

高等院校信息管理与信息系统专业系列教材

数据结构及应用算法教程 (修订版)

严蔚敏 陈文博 编著



清华大学出版社



高等院校信息管理与信息系统专业系列教材

数据结构及应用算法教程 (修订版)

严蔚敏 陈文博 编著

清华大学出版社
北京

内 容 简 介

本书从数据类型的角度,分别讨论了四大类型的数据结构的逻辑特性、存储表示及其应用。此外,还专辟一章,以若干实例阐述以抽象数据类型为中心的程序设计方法。书中每一章后都配有适量的习题,以供读者复习提高之用。第1~9章还专门设有“解题指导与示例”一节内容,不仅给出答案,对大部分题目提供了详尽的解答注释;其中的一些算法题还给出了多种解法。书中主要算法和最后一章的实例中的全部程序代码均收录在与本书配套的光盘之中。

本书内容丰富,概念阐述细致清楚,可作为高等院校计算机类专业和信息类相关专业“数据结构”或“软件基础”课程的本科教材。另外,对于准备参加计算机类研究生专业课统考的考生,本书也可作为应试的解题指导。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

数据结构及应用算法教程/严蔚敏,陈文博编著. —修订本. —北京:清华大学出版社, 2011.5

(高等院校信息管理与信息系统专业系列教材)

ISBN 978-7-302-24390-8

I. ①数… II. ①严… ②陈… III. ①数据结构—高等学校—教材 ②算法分析—高等学校—教材 IV. ①TP311.12

中国版本图书馆 CIP 数据核字(2010)第 259012 号

责任编辑:战晓雷

责任校对:李建庄

责任印制:何 芊

出版发行:清华大学出版社

<http://www.tup.com.cn>

社 总 机:010-62770175

投稿与读者服务:010-62795954, jsjic@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

地 址:北京清华大学学研大厦 A 座

邮 编:100084

邮 购:010-62786544

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185×260

印 张:25.5

字 数:633 千字

(附光盘一张)

版 次:2011 年 5 月第 1 版

印 次:2011 年 5 月第 1 次印刷

印 数:1~4000

定 价:45.00 元

产品编号:032042-01

出版说明

20 世纪三四十年代,一直摸索着前进的计算技术与刚走向成熟的电子技术结缘。这一结合,不仅孕育了新一代计算工具——电子计算机,还产生了当时谁也没有料到的巨大效应:电子计算机——这种当初为计算而开发出来的工具,很快就超出计算的范畴,成为“信息处理机”的代名词。

信息能促成管理系统的优化,促进组织创新,绩效不断上升;信息能提高计划与决策的科学性和及时性,是信息时代组织生存、发展、竞争制胜的有力武器;信息能革新企业内部的生产力要素结构,使资源转换系统的生产率大幅度提高,并同时以不断增加的柔性适应市场需求结构和消费结构的快速变化。

随着信息技术的发展与广泛应用,人类开始能够高效率地开发并利用信息,信息资源对人类社会的作用得以有效地发挥,并逐步超过材料和能源成为人类社会的重要支柱,信息化成为一个时代的口号。与此同时,信息资源开发与管理人才越来越广受社会青睐。

信息管理与信息系统专业是一个培养信息化人才的专业,是一个培养信息资源开发与管理方面的专门人才的专业。从知识结构上看,它处在管理学、信息科学与技术 and 有关专业领域的交叉点上。它对技术有极高的要求,又要求对组织有深刻的理解,对行为有合理的组织,反映了科学与人本融合的特点。这种交叉与融合正是信息管理与信息系统专业最重要的特征,是别的学科或专业难以取代和涵盖的。但是,它从 20 世纪 70 年代末开始创办到 90 年代初,尽管国内设有该专业的院校已经上升到 150 多所,但还没有形成很好反映自己特色的一个教材体系。1991 年全国 10 所院校的信息管理专业的负责人在太原召开第一次研讨会,异口同声地谈起创建一套符合专业需要的教材体系话题。以后,又经过 1993 年在大连、1995 年在武汉,又有更多的院校参加了这一研讨之中。这些研讨活动得到了国家教委有关部门的赞许和支持。通过研讨,大家在建设具有专业特点的教材体系、改变简单照搬其他专业教材上取得了共识。1996 年正式启动这个项目,协商由张基温教授担任主编,由魏晴宇教授、陈禹教授担任顾问。在清华大学出版社的大力支持下,从 1997 年起这套我国信息管理与信息系统专业的第一套系列教材陆续问世。迄今已经 10 年多,当初规划的七八本教材已经扩展到 30 多本,形成了一套品种多样、影响面广的系列教材,不仅为信息管理和信息系统专业建设作出了贡献,而且也被许多计算机专业所选用。这些都是编委会全体同仁和作者、广大使用本系列教材的师生以及出版社的编辑们辛勤劳动的结果。

