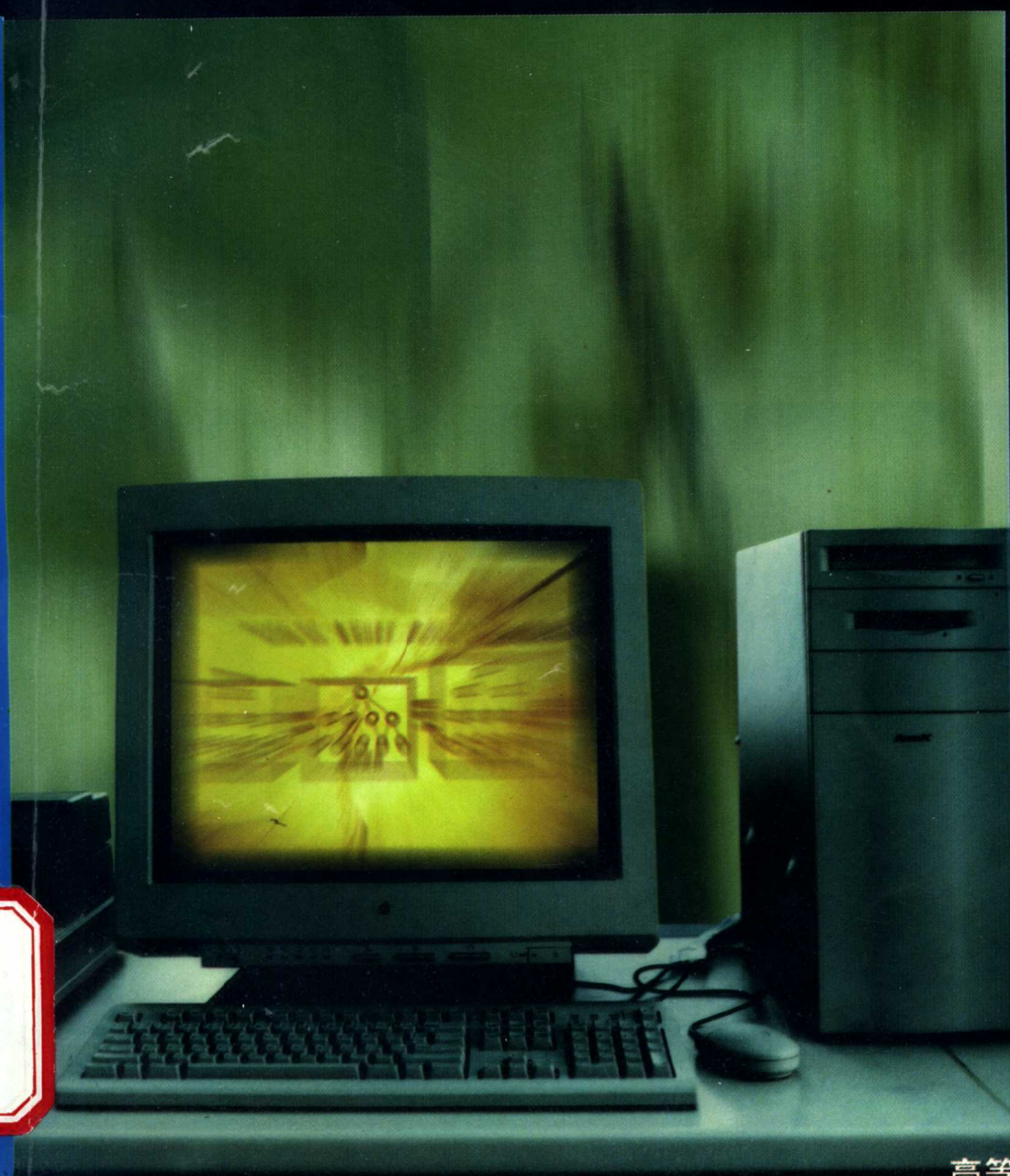


高等学校文科专业计算机基础课程教材

微型计算机 应用信息系统

主编
许卓群

编著
谢新洲
王益明
谢柏青



高等教育出版社

高等学校文科专业计算机基础课程教材

微型计算机应用信息系统

许卓群 主编

谢新洲 王益明 谢柏青 编著

高等教育出版社

内 容 提 要

本书在介绍了数据库系统等基本概念的基础上,详尽地讨论了 FoxBASE+ 数据库的使用方法及其在实际信息系统开发中的应用,同时介绍了 FoxPro 2.5 for Windows 的基本操作,书中最后还介绍了应用信息系统的设计与开发方法。

本书针对普通高校文科专业学生的特点而编写,突出实际操作与应用,内容安排循序渐进,语言叙述深入浅出,例子丰富,示图多样;可作为文科各专业教材,也可供各类培训班学员及广大计算机爱好者使用与参考。

