

现代信息管理与信息系统系列教材

数据结构与 程序设计

SHUJU JIEGOU YU

CHENGXU SHEJI

王裕明 李旭芳/主 编
李跃文 刘 升/副主编

清华大学出版社



上海市第四期教育高地（信息管理与信息系统）建设成果

数据结构与 程序设计

SHUJU JIEGOU YU
CHENGXU SHEJI

王裕明 李旭芳/主 编
李跃文 刘 升/副主编

清华大学出版社
北 京



内 容 简 介

本书比较系统地介绍了线性表、栈和队列、串、数组和广义表、树、图、查找、排序等内容,共分为9章,最后有一套实验题目。每章都配置了本章要点、本章小结和习题,便于读者理解、消化书中的内容。本书的最后,有一套实验题目,共有六个实验,每个实验又包含多道实验题目,每道实验题都从问题描述、基本要求、测试数据、实验提示、选作内容等环节给予解析,通过实验进一步提高读者运用数据结构知识解决实际问题的能力。全书融入了作者长期的教学经验,在编排上注意体现教学思路的完整性,力求做到深入浅出、通俗易懂,同时也考虑了自学者的学习方便,适于教学和自学。

本书的读者对象是学习数据结构课程的本、专科及高职类学生,也可作为信息类各专业考试、高校教师以及其他专业技术人员的参考用书。

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

数据结构与程序设计/王裕明,李旭芳主编.--北京:清华大学出版社,2010.10
(现代信息管理与信息系统系列教材)

ISBN 978-7-302-23478-4

I. ①数… II. ①王… ②李… III. ①数据结构 ②程序设计 IV. ①TP311

中国版本图书馆CIP数据核字(2010)第155270号

责任编辑:刘志彬

责任校对:宋玉莲

责任印制:杨 艳

出版发行:清华大学出版社

地 址:北京清华大学学研大厦A座

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

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

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

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185×230

印 张:19

字 数:367千字

版 次:2010年10月第1版

印 次:2010年10月第1次印刷

印 数:1~4000

定 价:35.00元

丛书编委会

主任 汪泓

副主任 吴忠 王裕明 史健勇

委员 唐幼纯 汪明艳 范君晖

刘升 朱君璇 李红艳

作为一种资源,信息是人类智慧的结晶和财富,是社会进步、经济与科技发展的源泉。信息同物质、能源一起,成为现代科学技术的三大支柱:物质为人类提供材料,能源向人类提供动力,而信息为人类奉献知识和智慧。

在人类发展史上,还没有哪种技术能够像信息技术这样对人类社会产生如此广泛而深远的影响。而现代信息技术,特别是采用电子技术来开发与利用信息是时代的需要,是世界性潮流、是人类社会发展的必然趋势,并正以空前的速度向前发展。

环顾当今世界,几乎每个国家都把信息技术视为促进经济增长、维护国家利益和实现社会可持续发展的最重要的手段,信息技术已成为衡量一个国家的综合国力和国家竞争实力的关键因素。

在国内,随着信息化建设的进一步深化,特别是电子商务和电子政务的兴起,社会各界对于信息管理人才的需求越来越多,要求也越来越高。这表明,“信息管理与信息系统”作为管理科学的一个重要分支,已经成为信息时代人才培养不可缺少的一个重要方面。

作为上海市优秀教学团队,上海工程技术大学信息管理与信息系统专业教师队伍在学科建设中,秉承面向国际、面向服务国家和地区经济建设的宗旨,坚持教学与研究相结合、理论与实践相结合,在近 20 年的专业建设中取得了一系列丰硕的教学与研究结果。

为了使读者进一步掌握信息管理理论和技术,也为了让研究成果更好地服务于社会,我们组织了一批长期从事信息管理与信息系统教学和研究的教师撰写了本系列教材。

本着培养“宽口径、厚基础、重应用、高素质”德才兼备、一专多能的信息管理类人才的原则,本系列教材以理论与实践相结合,注重系统性、基础性,突出应用性作为编写理念。因此,体现出以下三个方面的特点:

(1) 构建与人才培养目标相适应的教材体系。

教材建设的关键在于构建与人才培养目标相适应的知识内容体系。作为 21 世纪信息管理与信息系统专业的教材必须适应“以信息化带动工业化”的国家发展战略,以运筹学、系统工程等管理科学为研究方法,以计算机科学与技术为支持工具,构建培养读者掌握企业实施管理信息化所必需的知识体系。

本系列教材密切结合我国社会主义市场经济的发展对人才的需要,紧跟时代的发展,



不断补充和引进新的教学内容,增加信息技术方面的最新进展,紧紧围绕上述培养目标建设面向 21 世纪的信息管理与信息系统专业课程体系,并在此基础上进行教材体系的建设。

(2) 重视理论体系架构的完整性和鲜明性。

本系列教材可以使读者了解信息管理过程中,各个环节所应用的信息技术,了解信息管理系统的规划、开发和管理的内容,从而体会到信息管理的三大支撑学科——经济学、管理学和计算机科学在信息技术和信息系统所实现的信息管理中的内在联系和作用。

本系列教材由三个层次模块的教材组成,三个层次模块既有本身的核心知识内容,又紧密联系,形成了知识结构系统性的特点。其中:

- 信息管理的基础理论模块,如《信息资源管理》、《系统工程——方法应用》、《运筹学》等;
- 信息管理的技术模块,如《Java 语言编程实践教程》、《信息系统分析与设计》、《数据结构与程序设计》、《数据库系统原理及应用》等;
- 信息管理的应用模块,如《电子商务》、《管理信息系统理论与实践》等。

(3) 体现专业知识内容的应用性。

本系列教材强调理论联系实际,充分结合信息技术的实践和我国信息化的实际,注重理论的实际运用,全面提升“知识”与“能力”。在教材编写的过程中,教材案例编排的逻辑关系清晰,应用广泛,针对性强。本系列教材在注重理论与实践相结合的同时,也提高了实际应用的可操作性。

本套教材内容丰富,信息量大,章节结构符合教学需要和计算机用户的学习习惯。每章的开始,列出了“学习目标”和“本章重点”,便于读者提纲挈领地掌握各章知识点,每章的最后还附有“案例分析”和“习题”两部分内容,教师可以参照上机练习,实时指导学生进行上机操作,使学生及时巩固所学知识。

丛书编著做到了专业知识体系框架完整。在内容安排上,各教材内容广泛汲取了同类教材的精华,借鉴了本领域内众多专家和学者的观点和见解。

本套丛书在编写过程中参阅了大量的中外文参考书和文献资料,在此向国内外有关作者表示衷心的感谢。

由于编者水平和时间所限,如有错误和遗漏之处,敬请读者提出宝贵意见。

汪 泓

2010 年 4 月

于上海工程技术大学

数据结构课程是我国计算机教学中较早形成和完善的一门专业基础课程,也是计算机课程体系中的核心课程之一。近年来,数据结构课程不仅在计算机科学与技术类各专业开设,也陆续在信息类的许多本科专业和其他相关专业开设。该课程所介绍的各类数据的逻辑结构、存储方式及相关的算法既是程序设计,尤其是非数值性程序设计的基础,又是设计和实现系统软件及大型应用软件的重要基础。

本书以培养与提高读者的基本专业素质及综合应用能力为目标,注重体现教材的以下特色:

(1) 继承性。多年来,出现了许多很好的数据结构类书籍。本书在编写过程中认真总结和吸收与相关书籍的精华,并加以消化和创新。

(2) 适应性。对每一种抽象的数据类型先定义,然后再结合各种具体的存储结构加以实现,并将各种操作的应用实现辅以实例加以说明。与传统的实现方法相比,采用这种方案可使数据结构的抽象概念具体化、应用操作简单化,以适应我国现行普通高等院校的实际情况。

(3) 实用性。在人们学习的早期阶段,往往需要强化直接应用所研究的思想的能力,并且需要实际练习编写和运行程序,演示所学的每一个重要概念。在本书中,配合每一个重要概念都设置了相关的应用实例,这些实例在 VC++ 编辑环境下都能顺利运行,从而加强了学习的趣味性。

本书是在认真总结数据结构课程教学实践中的教学经验和体会的基础上,结合 21 世纪大学生的特点和学习需求编写的,以期通过简洁的语言,给学习数据结构课程的读者描绘出学习要点;通过实例分析,使读者对数据结构课程的学习能够完成从具体到抽象的一个提升过程。