同时,我们也欣喜地看到,10 年来,信息管理与信息系统专业也有了较大的发展,不仅其规模已经发展到 500 多个点,而且随着信息化的纵深推进,随着电子商务、电子政务和企业信息化的发展,专业的教学内容也与时俱进地深化和更新,从过去的围绕信息系统分析与设计,已经延伸到信息资源的开发与管理;专业的定位也逐步明晰,即为信息化建设与管理培养人才。同时,近年来围绕提高教学质量,许多学校开展了精品课程建设和教材建设。这些都标志这个专业正在走向成熟。

成熟的专业,需要优秀教材的支持。我们重新审视并修订这套教材。在这套教材问世10周年之际,我们再一次表示一个心愿:希望与全国的同行共勉,在教材和专业建设上齐心协力,作出更大贡献。我们将在原来的基础上,重新审视,不断补充,不断修改,不断完善。对于它的任何建设性意见,都是我们非常期盼的。为此,这一套教材将具有充分的开放性:每一本教材都是一个原型,每一位有志者对它的建设性意见都将会被采纳,并享有自己的知识产权,以使它们逐步成为精品。

《高等院校信息管理与信息系统专业系列教材》编委会

修订版前言

当今计算机应用已普及到各行各业,数据结构的知识内容越来越受到人们的重视。各类与计算机相关的考试,包括求职面试等,“数据结构”也都被列为首要的考试科目。学习数据结构的读者越来越多,市面上出版的数据结构书籍也不下几百种。然而,真正要学好数据结构并能融会贯通并非容易之事。本书自 2001 年 2 月出版以来,也陆续收到读者反映,感觉学习数据结构最大的难点是不会做题,特别是对算法设计题缺乏解题思路,希望本书能增加解题指导的内容。作者多年的教学实践也证明,阅读解题指导示例,揣摩其中算法的代码,也是提升学习数据结构效率的必经之道。

为此,作者在多年教学实践基础上,搜集整编,优化梳理,在本书的再版中增补了“解题指导与示例”的内容,共含习题百余道,为与教科书的内容融为一体,按题目所涉及的内容分别归入各章的正文之后。这些新增的题,不仅配有规范的答案,而且尽可能加上详尽的解答注释。对某些算法设计题还提供了多种解法,更有利于开阔解题的思路。

“解题指导与示例”共有 5 类题型,分别为选择题、填空题、解答题、算法阅读题和算法设计题,根据各章的内容需要而采用。算法阅读题和算法设计题偏重于难点集中的章节,如树、图、栈和队列的应用等。递归算法的跟读与设计等内容,读者也都能在“解题指导与示例”中找到相应细节内容的讨论。

掌握算法设计的要领和技巧是学习数据结构的基本要求,初学者面对辉煌而又森严的算法殿堂,往往会踌躇不前。究其原因,除了基本设计技法方面的欠缺外,很可能还存在设计思路贫乏的弊病。为此,对有关算法的解答,就不仅仅是给出冰冷的代码,而是呈献给读者带有启发性的思考过程,使读者从中看到设计灵感。其中有的题目给出了多个解答,以供读者有比对的想象空间:有的算法还提供了从简单到成熟的写作过程,以期体味发展的脉络;不少算法安排了扩展讨论的内容,有利于读者开阔思路。

跟读算法是深刻领会算法的必经之路,这次修订特意增加了这方面的内容,配有详尽的数据结构模型,逐步剖析算法的执行过程,多角度展露数据结构内容变化的“快照”。客观题型的题目也有一定的解题步骤和行文规则,在这次修订中,我们也试图通过众多的解答实例,启发读者养成规范的解题习惯。

上述尝试是否能适应读者的期望,还有待实践的检验。诚心为读者服务,虚心与同行切磋,不断改进和完善,是我们矢志不渝的初衷。

对本书初版书中的疏漏,有几位读者朋友向作者提出了很好的建议。作者借这次修订的机会已加以补充和修正,在此向热心的读者深表谢意!