图书在版编目(CIP)数据

微型计算机应用信息系统/许卓群主编;谢新洲等编著.
北京:高等教育出版社,1997
ISBN 7-04-005945-2

I. 微… I. ①许… ②谢… II. 微型计算机-数据库系
统-高等学校-文科(教育)-教材 N. TP311.13

中国版本图书馆 CIP 数据核字 (97) 第 08930 号

*

高等教育出版社出版

北京沙滩后街 55 号

邮政编码: 100009 传真: 64014048 电话: 64054588

新华书店总店北京发行所发行

化学工业出版社印刷厂印装

*

开本 787×1092 1/16 印张 15.5 字数 380 000

1997 年 7 月第 1 版

1997 年 7 月第 1 次印刷

印数 0001—1 092

定价 16.60 元

凡购买高等教育出版社的图书,如有缺页、倒页、脱页等
质量问题者,请与当地图书销售部门联系调换

版权所有,不得翻印

高等学校文科专业计算机基础课程教材出版说明

21 世纪是社会信息化的时代。信息技术的发展已使计算机及其网络深入到了人类生活中的各个方面。早在几十年前,计算机就已经走出纯科学计算领域,开始应用于人文科学中的各种数据处理。目前,计算机已成为人文科学研究的有效工具和手段,许多新的研究方法都是建立在计算机应用的基础之上。不仅如此,计算机还成为人文科学教学的有力工具和手段,尤其是多媒体技术的迅速发展,使计算机辅助教学变得越来越普通。计算机正改变着人们工作与生存的方式。正因为此,计算机知识与技能已成为 21 世纪人才的必备素质。

为了适应社会和时代对新型人才的需求,近几年来,全国高校各文科专业纷纷开设了计算机基础课程,并取得了良好的效果。为了推进高校文科专业计算机基础课程教学工作更上新台阶,国家教委高教司于 1995 年组织部分计算机专家和计算机基础课程教学工作者成立了高等学校文科专业计算机基础教学指导小组,由北京大学计算机系许卓群教授任组长。该指导小组在充分调查和认真研究的基础上,根据高等院校各类文科专业的特点和当前计算机基础教学的发展趋势,制订了《普通高等学校文科专业计算机基础课程教学大纲》。该大纲的制订目的是加强高等学校文科专业计算机基础教学工作,规范教学的基本内容,以进一步提高教学质量。

这套教材就是按照该大纲的规定内容,结合计算机应用的最新发展,由许卓群教授任主编,并组织有多年文科专业计算机基础教学经验的教师编写而成。根据大纲,全套教材共分两册,即《微型计算机应用基础》和《微型计算机应用信息系统》。这套教材的宗旨是针对高校广大文科专业的需要和特点,突出应用,强调实际操作。内容安排循序渐进,叙述方式深入浅出,以帮助读者更易更好地理解 and 掌握计算机的基础知识与基本技能。

由于时间仓促,加之作者水平有限,书中肯定存在许多不足,恳请社会各界和广大读者批评指正。

编著者

1997 年 4 月

前 言

《微型计算机应用信息系统》是国家教委高教司组织制订的《普通高等学校文科专业计算机基础课程教学大纲》所规定的第二门计算机基础课程的教材。本课程和第一门课程《微型计算机应用基础》既相互联系,又相对独立。在学习了《微型计算机应用基础》后,读者将很容易理解和掌握本课程的内容,但稍有一些微机基础的读者也可以直接学习本课程。

本书根据教学大纲的规定及微机数据库系统的最新发展,将内容分为四大部分。第一部分为基础部分,简要介绍数据库、数据库管理系统及关系数据库的基础知识。第二部分为数据库语言部分,是本书的重点,详细介绍和分析目前国内微机上比较流行且比较实用的数据库系统 FoxBASE+ 2.1 的常用命令及其使用方法,并结合 FoxBASE+ 2.1 的命令广泛讨论实际信息系统的开发、设计和建立。第三部分为第二部分的继续,介绍 FoxBASE+ 换代产品 FoxPro for Windows 版的基本操作,以使读者进一步学习 Windows 环境下数据库系统的操作方法。第四部分为提高部分,讨论应用信息系统开发和设计的方法,以提高读者在建立实际信息系统时的分析能力。

本书充分考虑了高校文科学生的特点,在内容安排上循序渐进,在叙述上深入浅出,并列举了丰富的例题。书中每章后面配有大量习题,以利读者复习、巩固和提高。

本书在编写和出版过程中,始终得到了高等学校文科专业计算机基础课程教学指导小组的关心和帮助,在此谨向诸位学者和同行表示诚挚的感谢!

由于时间仓促,作者水平有限,书中错误在所难免,还请广大同行和读者批评指正。

编 者

1997 年 4 月于北京

目 录

第一章 信息系统的认知模型	1	4.3 FoxBASE ⁺ 的启动与退出	34
1.1 三个世界	1	习题	34
1.1.1 观念世界	1	第五章 数据库文件的基本操作	36
1.1.2 数据世界	2	5.1 数据库文件的建立	36
1.2 概念模型	3	5.1.1 数据库文件的结构	36
1.2.1 概念间的关系	3	5.1.2 数据库文件结构的建立	37
1.2.2 信息系统的概念模型	4	5.1.3 数据库文件结构的显示与修改	39
1.3 数据模型	6	5.1.4 数据记录的录入	42
1.4 问题域	6	5.2 数据库文件的打开、关闭和显示	43
1.5 建立信息系统的基本过程	7	5.2.1 数据库文件的打开	44
习题	8	5.2.2 数据库文件的关闭	44
第二章 数据库与数据库管理系统	9	5.2.3 数据库文件的显示	45
2.1 数据管理技术的发展	9	5.3 记录的定位与插入	47
2.2 数据库的概念	12	5.3.1 记录的定位	47
2.3 数据库系统的体系结构	14	5.3.2 记录的插入	50
2.4 数据库管理系统	15	5.4 记录的修改与删除	51
习题	17	5.4.1 记录的修改	51
第三章 关系数据库的基本知识	18	5.4.2 记录的删除与恢复	56
3.1 基本数据模型	18	5.5 数据库文件的复制	59
3.1.1 层次模型	19	5.5.1 整个库文件的复制	59
3.1.2 网状模型	20	5.5.2 库文件结构的复制	60
3.1.3 关系模型	21	5.5.3 数据库文件的结构文件及其应用	61
3.2 关系的数学定义	22	5.6 常用的磁盘文件操作	64
3.3 关系数据库模型	24	5.6.1 磁盘文件目录的显示	64
3.4 关系代数	25	5.6.2 磁盘文件的复制、改名和删除	66
习题	28	5.6.3 外部命令的运行	67
第四章 FoxBASE⁺概述	29	习题	68
4.1 FoxBASE ⁺ 的特点与功能指标	29	第六章 数据库文件的排序、查询与统计	70
4.1.1 FoxBASE ⁺ 的特点	29	6.1 索引文件与记录排序	70
4.1.2 FoxBASE ⁺ 的运行环境	29	6.1.1 索引文件的建立	70
4.1.3 主要功能指标	29	6.1.2 索引文件的开闭与更新	72
4.2 FoxBASE ⁺ 的语言规则	30	6.1.3 记录的物理排序	75
4.2.1 命令	30	6.2 数据库文件的查询	77
4.2.2 量与表达式	31	6.2.1 数据库文件的顺序查询	77
4.2.3 文件	33	6.2.2 数据库文件的索引查询	82