全书共分为 9 章,包括绪论、线性表、栈和队列、串、数组和广义表、树、图、查找、排序,最后有一套实验题。各章都通过“本章要点”及“本章小结”对主要内容和知识点进行了概括总结,归纳出各章知识的主线条,有利于读者复习总结和理解掌握。每章配套有习题,题型包括选择题、填空题和算法设计题,选择题、填空题是针对各章基本内容给出的巩固题型,算法设计题侧重于各类基本数据结构的应用。本书的最后有一套实验题,共有 6 个



实验,每个实验又包括多道实验题,每道实验题都从问题描述、基本要求、测试数据、实验提示、选作内容等环节给予解析,通过实验进一步提高了读者运用数据结构知识解决实际问题的能力,使读者具备较复杂程序的初步设计能力。

本书的读者对象是学习数据结构课程的本科、专科及高职类学生,也可作为信息类专业考试、高校教师以及其他专业技术人员的参考用书。在使用过程中要注意,书中虽然给出了一些实例的具体程序,实验题中也给出了一定的实验提示,但是算法的设计具有不唯一性,切忌照搬照抄,要在学习、理解和领会的基础上改编程序,这样才能获得好的效果。

本书第1~3章、第6~7章由王裕明教授执笔编写,第4~5章、第8~9章以及实验题由李旭芳老师执笔编写。

本书在编写过程中,一直得到了国家教育部管理科学与工程教学指导委员会副主任委员、上海工程技术大学校长汪泓教授的关心和支持。初稿完成后,她又在百忙之中抽空审阅了全书。吴忠教授、刘升教授也对本书的编写提出过宝贵的修改意见,在此一并表示感谢!

本书在编写过程中,参阅了大量的中外文参考书和文献资料,在此向国内外有关作者表示衷心的感谢!

由于时间仓促和编者水平有限,书中难免存在疏漏或错误,敬请业内同人和广大读者不吝赐教。

编者
2010年8月

第 1 章 绪论	1
1.1 数据结构概述	1
1.2 算法及其复杂度	3
1.2.1 什么是算法	3
1.2.2 有关算法复杂度的概念	4
1.2.3 算法复杂度分析	4
本章小结	8
习题一	8
第 2 章 线性表	10
2.1 线性表的概念	10
2.2 顺序存储的线性表	11
2.2.1 顺序线性表的基本操作	11
2.2.2 顺序线性表应用举例	15
2.3 线性链表	17
2.3.1 向前链表	18
2.3.2 循环链表	25
2.3.3 双向循环链表	28
2.4 线性表应用举例	31
本章小结	42
习题二	42
第 3 章 栈和队列	46
3.1 栈	46
3.1.1 栈的定义和特性	46
3.1.2 栈的顺序存储方式	46



3.1.3 栈的链接存储方式	49
3.2 栈的应用举例	51
3.3 队列	56
3.3.1 队列的定义和特性	56
3.3.2 队列的顺序存储方式	56
3.3.3 队列的链接存储方式	61
3.4 队列应用举例	63
本章小结	68
习题三	68
第4章 串	72
4.1 串的基本概念和术语	72
4.1.1 串的定义	72
4.1.2 串的有关术语	72
4.2 串的存储结构	73
4.2.1 顺序存储结构	73
4.2.2 链式存储结构	74
4.3 串的基本操作	75
4.3.1 串的基本操作	75
4.3.2 顺序串的操作举例	76
4.4 串的模式匹配算法	78
4.4.1 简单的模式匹配算法	78
4.4.2 改进后的模式匹配算法	82
4.5 串操作应用实例	86
4.5.1 堆存储结构	86
4.5.2 文本编辑	86
本章小结	90
习题四	91
第5章 数组和广义表	92
5.1 数组的定义	92
5.2 数组的存储结构	93
5.3 特殊矩阵的压缩存储	96
5.3.1 对称矩阵	96



