1 IDENTIFIED BY CURRENCY	79
5.10.2	SET SELECTION IS THRU Set-name-1 SYSTEM.....	80
5.10.3	SET SELECTION IS THRU Set-name-1 OWNER IDENTIFIED BY db-id-1 IN record-name-1	80
5.11	查找码.....	81
5.12	SOURCE/RESULT	81
5.12.1	ACTUAL/VIRTUAL SOURCE	81
5.12.2	ALTUAL/VIRTUAL RESULT	82
5.13	SCHEMA DDL	83
5.14	SUBSCHEMA DDL	85
5.14.1	子模式与宿主语言的关系.....	85
5.14.2	子模式与模式的区别.....	86
5.15	DML	88
5.15.1	语句概念.....	88
5.15.2	程序的运行.....	89
5.15.3	REAOY 与 FINSH	90

5.15.4	FIND	91
5.15.5	GET	99
5.15.6	MDDIFY	100
5.15.7	CONNECT 与 DISCONNECT	101
5.15.8	STORE 与 ERASE	103
5.15.9	RETAINING	104
5.15.10	ORDER	105
5.15.11	ACCEPT	105
5.15.12	KEEP, FREE 与 REMDNIOR	106
5.16	程序的例	107
第六章	IMS	110
6.1	引言	110
6.2	IMS 系统的结构	110
6.3	IMS 数据模型	111
6.3.1	物理数据库	111
6.3.2	数据库描述 (DBD)	114
6.3.3	层次序列	116
6.4	IMS 外模型	118
6.4.1	逻辑数据库	118
6.4.2	程序通讯块	119
6.5	IMS 数据子语言	120
6.5.1	定义程序通讯块 (PCB)	120
6.5.2	DL/I 操作	122
6.5.3	DL/I 例题	123
6.5.4	SSA 命令码	126
6.6	IMS 存贮结构	127
6.6.1	引言	127
6.6.2	HSAM	129
6.6.3	HISAM	129
6.6.4	HDAM	133

6 · 6 · 5	HIDAM	135
6 · 7	IMS 逻辑数据库	137
6 · 7 · 1	逻辑数据库 (LDB)	137
6 · 7 · 2	例题	137
6 · 7 · 3	数据库的描述 (DBD)	139
6 · 7 · 4	双向逻辑关系	140
第七章	关系数据库	143
7 · 1	数学概念	143
7 · 2	关系模型	144
7 · 3	数据子语言—关系代数	149
7 · 3 · 1	传统的集合运算	149
7 · 3 · 2	关系运算	151
7 · 3 · 3	检索范例	153
7 · 3 · 4	存贮范例	154
7 · 4	数据子语言—关系演算	154
7 · 4 · 1	检索操作	155
7 · 4 · 2	存贮操作	160
7 · 4 · 3	库函数	161
7 · 5	数据子语言—SQL	164
7 · 5 · 1	检索操作	164
7 · 5 · 2	存贮操作	167
7 · 5 · 3	库函数	168
7 · 6	数据子语言—Query By Example	168
7 · 6 · 1	检索操作	168
7 · 6 · 2	树结构关系的检索	171
7 · 6 · 3	存贮操作	174
7 · 6 · 4	库函数	175
7 · 7	关系的规范形	175
7 · 7 · 1	函数依赖	176
7 · 7 · 2	第一、第二和第三规范形	178

7.7.3	具有多于一个候选码的关系	183
7.7.4	第四规范形	184
7.8	外模型	187
7.8.1	外关系	187
第八章	安全性	189
8.1	引言	189
8.2	标识和鉴定	190
8.3	存取控制	190
8.4	在IMS中的安全性	192
8.5	在DBTG中的安全性	193
第九章	完整性	194
9.1	引言	194
9.2	完整性限制	195
9.3	数据共享	199
9.4	维持程序	200