6.3 数据库文件的统计	85	8.5.2 EXIT 和 LOOP	140
6.3.1 库文件中记录数量的统计	85	8.6 过程文件及其调用	141
6.3.2 数值型字段的求和与求平均	87	8.6.1 过程文件的建立	141
6.3.3 记录的汇总	88	8.6.2 过程文件的调用	143
习题	89	8.6.3 过程调用中的参数传递	145
第七章 多个数据库文件的操作	91	习题	147
7.1 内存工作区的选择与互访	91	第九章 屏幕与菜单设计	151
7.1.1 工作区的选择	91	9.1 数据的定位显示和输入	151
7.1.2 数据库文件间的互访	92	9.1.1 数据的定位显示	151
7.2 数据库文件间的关联与更新	95	9.1.2 数据的定位输入	153
7.2.1 数据库文件的关联	95	9.1.3 PICTURE 和 FUNCTION 子句	155
7.2.2 数据库文件间的更新	98	9.1.4 RANGE 和 VALID 子句	157
7.3 数据库文件的连接与成批追加	100	9.2 清屏、画框和屏幕颜色设置	158
7.3.1 数据库文件的连接	100	9.2.1 清屏	158
7.3.2 数据文件的成批追加	102	9.2.2 画框	159
习题	106	9.2.3 屏幕颜色的设置	160
第八章 程序设计基础	107	9.3 格式文件	161
8.1 程序文件	107	9.3.1 格式文件的建立	162
8.1.1 程序文件的建立与运行	107	9.3.2 格式文件的调用	163
8.1.2 程序文件中的状态设置命令	109	9.4 亮条式菜单	164
8.1.3 程序文件中常用的辅助命令 与运行控制命令	110	9.4.1 亮条式菜单的建立	164
8.2 内存变量与数组	111	9.4.2 亮条式菜单应用举例	164
8.2.1 内存变量	111	9.5 下拉式菜单与弹出式菜单	166
8.2.2 内存变量的存取	112	9.5.1 下拉式菜单	166
8.2.3 内存变量的定义与释放	117	9.5.2 弹出式菜单	169
8.2.4 数组	118	习题	170
8.3 函数	121	第十章 FoxPro 及其系统菜单	171
8.3.1 数学运算函数	121	10.1 FoxPro 的特点	171
8.3.2 字符函数	123	10.1.1 FoxPro 的运行环境	171
8.3.3 日期函数	125	10.1.2 FoxPro 的性能指标	172
8.3.4 转换函数	126	10.1.3 FoxPro 和 FoxBASE+ 的兼容性	173
8.3.5 测试函数	127	10.2 FoxPro 的文件与命令	174
8.3.6 其他函数	131	10.2.1 FoxPro 的文件	174
8.4 条件判断语句	133	10.2.2 命令与函数	174
8.4.1 简单条件判断语句	133	10.2.3 FoxPro 的启动与退出	175
8.4.2 多分支条件判断语句	135	10.3 FoxPro 2.5 的系统菜单	176
8.5 循环语句	137	10.3.1 File(文件)菜单项	176
8.5.1 循环语句的格式与使用方法	137	10.3.2 Edit(编辑)菜单项	177
		10.3.3 Database(数据库)菜单项	178

10.3.4 Record(记录)菜单项	178	第十二章 应用信息系统开发	203
10.3.5 Program(程序)菜单项	179	12.1 数据库应用系统设计的基本内容和要求	203
10.3.6 Run(运行)菜单项	179	12.2 数据库应用系统的设计过程	205
10.3.7 Text(文本)菜单项	179	12.3 用户需求分析	206
10.3.8 Windows(窗口)菜单项	180	12.4 E-R 模型与概念设计	208
10.3.9 Help(帮助)菜单项	180	12.4.1 建立概念模型的原因	208
习题	180	12.4.2 E-R 模型	209
第十一章 FoxPro 的数据库基本操作 ...	182	12.4.3 局部概念模式设计	212
11.1 数据库的建立与修改	182	12.4.4 全局概念模式集成	213
11.1.1 数据库的建立	182	12.5 逻辑设计与物理设计	214
11.1.2 数据库结构的显示与修改	183	12.5.1 逻辑设计	214
11.2 记录的录入与编辑	185	12.5.2 物理设计	216
11.2.1 记录的录入	185	12.6 数据库的实施与维护	217
11.2.2 记录的浏览与修改	187	习题	217
11.2.3 记录的删除与替换	189	附录	218
11.3 记录的排序与查询	191	附录一 FoxBASE ⁺ (V2.10)命令一览表	218
11.3.1 记录的排序	191	附录二 FoxBASE ⁺ (V2.10)函数一览表	227
11.3.2 记录的查询	194	附录三 FoxBASE ⁺ (V2.10)出错信息英汉对照及处理方法	232
11.4 记录的统计与汇总	195	参考文献	234
11.4.1 数值型字段的求和与求平均	195		
11.4.2 记录的统计与计算	196		
11.4.3 记录的汇总	198		
11.5 FoxPro 的多库操作	199		
习题	202		