严蔚敏 清华大学计算机科学与技术系
ywm-dcs@tsinghua.edu.cn

陈文博 北京工业大学计算机学院软件工程系
cscwmm@bjpu.edu.cn

2011 年春

初版前言

“数据结构”是计算机程序设计的重要理论基础,它所讨论的知识内容和提倡的技术方法,无论对进一步学习计算机领域的其他课程,还是对从事软件工程的开发,都有着不可替代的作用。“数据结构”是公认的计算机学科本科和大专的核心课程,也是计算机类专业“考研”和等级水平考试的必考科目,而且正逐渐发展成为众多理工专业的热门选修课。本书正是针对这一背景和社会需求编写的教材性读物,在内容选材方面,更多地考虑了普通高等院校计算机专业和信息类相关专业的读者的实际需要。

为了便于读者理解,书中对数据结构众多知识点的来龙去脉都做了详细的解释和说明,并配有大量的算法实例穿插其间。书的最后还专门辟出一章,用来讲解数据结构在解决实际问题中的应用示例,便于读者举一反三。考虑到计算机技术的发展和进步,在内容的编排方面尽量做到推陈出新,实例也力求新颖,以适应技术发展的潮流。

本书的第1章综述数据、数据结构和抽象数据类型等基本概念和算法;第2章、第4章至第7章从数据类型的角度,分别讨论线性表、栈和队列、串和数组、二叉树和树以及图和广义表等数据结构的逻辑特性、存储表示及其应用;第3章和第8章分别讨论排序和查找表的各种实现方法,其中除介绍各种实现方法外,并着重对算法的时间效率做了定性的分析,对算法的应用场合及适用范围进行了比较和介绍;第9章讨论常用的文件结构;第10章则以8个数据结构的综合应用为例,阐述以抽象数据类型为中心的程序设计方法。书的每一章都配有适量的习题,供读者复习提高之用。

本书在编排方面注意了数据结构本身的内在联系和从易到难的学习规律。例如,将排序安排在第3章,因为对读者来说,排序的内容比较容易理解,而且所涉及的数据结构主要是线性结构;又如对栈和队列的学习重点是它们的应用,因此在第4章里更多地列举了栈和队列的应用例子;在第5章中,结合C语言的串类型讲解串结构的知识内容,以使实际和理论在应用中和谐统一起来,等等。虽然广义表属线性结构,但由于它的“递归”特性,使得涉及广义表操作的算法和树更相似,因此将它放在图之后进行讨论,以降低理解难度。第10章的内容相当于“数据结构实习指导”,本意是为学生提供一个“综合利用数据结构知识编制小型软件”的规范示例。

全书采用了类C语言作为数据结构和操作算法的描述工具,它是C语言的一个精选子集,同时又采用了C++对C的非面向对象的增强功能。例如,动态分配和释放顺序存储结构的内存;利用引用参数传递函数运算的结果;使用默认参数以简化函数参数表的描述等。这些措施使数据类型的定义和数据结构相关操作算法的描述更加简明清晰,可读性更好,转变成C程序也极为方便。另一方面又埋下了伏笔,把类型定义和操作算法稍加技术处理,就很容易将其封装成类,并进一步转化成面向对象的程序模型。

从课程性质上讲,“数据结构”是一门专业技术基础课。它的教学要求应当是:学会从问题入手,分析研究计算机加工的数据结构的特性,以便为应用所涉及的数据选择适当的逻辑

结构、存储结构及其相应的操作算法,并初步掌握时间和空间分析技术。另一方面,本课程的学习过程也是进行复杂程序设计的训练过程,要求学生书写符合软件工程规范的文件,编写的程序代码应结构清晰、正确易读,能上机调试并排除错误。数据结构比高级程序设计语言课有着更高的要求,它重在培养学生的数据抽象能力。事实一再证明,任何具有创新成分的软件成果都离不开数据的抽象和在数据抽象基础上的算法描述。数据抽象能力是一种创造性的思维活动,是任何软件开发工具也无法取代的。本书将通过不同层次的应用示例培养学生逐步掌握数据抽象的能力,学会数据结构和数据类型的使用方法,为今后的学习和提高编程水平打下扎实的基础。