第一章 数据库的概念

1.1 什么是数据库？

为了便于理解这一重要概念，下面给出一个图。

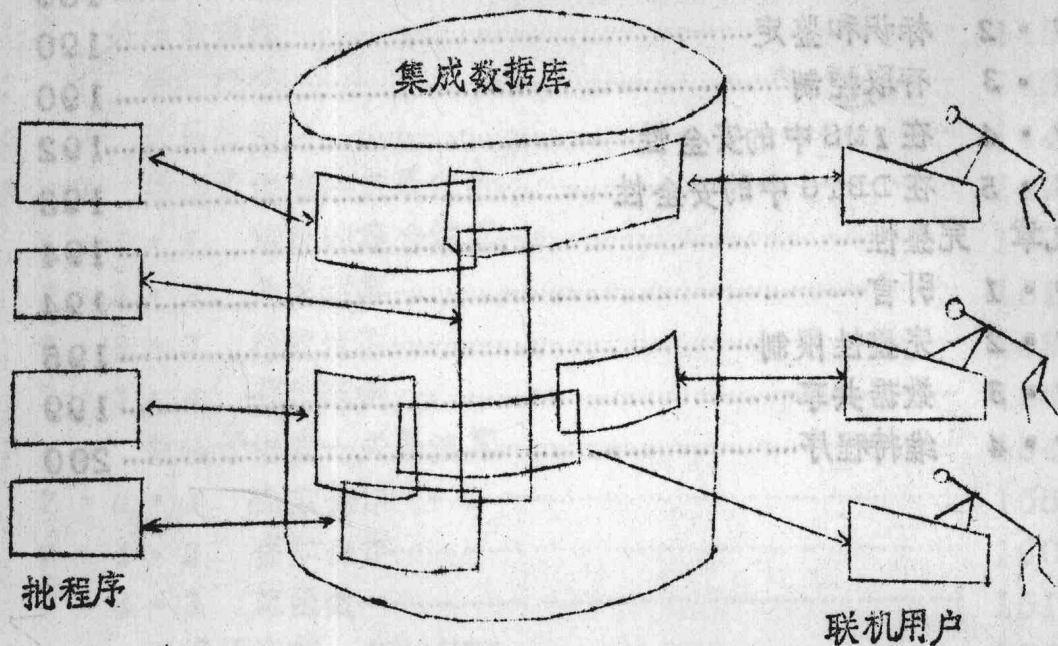


图 1.1 数据库系统的简化视图

下面对图进行简单的解释。

首先，数据库——存贮在磁盘、磁鼓或其它辅助存贮器上的数据（可能是很大的）集合。

其次，对于这些数据运行的通常是一组批应用程序，对数据库的操作通常是——检索、修改、插入、删除。此外，可以有一组联机用户，从遥远的终端与数据库联系。一般说来，它也可实现检索、修改、插入和删除的全部功能。然而在这种情况下检索是最常用和最重要的。

第三，由于数据库是“集成的”。那么这就意味着数据包着不是

一个用户的数据，而是很多用户的数据。这同样意味着，任何一个用户（批或联机）只与数据库的一小部分发生关系；此外，不同用户的数据部分可以以各种方式重迭，也就是，单独一块数据可以为许多不同用户共享。

下面我们着手数据库通用的定义。

• 数据库——是某些特定企业应用系统所使用的存贮操作数据的集合。这个定义需要作些解释。

“企业”——仅仅是对于任何相当大的商业、科学、技术或其它操作的一个方便用语。例如：

制造公司

银行

医院

大学

政府部门

任何企业必然需要维持关于它自身操作的大量数据，即为上述的“操作数据”。而对于上述企业的操作数据包括如下各项。

产品数据

帐目数据

病人数据

学生数据

计划数据

然而，需要指出的是，操作数据并不包括输入或输出数据。工作查询或任何纯粹中间信息。“输入数据”指的是从外部世界（典型是从卡片或者终端）输入到系统的信息；这样的信息可以使得操作数据的改变，但它并不是数据本身的一部分。类似的，“输出数据”指的是从系统发出的信息和报告（打印或在终端上其它显示）。同样，这样的报告所包含的信息是从操作数据推导出来的，但它仍然不是数据库的一部分。

作为操作数据概念的说明，让我们稍微详细考虑一个制造公司的情况。这个企业希望保留如下的信息：现有的工程；这些工程使用的另件；提供另件的供应者；存放另件的仓库；在这个工程中工作的雇员等等。这些就是记录到数据库中的数据的基本实体。见图 1·2。