第一章 信息系统的认知模型

人类活动的整个历史过程中,始终贯穿着对信息的收集、保存、利用、处理和传递。20 世纪 50 年代,计算机第一次被应用于信息处理,标志着现代信息系统的诞生。从那时起,计算机成为信息处理的关键设备,以计算机技术为核心的信息技术也随后得到了迅猛的发展。现在,人类社会已经进入了“信息社会”阶段,信息系统已经成为社会活动的一种必不可少的设施,其作用也越来越重要。

在现代社会里,信息在各种社会活动中居越来越重要的地位。信息量的急剧增加,信息的形式多样、结构复杂以及人们对信息要求的及时性,使得信息处理成为一个极其重要的问题。理论上,信息论已发展为信息科学,被视为现代科学技术的三大支柱之一;实践上,计算机所引起的重大技术革命为信息化社会的形成提供了强有力的现代化信息处理工具。

1.1 三个世界

开发和利用信息系统的目的,是要解决人们在社会实践中所面临的问题。人类社会活动是在现实世界中进行的,因此,信息系统与现实世界必定存在着某种联系。从一种简单的观点看,信息系统所处理的对象应当能够在某种抽象和精确的程度上,反映出与问题域相关的那一部分现实世界。在以计算机为核心的信息系统中,信息必须表示为计算机能够处理的数据,而数据本身并不是客观存在于现实世界中的,数据只是现实世界事物的一种抽象描述。为了理解数据与现实世界事物之间的关系,有必要在现实世界和数据之间再引入一个抽象的层次,即观念世界。

观念世界是现实世界的抽象。观念世界是由实体构成的,实体是对现实世界事物的抽象,现实世界中的每一个事物都与观念世界中的一个实体相对应,现实世界中事物之间的关系也映射到观念世界中实体间的关系上。人们在改造现实世界的实践活动中所产生的问题也是对现实世界的抽象理解,因此,问题域是观念世界的一部分。

数据是对实体的抽象描述,数据世界中的每一项数据,都对应于观念世界中的一个实体,数据间的关系也对应于实体间的关系。

信息系统的活动是在数据世界的范围内进行的。信息系统通过观念世界获得客观事物的抽象描述,进一步抽象为数据,对数据进行处理,将处理结果解释为问题域的解,最终通过人的活动来改变现实世界。

三个世界的观点,是数据库系统方法论的起始点。对于信息系统而言,它也同样是一种合理的立场。现代的信息系统是以数据库为核心的,在数据库方法论中引入三个世界的观点是为了便于阐明数据模型、模式设计等首要的数据库理论问题。

1.1.1 观念世界

在现实世界中,事物是千变万化的,没有两件事物是完全一样的,也就是说,事物是彼此能

够区分的,事物之间互相区分的依据是事物特征的差异。例如,“张三”是一个人,“李四”也是一个人,他们之间是能够互相区分的,因为他们的名字不同,姓名是他们的一个特征。但客观事物又是普遍联系的,事物之间存在着复杂的关系。例如,张三和张四是兄弟俩,他们之间则存在着兄弟关系。

概括地说,观念世界是现实世界的抽象映像,现实世界中每一个事物都可以在观念世界中找到一个相应的实体,现实世界事物间的每一种实际存在的关系也可以在观念世界中的对应实体之间找到一种与之相对应的实体关系。再进一步讲,现实世界与观念世界的一部分——实体域,也存在着——对应关系,只是由于观念世界中还存在着另一部分——概念域和问题域,这是人类思维的结果,在现实世界中找不到对应的客观存在物。

因此,观念世界是由实体域、概念域和问题域三部分构成的。划分观念世界的三个域是很有好处的。为了弄清这种划分的目的,下面进一步讨论三者之间的关系。

概念域由实体域抽象而来。概念域由概念构成,概念是对实体域中实体进行抽象得出的。我们知道,人类认识活动的成果,都是通过各种概念来加以概括和总结的。科学研究的成果,也会产生一系列新的概念,学科知识是通过相关概念所构成的概念体系表达、保存和交流的。一般认为,概念是事物本质属性的抽象和概括,但在这里,以三个世界的观点来看待概念时,概念是对实体的本质属性的概括和抽象。概念域中的一个概念与实体域中的一组实体是相对应的,这组实体都具有某些共同的本质属性。概念与实体间的关系非常类似于在数据库方法中经常提到的“型与值”的关系。例如,“人”是一个概念,而“张三”则是一个实体。“人”这个概念是从“张三”、“李四”等许多具有共同的本质属性的实体中抽象出来的,这些共同的属性包括“能使用工具”、“能使用语言”、“是哺乳类动物”,等等。

问题域由问题构成,问题是人们用概念表达出来的对信息系统的需求。例如某个档案管理系统的用户表示,“我希望通过人名查找档案材料”,从这个问题入手,可以分析出一个问题模型,这个模型包含一个“人”的概念,以及“人”的各方面属性概念,如姓名、住址、年龄、性别、职业等,这些属性概念都是预期与问题相关的,其中“姓名”已经被明确地指出。