本书可作为计算机类专业的本科教材,也可以作为电子信息类相关专业的选修教材,教授可为 40 至 60 学时,另外应留有一定的时间供学生完成适量的上机作业。本书在编写方面以通俗易懂为其宗旨,特别注意了技术细节的交代,以便于自学,故也可作为从事计算机应用等工作的科技人员参考和查阅用书。在学习本书时应至少掌握一门高级程序设计的知识,如掌握的是 C 语言则最为理想;若能具有初步的离散数学和概率论的知识,对书中某些内容的理解会更容易。学习本书的同时还可把《数据结构》(C 语言版)作为配套参考用书。

与本书配套的光盘中含有书中所有算法和最后一章应用示例的全部源程序,可在 Visual C++ 5.0 或 6.0 的环境下编译执行,读者还可改变其中的输入数据,以观察程序对不同输入的执行结果。为了便于读者理解算法,在光盘中还为部分算法配有执行过程的示例演示。

应当感谢因特网,在本书的写作过程中,通过 E-mail 传送书稿使不在同一地方工作的两位作者可以做到随时交换意见并频繁修改书稿,以便使本书内容尽可能地做到令读者满意。但因时间仓促,仍有不尽如人意之处,请读者和同行赐教。

在写作本书的过程中,刘巍、钱大智、李莉、楼健、徐佳、金颖、林京秀、王福建等同学参加了第 10 章有关程序的调试工作,在此表示感谢。

严蔚敏

清华大学计算机科学与技术系
ywm-dcs@tsinghua.edu.cn

陈文博

北京工业大学计算机学院
cscwmm@bjpu.edu.cn

2000 年 7 月

目 录

第 1 章 绪论	1
1.1 数据结构讨论的范畴	1
1.2 与数据结构相关的概念	2
1.2.1 基本概念和术语	3
1.2.2 数据结构(data structures)	4
1.2.3 数据类型和抽象数据类型	5
1.3 算法及其描述和分析	6
1.3.1 算法	6
1.3.2 算法的描述	6
1.3.3 算法效率的衡量方法和准则	9
1.3.4 算法的存储空间需求	10
解题指导与示例	11
习题	16
第 2 章 线性表	18
2.1 线性表的类型定义	18
2.1.1 线性表的定义	18
2.1.2 线性表的基本操作	19
2.2 线性表的顺序表示和实现	22
2.2.1 顺序表——线性表的顺序存储表示	22
2.2.2 顺序表中基本操作的实现	23
2.2.3 顺序表其他算法举例	28
2.3 线性表的链式表示和实现	31
2.3.1 单链表和指针	31
2.3.2 单链表的基本操作	32
2.3.3 单链表的其他操作举例	36
2.3.4 循环链表	39
2.3.5 双向链表	40
2.4 有序表	42
2.5 顺序表和链表的综合比较	47
解题指导与示例	47
习题	57
第 3 章 排序	60
3.1 排序的基本概念	60

3.2	简单排序方法	63
3.2.1	插入排序	63
3.2.2	起泡排序	65
3.3	先进排序方法	67
3.3.1	快速排序	67
3.3.2	归并排序	69
3.3.3	堆排序	71
3.4	基数排序	72
3.5	各种排序方法的综合比较	76
	解题指导与示例	78
	习题	84
第4章	栈和队列	86
4.1	栈	86
4.1.1	栈的结构特点和操作	86
4.1.2	栈的表示和操作的实现	87
4.2	栈的应用举例	90
4.3	队列	98
4.3.1	队列的结构特点和操作	98
4.3.2	队列的表示和操作的实现	100
4.4	队列应用举例	104
	解题指导与示例	109
	习题	113
第5章	串和数组	115
5.1	串的定义和操作	115
5.2	串的实现和表示	118
5.2.1	定长顺序存储表示	118
5.2.2	堆分配存储表示	119
5.2.3	块链存储表示	120
5.3	正文模式匹配	121
5.4	正文编辑——串操作应用举例	123
5.5	数组	124
5.5.1	数组的定义和操作	124
5.5.2	数组的顺序表示和实现	125
5.5.3	数组的应用	127
5.6	矩阵的压缩存储	130
5.6.1	特殊形状矩阵的存储表示	130
5.6.2	随机稀疏矩阵的存储压缩	131
	解题指导与示例	137