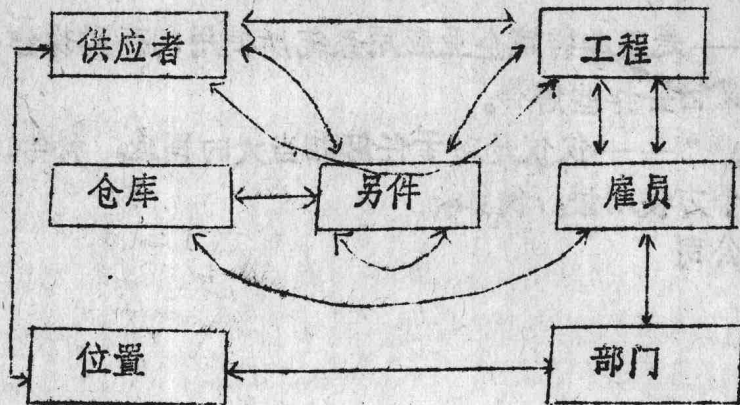


图 1·2 一个操作数据的范例

注意，链接这些基本实体到一起的称为联系或关系。这些在图 1·2 中，用箭头来表示。例如，供应者和另件之间有这种关系：每个供应者提供某些另件，反之，每个另件由某些供应者所提供。相似地，另件用于工程中，反之，工程使用另件等等。还要指出，所有关系都是双向的。更重要的是，被联系的这些实体的关系正好是操作数据的一部分。因此，这些关系必须表示在数据库中。至于这种表示的方法可以用指针，物理邻接，或者其它更有效的方法，但不管那种方法必须得作这一步。

图 1·2 还说明了另外一些问题。

1、在图中大量的关系虽然是联系两个实体型。但也有例外。在我们的例子中，有一个箭头联系三个实体（供应者—另件—工程）。代表这样的事实，某些供应者供应某些另件到某些工程中去。而一般

的並不是供应者—另件和另件—工程这两种关系的组合。

2、另外在我们的例子中，还有一个箭头仅仅包含实体型（例如另件）。这说明一些另件是另外一些另件的部件。

3、相同实体可以有若干种关系联系着。例如工程和雇员以两种关系相联系。一种是“工作”关系（雇员为工程工作），另一种关系是管理者关系（雇员是工程的管理者）。

1·2 为什么需要数据库？

目前大多数企业很流行的情况是—每个应用者都有它自己的私有文件，往往还有它自己的磁带和磁盘组，因而操作数据非常分散，几乎无法对它进行系统的控制。这就产生了许多问题。而企业将它自己的操作数据存贮到集成数据库中，这就提供企业对它的操作数据进行集中控制。这对企业来说是很宝贵的。这与第前一情况相比有许多优点。

- 可以减少存贮数据的冗余量。
- 在一定程度上避免存贮数据的不一致性。
- 共享存贮数据。
- 可以实施标准化。
- 可以应用保密限制。
- 维护数据的完整性。
- 可以平衡冲突的要求。

1·3 数据库系统结构

在结束这章之前，我们简单叙述一下数据库的结构（图1·3）。我们提出这个结构的主要目的是为以后的章节打下基础。

结构分成三层：内层，概念层和外层。概括说来，内层是最接近物理存贮的一层，而涉及数据实际存贮方式的一层。外层是最接近用户的一层，即涉及各个用户观察数据方式的一层。概念层是上述中间的“间接层”。如果说外部层涉及的是个别用户的视图，那么概念层