在传统的数据库方法中,实体有型与值的区别。也就是说,实体类型和实体值都是实体。当实体模型映射到数据库模型中时,相应地就产生这样一种观念:数据类型和数据值都是数据。我们知道,型和值的抽象层次是不同的,把二者混为一谈,会引起误解。引入概念域,是为了把观念世界中不同的抽象层次区分开来。引入问题域,则是为进行需求分析提供概念基础。系统分析的工作主要是在概念域和问题域中进行的。

1.1.2 数据世界

数据世界由数据构成。数据是客观事物通过两级抽象后在计算机中呈现出的表现形式,也就是观念世界中实体通过一级抽象而在计算机中呈现出的表现形式。数据世界是所有信息系统(现存的和潜在的)的数据的总和,包含一切可能包含的数据,从观念世界的实体域到数据世界中的数据域存在一个——对应的关系。同观念世界一样,数据世界中也存在着不同的抽象层次,即类型域,类型域中的数据类型是对数据域中数据本质属性的抽象。类型域中的一个数据类型对应于数据域中一组数据,这组数据具有某些共同的本质属性。

数据世界包含了所有信息系统的数据库内容。而每一个具体的信息系统,都占有数据域中的特定部分,这些特定的数据抽象后对应于类型域中的特定部分。一个信息系统所占有的类型域

中的那一部分,就相当于传统的数据库环境中的数据库模式。

1.2 概念模型

现实世界中事物及事物间的关系,经过一级抽象后在实体域中完整地反映出来。实体域又经过一级抽象成为概念域。那么,概念域中是否如实反映了实体域中存在着的实体间关系呢?

1.2.1 概念间的关系

概念都有内涵和外延,概念的内涵是它在实体域中所对应的那一组实体共同具备的属性的集合。概念的外延则是它在实体域中所对应的那一组实体的集合。

概念的内涵有深浅,即它所概括的实体的属性有多少;概念的外延有宽窄,即它所对应的实体集合有大小。概念的内涵越浅(即它所概括的实体属性越少),其外延越宽(即它所对应的实体集合所包含的实体越多),反之,概念的内涵越深,则外延越窄。为了说明这种关系,我们来看一个例子,“图书馆”是“从事图书资料的搜集、整理、保存和以借阅方式提供利用的机构”。这个概念的内涵是:1)它是一个组织机构;2)它的处理对象是图书资料;3)它的任务是搜集、整理和保存图书资料;4)它的服务方式是借阅。这个概念的外延包括“北京图书馆”、“北京大学图书馆”等。现在为“图书馆”的内涵增加一个属性:它的服务对象是全体公民,就形成了另一个概念:“公共图书馆”,这个概念的外延是“图书馆”外延的子集合,例如“北京大学图书馆”就不在“公共图书馆”概念的外延之中。

概念同概念的关系,可以从外延和内涵这两个角度来考察。从外延的角度来考察时,可以用集合论的方法。设概念 A 的外延为 $E(A)$,概念 B 的外延是 $E(B)$,则:

(1) $E(A) \cap E(B) \neq \phi$,则称 A 与 B 为相容关系。

① 若 $E(A) \subset E(B)$,则称 A 、 B 为属种关系,其中 A 为下位概念, B 为上位概念;

② $E(A) = E(B)$,则称 A 与 B 为同一关系;

③ $E(A) \neq E(B) \wedge E(A) \cap E(B) \neq E(A) \wedge E(A) \cap E(B) \neq E(B)$,称 A 与 B 为交叉关系。

(2) $E(A) \cap E(B) = \phi$,称 A 和 B 为不相容关系。如果 A 和 B 是不相容关系,并且都是同一个概念 C 的下位概念,则 A 和 B 之间的关系称为并列关系。

当从概念的内涵角度来考察概念间关系时,首先要区分两类属性。一类是共性属性,是从一组实体抽象出一个概念的依据。在一个概念所对应的实体集合中的所有实体,其共性属性都是共同的。例如,“工人”这个概念有一个共性属性:工作场所是工厂,“工人”外延中的所有个体,其“工作场所”属性的值都是工厂。另一类属性是特性属性,是在概念的外延中对实体进行区分的依据。例如,“人”这个概念,“姓名”就是它的一个特性属性。当一组实体已经被抽象为某一概念时,概念本身已经隐含了所有这一组实体的共性属性,这个时候我们在谈论概念的属性时,就是指特性属性。这样用“内涵”指代共性属性,用“属性”来指代特性属性。

概念之间存在着一类关系,即全面与方面的关系,整体与部分的关系。例如“人体”和“人的心脏”是整体与部分的关系,“产品”和“制造材料”是全面与方面的关系,这一类关系统称为属性关系,即一个概念是另一个概念的某种属性。例如“姓名”、“年龄”、“职业”、“住址”等都是“人”的属性概念。

如前所述,概念是对实体的本质属性的抽象和概括。实体之间是普遍联系的,概念之间也是普遍联系的,实体间的联系也通过概念间的关系反映了出来。但到此为止,我们只讨论了实体间和概念间的静态联系,这些联系都是与实体和概念结构、属性相关的。客观事物除了在结构方面静态地互相联系外,还通过人类的活动而动态地联系。例如一个农民在耕种土地,他的这项活动就把土地、种子、农具和他自己互相关联了起来。客观事物也可以通过自己的活动将自己与其他客观事物动态地联系起来,例如,一块带正电荷的云块与另一带负电荷云块相碰撞产生闪电和雷声。客观事物的行为在实体域中反映为实体的行为,在概念域中又被抽象为“行为概念”。行为概念是联系行为的客体概念、主体概念、结果概念等的纽带,同时,行为概念可以作为客体概念、主体概念等的属性。