习题	142
第 6 章 二叉树和树	144
6.1 二叉树	144
6.1.1 二叉树的定义和基本术语	144
6.1.2 二叉树的几个基本性质	147
6.1.3 二叉树的存储结构	148
6.2 二叉树遍历	150
6.2.1 问题的提出	150
6.2.2 遍历算法描述	152
6.2.3 二叉树遍历应用举例	153
6.2.4 线索二叉树	157
6.3 树和森林	159
6.3.1 树和森林的定义	159
6.3.2 树和森林的存储结构	161
6.3.3 树和森林的遍历	164
6.4 树的应用	168
6.4.1 堆排序的实现	168
6.4.2 二叉排序树	171
6.4.3 赫夫曼树及其应用	173
解题指导与示例	179
习题	198
第 7 章 图和广义表	201
7.1 图的定义和术语	201
7.2 图的存储结构	204
7.2.1 图的数组(邻接矩阵)存储表示	204
7.2.2 图的邻接表存储表示	205
7.3 图的遍历	207
7.3.1 深度优先搜索遍历图	208
7.3.2 广度优先搜索遍历图	210
7.4 连通网的最小生成树	215
7.5 单源最短路径	217
7.6 拓扑排序	221
7.7 关键路径	223
7.8 广义表	225
7.8.1 广义表的定义	225
7.8.2 广义表的存储结构	226
7.8.3 广义表的遍历	227
解题指导与示例	228

习题	246
第8章 查找表	248
8.1 静态查找表	249
8.1.1 顺序查找	250
8.1.2 折半查找	251
8.1.3 分块查找	254
8.2 动态查找表	255
8.2.1 二叉查找树	256
8.2.2 键树	260
8.3 哈希表及其查找	265
8.3.1 什么是哈希表	265
8.3.2 构造哈希函数的几种方法	267
8.3.3 处理冲突的方法和建表示例	268
8.3.4 哈希表的查找及其性能分析	269
8.3.5 哈希表的应用举例	272
解题指导与示例	274
习题	285
第9章 文件	287
9.1 基本概念	287
9.1.1 外存储器简介	287
9.1.2 有关文件的基本概念	288
9.2 顺序文件	289
9.2.1 存储在顺序存储器上的文件	289
9.2.2 存储在直接存储器上的文件	291
9.3 索引文件	291
9.3.1 B树	291
9.3.2 B ⁺ 树和索引顺序文件	293
9.4 哈希文件	295
9.4.1 文件组织方式	295
9.4.2 文件的操作	296
9.5 多关键码文件	296
9.5.1 倒排文件	297
9.5.2 索引链接文件	297
解题指导与示例	299
习题	300
第10章 数据结构程序设计示例	301
10.1 抽象数据类型	301
10.2 从问题到程序的求解过程	304

10.2.1	建立数据结构模型设计抽象数据类型	305
10.2.2	算法设计	306
10.2.3	实现抽象数据类型	307
10.2.4	编制程序代码并进行静态测试和动态调试	308
10.3	程序的规范说明	309
10.4	应用示例分析	311
10.4.1	含并、交和差运算的集合类型	312
10.4.2	最佳任务分配方案求解	322
10.4.3	排队问题的系统仿真	330
10.4.4	十进制四则运算计算器	339
10.4.5	自行车零部件库的库存模型	345
10.4.6	教务课程计划的辅助制定	353
10.4.7	一个小型全文检索模型	360
10.4.8	汽车牌照的快速查找	369
	实习题	377
	实习一 链表的维护与文件形式的保存	378
	实习二 用回溯法求解“稳定婚配”问题	378
	实习三 以队列实现的仿真技术预测理发馆的经营状况	379
	实习四 利用树形结构的搜索算法模拟因特网域名的查询	380
	实习五 管道铺设施工的最佳方案选择	381
	实习六 使用哈希表技术判别两个源程序的相似性	381
	附录 算法一览表	384
	参考文献	391