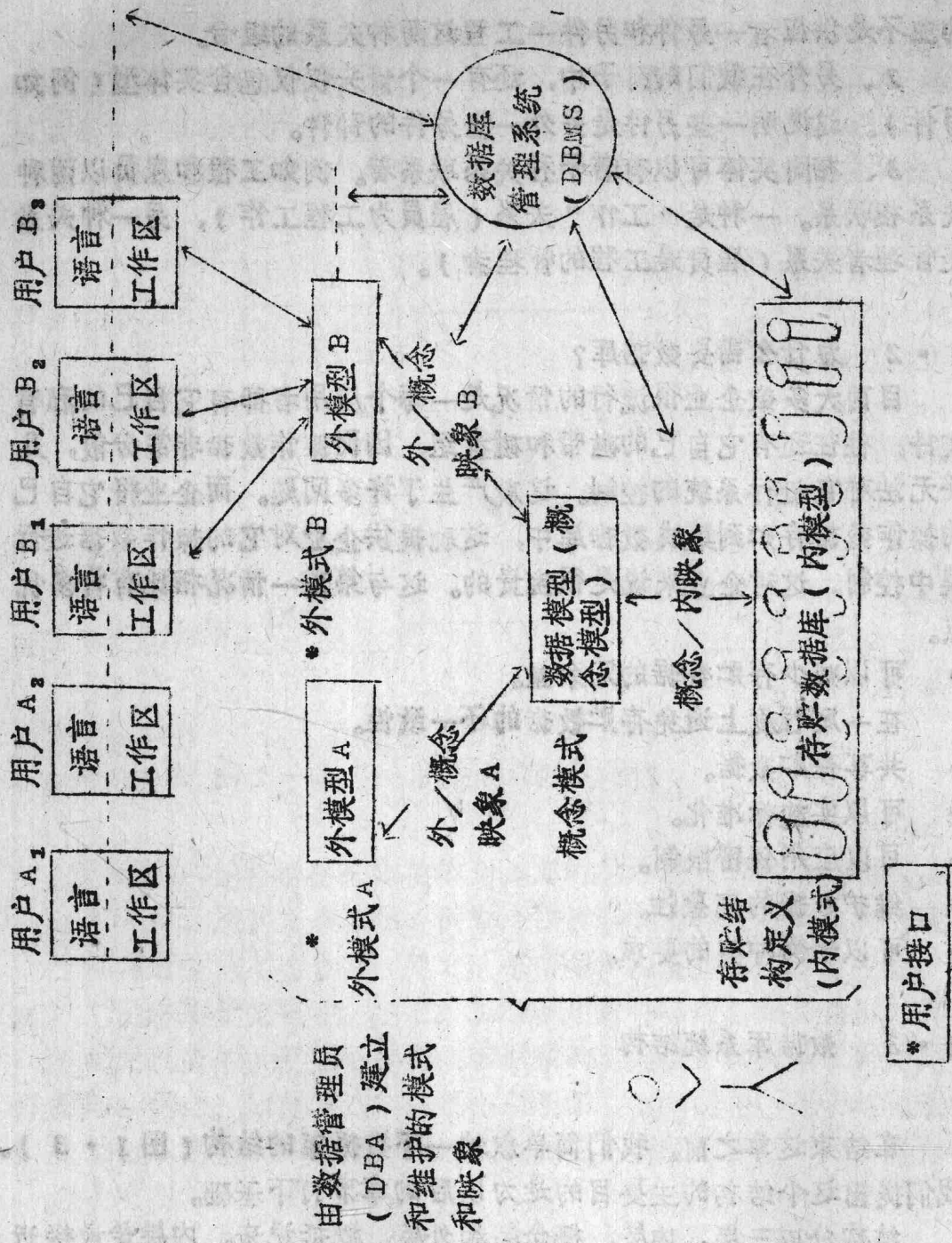


图 1 · 3 数据库系统的结构

可以认为确定了总体的用户视图。换句话说，有很多的“外视图”，每一个外视图都是由数据库某些部分的抽象描述所组成。而只有一个“概念视图”，类似的，概念视图是由数据库全体的抽象描述所组成。同样，只有一个“内部视图”，描述了整个数据库的实际存贮。

现在我们更详细的研究系统的各组成部分。

用户或者是应用程序员，或者是远程终端用户。对于每个用户提供一个工作空间，它的作用是对用户和数据库之间全部转换数据作为接收和或发送区。对于应用程序员，这个工作空间仅仅作为输入/输出区。对于终端用户来说，指定给它的工作存贮空间，或者显示设备组成。

外模型由多个外记录型（和存贮记录不同）的多个值组成。

外模式用来定义外模型，是由对外模型中的各种外记录型的描述组成。

概念模型是数据库的全部信息内容的描述。是由多个概念记录型的多个值组成。

概念模式用来定义概念模型的，包括对各个概念记录型的定义。

内模型是由多个内记录组成（存贮记录）。它与物理层仍然有一个距离。

内模式用来定义内模型的。它不仅定义各存贮记录型，而且确定存在什么索引，存贮记录按什么顺序等等。

图 1·3 中还有两级映象；

概念/内映象定义了数据模型和存贮数据库间的对应关系；它确定了概念记录和字节怎样映成存贮记录和字节。若改变数据的结构——即改变存贮结构的定义——则概念/内映象要相应的改变。要使得概念模式可以保持不变。换言之，这种改变的影响保持概念层之下，从而获得数据独立性。

外/概念映象定义指定外模型和数据层之间的对应关系。

从图 1·3 中可知还有三个问题需要讨论。

数据管理系统（DBMS）是处理对数据库所有存取的软件。还负责合法检验及有效性过程。

数据库管理员 (DBA) 是对数据库全局负责的人。

用户接口是系统的界面，在它之下各种事情用户是看不到的。

1.4 基本术语

(1) 实体—对之存贮信息的项目称为实体。

一个实体可以是一个可触及对象，如一个雇员，一个另件，或一个地方。也可以是一个不可触及的对象，如一个事件，一个工作名称，一个顾客帐目，或一个抽象概念。在数据处理中，我们常常涉及一组类似的实体，如雇员，并希望记录他们每一个相同特点。我们称这样的相似实体的集合为一个实体结合。