综上所述,概念分为两类,一类是实体概念,即实体域中实体的抽象;一类是行为概念,即实体行为的抽象。概念间的关系也可以分为两大类,一类是静态关系,包括:

- (1) 属性关系,即一个实体概念是另一个实体概念的属性;
- (2) 层次关系,即一个实体概念是另一个实体概念的上位概念;
- (3) 并列关系,即有两个实体概念都是同一实体概念的下位概念,并且这两个实体概念的外延不相交。

另一类是动态关系,由行为概念反映出来,其中行为概念与实体概念之间存在属性关系。

此外,实体与概念之间又存在一种“值与型”的关系,称作实例关系。例如,实体“张三”是概念“人”的一个实例,它们之间存在实例关系。

1.2.2 信息系统的概念模型

信息系统本身和信息系统环境都是处于现实世界中的,信息系统的处理对象和运行环境则是处于数据世界中的。这样看来,信息系统与观念世界并没有直接的关系。事实上,引进观念世界这个抽象层次对研究信息系统是很有必要的。为了说明这一点,先讨论图 1.1 所示的一个信息系统研究模型。

根据这个模型,在利用、开发和运行三个信息系统过程中,有两个过程涉及到了人的因素:信息系统的开发过程是由用户和系统开发人员共同完成的;利用过程的主体是用户。信息系统一旦涉及到人,就不可避免地与人认知活动发生关系,而人的认知活动根本离不开观念世界。

在信息系统的开发过程中,用户和开发人员必须对信息系统的任务达成共识。这种共识决不是轻易可以达到的,而是需要用户和开发人员互相交流,使开发人员对用户的信息需求和系统的应用背景深入理解。交流是要求有工具的。一旦用户与开发人员达到了某种程度的共同认识,就需要以某种形式记录和表达这种共识,作为进一步交流的基础,这种交流工具和认识表达工具就是概念模型。所谓概念模型,就是用户和系统开发人员对系统所涉及到的客观事物及其关系的理解,表现为一组相关的概念。概念模型是一个由一组概念通过 1.2.1 小节所述的各种概念间关系组成的概念体系,抽象地描述了信息系统对象的结构和行为两方面的特征和关系。

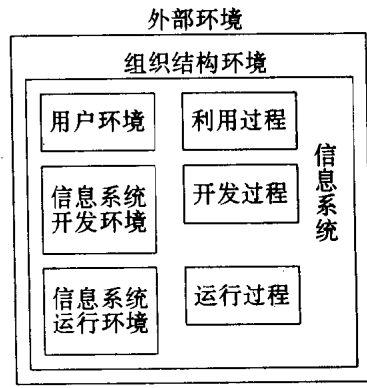


图1.1 一个信息系统研究的模型

信息系统的利用过程,也同样离不开观念世界的作用。信息的作用是通过改变人的认识或者意向而发挥出来的。要充分发挥信息系统的作用,必须使信息系统输出的数据符合当前用户的问题和所处的环境,那么,概念模型又将成为解释信息系统数据的最适用的工具。

概念模型不是凭空构造出来的,而是依赖于特定的建模方法。下面我们只讨论概念模型的要素和描述形式(详细介绍见 12.4 节)。

先来看 E-R 模型。E-R 模型,即实体-关系模型是一种广泛使用的概念模型。E-R 模型有三个构成要素:实体集合、属性和关系。E-R 模型是以 E-R 图描述的。在 E-R 图中,有三个基本符号,即长方形、椭圆形和菱形,分别用以描述实体集合、属性和联系。

(1) 用长方形表示实体集

每一个实体集合用一个长方形表示,方框内记上该实体集合名,如实体集合“教师”(teachers)、“学生”(students)和“课程”(courses)可以用图 1.2 中的三个方框表示。

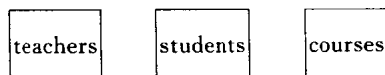


图 1.2 实体集表示法

(2) 用椭圆表示属性

每一属性用一个椭圆表示,属性名记在椭圆中。如学生的属性有学号(SNO)、姓名(name)、年龄(age)等,它们可以用图 1.3 中的三个椭圆表示。



图 1.3 属性表示法

(3) 用菱形表示联系

每一种联系用一个菱形表示,关系名记在菱形中,例如教师与学生之间的联系 TS、教师与课程的联系 TC 和学生与课程间的联系 SC 可用图 1.4 中的三个菱形表示。