第1章 绪 论

1.1 数据结构讨论的范畴

“算法+数据结构=程序设计”原是瑞士的计算机学者 Niklaus Wirth 早在 1976 年出版的一本书的书名,很快就成了在计算机工作者之间流传的一句名言。斗转星移至今,尽管新的技术方法不断涌现,这句名言依然焕发着无限的生命力,它借助面向对象知识的普及,使数据结构技术更加完善和易于使用。由此,也说明了数据结构在计算机学科中的地位 and 不可替代的独特作用。

程序设计的实质是为计算机处理问题编制一组指令集。首先应该解决两个问题,即提出问题的数学模型和设计相应的算法。例如编制梁架结构应力计算的程序,所依据的算法是结构静力分析,而它的数学模型是一组线性代数方程组;又如预报人口增长率的数学模型是微分方程。在进行非数值计算问题的程序设计时,同样需要建立问题的数学模型和设计相应的算法。

例 1.1 图书馆的书目检索问题。

当你想借阅一本参考书,但不知道书库中是否有此书的时候;或者,当你想找某一方面的参考书而不知图书馆内有哪些这方面的书时,都需要到图书馆去查阅图书目录卡片。在图书馆内有各种名目的卡片:有按书名编排的;有按作者编排的;还有按分类编排的等。若要利用计算机实现自动检索,则计算机处理的对象便是这些目录卡片上的书目信息。列在一张卡片上的一本书的书目信息可由登录号、书名、作者名、分类号、出版单位和出版时间等若干项组成,每一本书都有唯一的一个登录号,但不同的书目之间可能有相同的书名,或者有相同的作者名,或者有相同的分类号。为了实现快速查询,需要为整个书库的书目信息设计恰当的数学模型和相应的查询算法。例如最简单的做法是建立一个按登录号顺序排列的书目文件和 3 个分别按书名、作者名和分类号顺序排列的索引表,如图 1.1 所示。

例 1.2 酒店管理系统中的客房分配问题。

在酒店的客房房态管理中,希望同类房中各间客房的出借率机会均等,以保证维持一个平均的磨损率。为此,分配客房采用的算法应该是“先退的房先被启用”。相应地,所有“空”的同类客房的管理模型应该是一个“队列”,即酒店前台每次接待客人入住时,从“队头”分配客房;当客人结账离开时,应将退掉的空客房排在“队尾”。由于“排队”是日常生活中经常需要的一种行为,因此队列也是这样一类活动的模拟程序中经常用到的一种数学模型。

例 1.3 铺设煤气管道问题。

假设要在某个城市的 n 个居民区之间铺设煤气管道,则在这 n 个居民区之间只要铺设 $n-1$ 条管道即可。假设任意两个居民区之间都可以架设管道,但由于地理环境的不同,所需经费也不同,则采用什么样的施工方案能使总投资尽可能少,这个问题即为“求图的最小生成树”的问题。其数学模型为如图 1.2 所示的“图”,图中“顶点”表示居民区,顶点之间的

连线及其上的数值表示可以架设的管道及所需经费。求解的算法为：在可能架设的 m 条管道中选取 $n-1$ 条，既能连通 $n-1$ 个居民区，又使总投资达到“最小”。

登录号	书名	作者	分类号	...
:	:	:	:	:
172832	高等数学	樊映川	S01	...
172833	理论力学	罗远祥	L01	...
172834	高等数学	华罗庚	S01	...
172835	线性代数	漆汝书	S02	...
:	:	:	:	:

书名	登录号	作者	登录号	类别	登录号
:	:	:	:	:	:
高等数学	172832,172834, ...	樊映川	172832,...	L	172833,...
理论力学	172833,...	华罗庚	172834,...	S	172832,172834,...
线性代数	172835,...	漆汝书	172835,...	:	:
:	:	:	:	:	:

图 1.1 书目文件和索引表示例

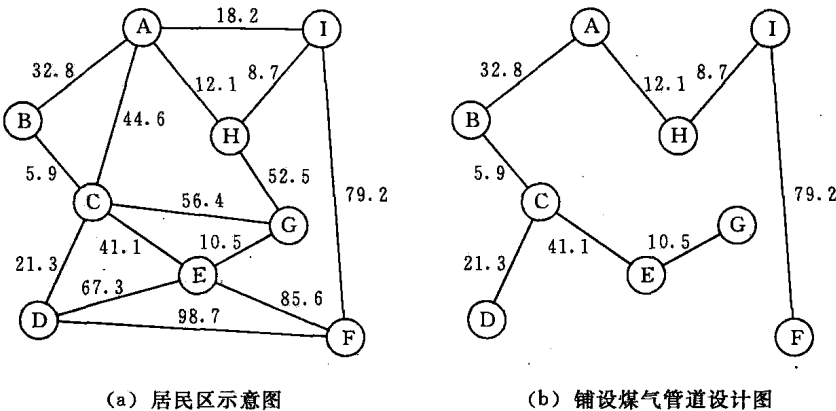


图 1.2 图及最小生成树示例

诸如此类的问题很多，在此不再一一列举。总的来说，这些问题的数学模型都不是用通常的数学分析的方法得到，无法用数学的公式或方程来描述，可称这些程序设计问题为“非数值计算”的程序设计问题。数据结构正是讨论这类程序设计问题所涉及的现实世界实体对象的描述、信息的组织方法及其相应操作的实现。

1.2 与数据结构相关的概念

在本节中，我们将对一些概念和术语赋予确定的含义，以便和读者取得共识，这些概念和术语将在本书的各章节中经常出现。

1.2.1 基本概念和术语

集合(set)是无法精确定义的基本概念。通常认为,集合就是若干具有共同可辨特征的事物的“聚合”,其中每个事物称为集合的元素或成员。例如,一个教室内的物件“黑板”、“讲台”、“课桌”和“椅子”可以构成一个“教室设施”的集合,其中每个物件称为该集合中的一个元素。类似地,教室内的人员“教师”和“学生”也可以构成一个“教室内人员”的集合。

集合的概念约定集合中不含有相同的元素,例如上述“教室内人员”的集合,若就人员的类别而言,则集合内只有两个元素“教师”和“学生”;若就人员的个体而言,则集合内的元素有“张老师”、“王国庆”、“刘晔”、“田华明”等。

集合有两种表示方法:一种是直接列出集合中的元素,元素之间以逗号分隔,如“教室设施集合” $C = \{\text{黑板, 讲台, 课桌, 椅子}\}$;另一种是规定集合中元素所共同具有的特征,如“1101 教室的学生集合” $S = \{p | p \text{ 是在 1101 教室听课的学生}\}$ 。在集合论中还规定集合内的元素无次序之分,例如, $\{\text{教师, 学生}\}$ 和 $\{\text{学生, 教师}\}$ 表示的是同一个集合。

数据(data)是对客观信息的一种描述,它是由能被计算机识别与处理的数值、字符等符号构成的集合。数据只是信息的一种特定的符号表示形式,是计算机程序进行“加工”的原料的总称。因此,对计算机科学而言,数据的含义极为广泛,并且随着技术的进步,数据所能描述的信息越来越丰富,如多媒体技术中涉及的视频和音频信号,经采集转换后都能形成计算机可操作的数据。

数据元素(data element)是数据的基本单位,在计算机程序中通常作为一个整体进行考虑和处理。数据元素可以是不可分割的“原子”,例如一个整数“6”或一个字符“A”;也可以由若干款项组成,例如上节所举的书目信息可由登录号、书名、作者名、分类号和出版时间等若干款项组成,其中每个款项称为一个“数据项(data item)”。如果一个数据项由若干款项组成(如出版时间),则称为组合项,否则称为原子项(如书名)。非原子的数据元素也是一种组合项。图 1.3 所示即为上述数据元素的内部结构。有时也称数据元素为记录、结点或顶点。

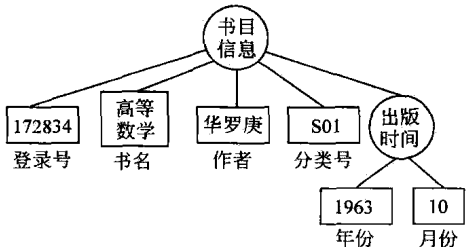


图 1.3 数据元素的内部结构

关键码(key)指的是数据元素中能起标识作用的数据项,例如,书目信息中的登录号和书名等。其中能起唯一标识作用的关键码称为“主关键码(简称主码)”,如登录号;反之称为“次关键码(简称次码)”,如书名、作者名等。通常一个数据元素只有一个主码,但可以有多个次码。