(2) 属性—属性是事物某一方面的特征，如姓名、年龄、性别、政治面貌等是人的几个属性。某人叫王一，22岁，男性，团员则分别为上述属性的值。而该属性必须与有关实体联系起来。属性所取值的范围称为属性域。如性别的属性域为(男，女)。域有理论上的域与实际上的域之分，而后者为前者的子集。如年龄这个属性的域，理论上可以是(0~200)，但某集团中实际上有一个比较小的范围。例大学生这个集团不会小于15岁，也不会大于32岁。由此可见属性是个变量，属性值是变量所取的值，而域是变量取值范围。

在数据处理中需要分清型与值的概念。通常用符号对：属性/属性值的形式来表示属性的型与值，如姓名/王一。年龄/22等等。

(3) 个体与总体

由若干个属性的属性值的集合所表征的对象(这些对象可以是物、事件或概念。)叫做个体。

例如由姓名/王一，政治成绩/77，语文成绩/82、数学成绩/92、物理成绩/91、化学成绩/98组成的集合表征了“王一考试成绩”这样一个个体。另外还可以组成“王一健康状况”这样一个个体。

由个体组成的集合称为总体。例如“王一考试成绩”“丁二考试成绩”……等等组成的集合为“考生考试成绩”总体。

对于个体也要分清型与值两个概念。因此几个属性的组合为描述个体的型,如姓名、年龄、身高、体重、血压、心率等为健康情况这种个体的型,而相应的属性值的集合则为描述某一个体的值,如(王一、18、1.70m、45Kg、70~110mm、74次/分)为“王一健康状况”这个个体的值。

总体中含有许多个体,为了索引一个个体,需要一个属性来标别个体。即当属性取某定值时,就可以确定某一个体,则此属性称为标别码。例如在“考生考试成绩”这种总体中,考生的姓名可以作为标别码,因为考生姓名一确定,那么该“考生考试成绩”这个个体随之而确定。若出现同名现象时,只能用号码作为标别码。

有时不是用一个属性作标别码,而是用若干属性组合作为标别码。例如火车班次可以用车次和日期的组合作为标别码。

有的个体有两个以上的标别码,如“考生成绩”中,若姓名不重复,而且又有报考号,那么姓名和报考号都可以作为标别码。在进行数据处理时,选定其中之一作为主标别码,简称为主码。

数据处理中除主码外也经常用不具有唯一性的属性作为码。所谓不具有唯一性的属性是指这样的属性,当它取一个定值时,个体有多个值与之对应。如在“考生健康状况”的总体中年龄等于18岁的不止王一一一个。称这种码为辅助码。

第二章 信息系统

2.1 信息系统

首先我们略述信息和数据这两个基本概念。这可以从两个不同的角度来叙述,可以说:

信息是物理状态反映的组合,数据是载荷信息的各种符号,即数据是表示信息的。

也可以说：

数据是从观察或测量中所得到的事实，而信息是数据有意义的表现，即信息是消化了的数据。

总之，信息用数据表示，对于数据进行综合推导得出新的数据。这些新的数据表示新的信息，可以作为决策的依据。

由此可见信息对人类活动，特别是管理工作是非常重要的，因此需要一个信息系统作为提供一个企业或组织所需的信息的工具。信息系统的工作有如下六个过程：即对于信息进行收集、组织、存贮、加工、抽取和传播。

一个企业或组织中有许多执行各种具体功能的子系统，这种系统通常称为应用系统。如会计系统、库存系统、生产系统等等。信息系统则与此不同，它不执行具体功能，而是为各种功能系统服务贯穿全局的系统。

信息系统可以是手工的，也可以是机械化的，通常写作 MDPS (Mechanized Data Processing System)，也可以是电子计算机化的。通常简写作 EDPS。只有计算机化才能提高信息系统的效能，这里我们只讲计算机化的信息系统。

一个计算机化的信息系统可以看成是一个人—机系统。在操作一级上，人将数据输入机器，由机器加工后输出数据，人再根据输出的数据所表示的信息在管理一级作出决策。

由于是人——机系统，所以我们对信息系统有两种观点，或者说可以从两个角度看待这个系统。一个是从机器的角度看待系统。我们称为机器观点。另一个是从人的角度看待系统。我们称为用户观点。前者是关于信息系统中数据处理方式问题。后者是关于信息系统中数据管理方式问题。数据库系统是研究后一问题的。但为了把两种观点联系起来，我们先对前一问题作一概述。

2.2 数据处理系统：

数据处理方式可分为三种系统：批处理系统，联机处理系统和分布处理系统。