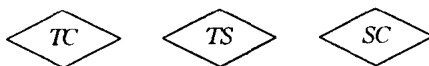


图 1.4 联系表示法

基本符号通过线段互相连接,就构成了 E-R 图。

通过上述基本符号和联接方法,可以画出 E-R 图表示信息系统的 E-R 模型。例如一个简单的大学数据库的 E-R 模型如图 1.5 所示。

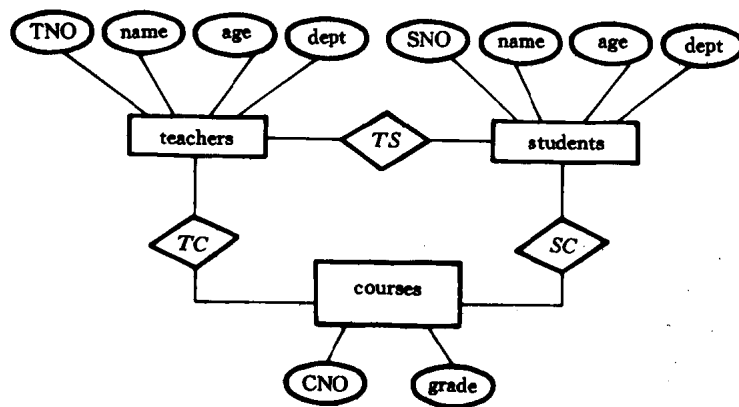


图 1.5 大学数据库 E-R 图

从以上的简单介绍中我们可以看出 E-R 模型没有超出概念域。但是 E-R 模型只反映了概念间的静态关系,而没有引入行为概念来反映概念间的动态关系。

1.3 数据模型

数据模型是数据库系统方法论要研究的首要问题。要指出的是,我们这里所说的数据模型是超脱于数据库系统方法的,我们的讨论是在信息系统背景中展开的。要理解数据模型是什么,有一个简单的办法,就是把它看成是概念模型在数据世界中的对应物。

数据世界包含了数据域和类型域,同概念与实体的关系一样,类型也是对数据的抽象,这种抽象关系同样地保持了概念对实体的抽象关系。更精确地说:

如果在观念世界中有实体 $e \in$ 实体域,概念 $c \in$ 概念域, $e \in E(c)$, $E(c)$ 表示 c 的外延,并且在数据域中有数据 d , d 是 e 在数据世界中的抽象描述,在类型域中有类型 t , t 是 c 在数据世界中的抽象描述,则 $d \in v(t)$, $v(t)$ 是类型 t 的值域;如果 $d \in (t)$,则也一定有 $e \in E(c)$ 。

同时,以下命题也是成立的:

- (1) 若概念 $c \in$ 概念域,则必有类型 $t \in$ 类型域,使 t 是 c 在数据世界中的抽象描述。
- (2) 若概念 c_1, c_2 之间存在关系 R , c_1, c_2 在类型域中的抽象描述 t_1, t_2 之间也必存在关系 R 。

这样看来,数据模型是概念模型在类型域中的镜像,或者说,是用数据世界中的语言来描述概念模型的产物。

一个信息系统在数据域中占有着自己的一部分,不妨记作 $D(I)$,其中 I 表示这个信息系统。 $D(I)$ 当中的数据在物理上如何存储和组织,并不是信息系统的用户所关心的问题,甚至也不是一部分开发人员(比如说,系统分析员)所关心的问题,他们对数据的理解和使用数据的基础,是数据模型。因此,数据模型是一个必不可少的抽象层次。与数据库系统相比较,信息系统的数据库模型相当于数据库模式,它同样也是保证数据独立性的一个必不可少的层次。

1.4 问题域

之所以要开发和使用信息系统,就是为了解决用户的问题,或者说是为了满足用户的信息需求。因此,开发好的信息系统的先决条件,就是要弄清用户的信息需求。80 年代的信息系统方法论是强调用户需求分析、强调与用户交流,强调用户参与的,“用户观点”是首要原则。

问题域就是由用户的信息需求构成的。问题域处于观念世界中,然而却并不是孤立作用于观念世界的,它也作用于数据世界。

事物的存在是客观的,这种客观性也反映在实体域中,也就是说,现实世界里的事物和观念世界里的实体都有客观性,不随人的意志而转移。但是,把实体抽象成概念就是一种人的认识活动,就会因人观察和理解实体的角度不同而不同。例如,“机动车”是一个概念,它的外延中每一件实体都有一个共同的属性,就是它们都是以机械能为动力的,这一属性已经包含在“机动车”的内涵之中。但是,“机动车”的特性属性,就随人的眼光不同而不同了。比方说,交通管理部门是把“机动车”看成一种运载工具,只关心它的“载重量”、“载客数目”、“自重”等属性,而商业部门却把它作为商品看待,关心其“价格”、“生产厂家”等属性。进一步讲,认识角度又取决于人

们要求解决的问题。因此,问题就好像是一个过滤器,它使人们只关心概念的某些属性而不是全部属性。问题的作用体现在两方面:

(1) 是建立概念模型的依据;

(2) 当一个信息系统不是为单一用户开发,而是为有不同问题的不同用户开发的时候,建立概念模型要兼顾所有用户的问题,而在利用信息系统时,每一个特定的用户只关心整个信息系统中与自己问题有关的那一部分,不同的问题决定了不同的“用户视图”。用户视图是处于数据世界中,作用于信息系统的数据模型。所有的用户视图在数据世界中形成了一个“用户视图域”。用户与信息系统打交道是通过用户视图的。

1.5 建立信息系统的基本过程

在 1.1 节中,我们讨论了看待信息系统的一种方法——即划分三个世界的方法。划分三个世界的依据,是信息系统开发、利用和运行过程中人的认识活动的抽象层次。下面我们用图 1.6 来表示三个世界思想的框架。