关系(relationship)指的是集合中元素之间的某种相关性。在集合中的元素之间可能存在一种或多种关系。例如,在教师和学生之间存在“教学”关系,在某两个学生之间存在“互为同桌”的关系等,在本书中将用如下的数学符号表示这种关系: $\{\langle \text{教师, 学生} \rangle\}$, $\{\langle \text{王国庆, 刘晔} \rangle, \langle \text{刘晔, 王国庆} \rangle\}$ 等。在集合论中, $\langle x, y \rangle$ 表示 x 相对于 y 的“顺序”关系。

1.2.2 数据结构(data structures)

若在特性相同的数据元素集合中的数据元素之间存在一种或多种特定的关系,则称该数据元素的集合为“数据结构”。换句话说,数据结构是带“结构”的数据元素的集合。在此,“结构”指的就是数据元素之间存在的关系。

数据结构包括(数据)逻辑结构和(数据)物理结构两个层次。数据的逻辑结构是对数据元素之间存在的逻辑关系的一种抽象描述,它可以用一个数据元素的集合和定义在此集合上的若干关系来表示;数据的物理结构则为其逻辑结构在计算机中的表示或实现,故又可称为存储结构。

例如,为了用计算机管理学生的课外活动小组,首先要为“小组”设计一个数据结构。假设每个小组只有 7 名成员(为了便于陈述,分别赋予他们代号为: A, B, C, D, E, F, G),其中 A 是组长,其余 6 人又分两个小小组,每组各有一个召集人。假设 B, C, D 同属一个小小组,召集人是 D,另一个小小组的召集人是 G。则表示这个“小组”的逻辑结构可以用一个数据元素的集合 S 和定义在该集合上的一个关系 R 来表示,其中

$$S = \{A, B, C, D, E, F, G\}$$

$$R = \{ \langle A, D \rangle, \langle A, G \rangle, \langle D, B \rangle, \langle D, C \rangle, \langle G, E \rangle, \langle G, F \rangle \}$$

如图 1.4 所示。

按照数据元素之间存在的逻辑关系的不同数学特性,通常有下列 4 类数据结构:

- (1) **线性结构**:指的是数据元素之间存在着“一对一”的线性关系的数据结构;
- (2) **树形结构**:指的是数据元素之间存在着“一对多”的树形关系的数据结构;
- (3) **图状或网状结构**:指的是数据元素之间存在着“多对多”的网络关系的数据结构;
- (4) **纯集合结构**:指的是在数据元素之间除了“同属一个集合”之外,别无其他关系。

上述结构的逻辑关系图分别如图 1.5(a), (b), (c) 和 (d) 所示。

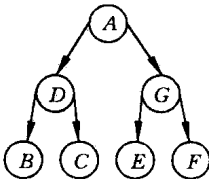


图 1.4 7 人小组的数据结构

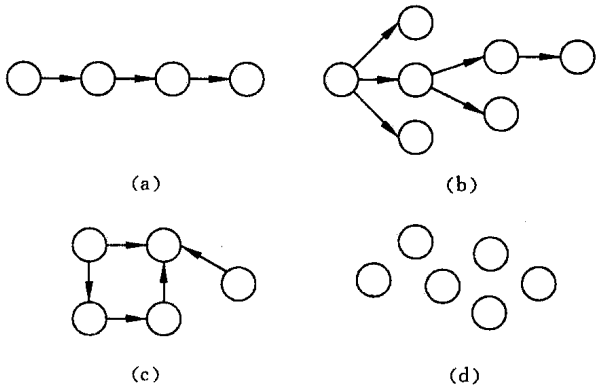


图 1.5 逻辑结构示意图

为了用计算机管理上述小组的人事关系,需要将这个数据结构存入计算机的存储器中,即考虑数据的存储结构。与逻辑结构相对应,存储结构包括数据元素的表示和关系的表示两个方面。

计算机的内存储器由顺序邻接的机器字组成。对 32 位机而言,一个机器字由 4 个字节