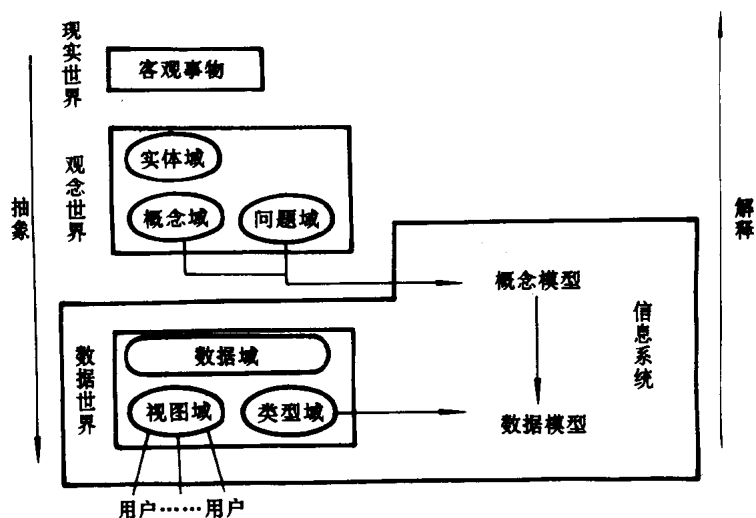


图 1.6 三个世界和信息系统

我们利用三个世界的思想,来简单讨论一下建立一个信息系统的基本过程。

有一个小型的图书馆,在这个图书馆中,我们按某种规则(比如说,按图书采购的时间先后)排架,给每本书标一个排架号,每个排架号都代表书架上的一个特定位置。现在,要建立一个信息系统,用这个信息系统,要求能够通过书名找到书的排架号。

(1) 现实世界:图书

(2) 观念世界

① 实体域:图书,由用排架号相互区分的实体构成。

② 概念域:图书、书名、作者、排架号……。

③ 问题:从书名获得排架号。

由于问题只关心书名和排架号,因此,在建立概念模型时,除了“书名”外,不考虑图书的其

他属性,例如“作者”等。在这个例子中,概念模型极其简单,用 E-R 图表示为图 1.7 所示。

(3) 数据世界

① 数据域:每本书有一条数据记录,记录其书名和排架号,例如:

(软件工程,1003)

(程序设计语言,1004)

② 数据模型

```
book{title: char-string;  
      shelf-no: integer;}
```

③ 视图:同上。

这个例子说明了在简单的情况下开发一个信息系统的基本过程:

- (1) 根据问题建立概念模型;
- (2) 将概念模型转换为一定的数据模型;
- (3) 根据数据模型组织数据。

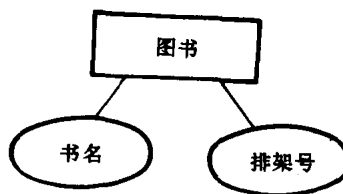


图 1.7 图书馆的简单 E-R 图

习 题

1. 三个世界的观点的主要内容是什么?
2. 三个世界的思想在信息系统开发中起什么作用?
3. 问题域在信息系统开发中的作用体现在哪些方面?
4. 概念间存在哪几类关系?
5. 建立一个信息系统包括哪些步骤? 举例说明。

第二章 数据库与数据库管理系统

2.1 数据管理技术的发展

自从计算机诞生以来,它就被广泛应用于各学科领域,进行科学计算和各类信息、数据的加工处理。在计算机发展最初的一段时间里,它主要应用于科学和工程领域的数值计算,但随着生产力的发展,信息在人类社会活动中起的作用越来越重要,人们经常需要收集、加工和处理大量的各类数据,如情报资料、银行帐目、人口统计、航空订票等。1950年,美国统计局利用计算机进行人口普查,1952年美国CBS电台利用计算机进行总统选票分析,从此,拉开了计算机进行数据处理的序幕。40多年来,随着计算机软、硬件技术的不断发展,数据管理技术也在不断地进步,其发展大体经过了人工管理、文件系统和数据库系统三个阶段。

1. 人工管理阶段(1946年~50年代中期)

在这一阶段,计算机首先被应用于科学计算。也正因为刚刚问世,所以计算机还很不完善。硬件方面,外存储器只有磁带、纸带、卡片。软件方面,没有操作系统和高级语言,没有对数据库管理的软件。用户只能用机器语言或汇编语言编程,且用户必须和计算机硬件打交道。程序员不仅要规定数据的逻辑结构,还要

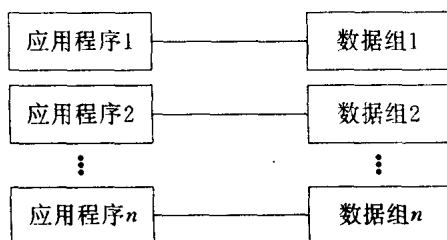


图2.1 手工管理阶段数据与程序间的关系

规定其存储结构、存取方法、输入输出的方式及其数据转换。因此,数据和其相应的应用程序是不可分的,是一种一一对应的关系(如图2.1所示),数据对应用程序来说不具有独立性。如果几个应用要用到同一数据,这些数据也要重复存储,数据的冗余度很大。在这一阶段中,数据的管理还是手工性质的,处理效率很低。

2. 文件系统阶段(50年代后期~60年代中期)

在这一阶段,计算机不仅应用于科学计算,同时也应用于数据处理。大量的数据需要长期地保存在外存储器中,并需要反复地进行查询、修改、插入、删除等处理,极大地促进了硬件和软件的发展。硬件方面,出现了大存储量的磁盘、磁鼓等直接存储设备;软件方面,出现了高级语言、编译系统、操作系统以及在操作系统支持下的数据管理软件——文件管理系统;处理方式也有了改进,既可以批量处理,又可以联机实时处理。

文件的应用,是计算机数据处理的重大发展。数据文件可以按名引用,应用程序通过文件管理系统与数据文件发生联系(如图2.2所示),数据库的物理结构与逻辑结构间有了简单变换。这样,在应用程序中,可以不必过多地考虑数据的物理存储细节,因此简化了程序员的数据管理工作。同时,一个应用程序可以和几个数据文件发生联系,增加了数据处理的灵活性。但是,数据仍是分散的,是面向应用的,用文件系统处理数据存在着以下几个问题:

(1) 数据冗余度大

由于数据是面向应用程序的,一个数据文件只为某一用户的特定用途服务,其他应用所需