5.3.2	三角矩阵	97
5.3.3	带状矩阵	98
5.4	稀疏矩阵	99
5.4.1	稀疏矩阵的三元组表存储	99
5.4.2	稀疏矩阵的十字链表存储	105
5.5	数组应用实例	111
5.6	广义表	117
5.6.1	广义表的基本概念和性质	117
5.6.2	广义表的存储	118
5.6.3	广义表的基本操作	119
	本章小结	121
	习题五	121
第6章 树		123
6.1	树的基本概念和术语	123
6.1.1	树的定义	123
6.1.2	树的有关术语	124
6.2	二叉树	125
6.2.1	二叉树的定义	125
6.2.2	二叉树的性质	126
6.2.3	二叉树的存储结构	127
6.2.4	二叉树的建立	128
6.3	遍历二叉树	131
6.3.1	遍历二叉树的顺序	131
6.3.2	遍历顺序与二叉树的结构	131
6.3.3	遍历二叉树的递归算法	132
6.3.4	遍历二叉树的非递归算法	134
6.3.5	按层次遍历二叉树	138
6.3.6	遍历二叉树的应用	141
6.4	线索二叉树	144
6.4.1	线索二叉树的建立	145
6.4.2	单步遍历线索二叉树	148
6.5	哈夫曼树及应用	155
6.5.1	哈夫曼树的定义	155



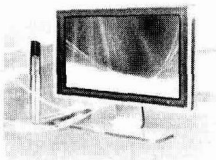
6.5.2	哈夫曼树的构造算法	155
6.5.3	哈夫曼树的应用	158
6.6	树和森林	161
6.6.1	树的存储结构	161
6.6.2	森林与二叉树的转换	164
	本章小结	165
	习题六	165
第7章 图		169
7.1	图的基本概念和术语	169
7.1.1	图的定义	169
7.1.2	图的基本概念和术语	170
7.1.3	带权图	171
7.2	图的存储结构	171
7.2.1	邻接矩阵	171
7.2.2	邻接表	172
7.3	图的遍历	172
7.3.1	深度优先搜索	172
7.3.2	广度优先搜索	173
7.3.3	图的遍历程序	174
7.4	图的生成树	182
7.4.1	图的生成树和最小生成树	182
7.4.2	Prim 算法	183
7.4.3	Kruskal 算法	186
7.5	最短路径问题	189
7.5.1	单源点最短路径问题	189
7.5.2	每对顶点间的最短路径问题	192
7.6	拓扑排序	195
	本章小结	199
	习题七	199
第8章 查找		204
8.1	基本概念	204



8.2	顺序表的查找	205
8.2.1	顺序查找	205
8.2.2	折半查找	207
8.2.3	分块查找	210
8.3	二叉排序树	211
8.3.1	二叉排序树的定义	211
8.3.2	二叉排序树的基本操作	212
8.4	哈希表	220
8.4.1	哈希表的概念	220
8.4.2	哈希函数的构造方法	221
8.4.3	解决冲突的方法	223
8.4.4	哈希表程序举例	225
	本章小结	227
	习题八	227
第9章 排序		229
9.1	排序的基本概念	229
9.2	插入排序	230
9.2.1	直接插入排序	230
9.2.2	折半插入排序	234
9.2.3	希尔排序	236
9.3	交换排序	240
9.3.1	冒泡排序	240
9.3.2	快速排序	242
9.4	选择排序	247
9.4.1	简单选择排序	247
9.4.2	树形选择排序	250
9.4.3	堆排序	251
9.5	归并排序	256
9.6	基数排序	259
9.6.1	多关键字的排序	259
9.6.2	链式基数排序	260
9.7	各种排序方法的综合比较	266



9.8 外部排序	268
9.8.1 外存信息的存储介质	268
9.8.2 多路平衡归并	269
9.8.3 平衡归并排序	276
本章小结	278
习题九	278
实验题	280
实验一 线性表	280
实验二 栈、队列的算法设计	282
实验三 串及其应用	284
实验四 树及其应用	285
实验五 图及其应用	286
实验六 查找和排序	287
参考文献	289



第 1 章

绪论

本章关键词

数据结构(data structure) 数据(data) 数据元素(data element) 算法(algorithm)
时间复杂度(time complexity) 空间复杂度(space complexity)

本章要点

本章主要介绍数据结构的研究内容和研究方法。

重点掌握：数据结构的有关定义；算法及其复杂度分析。

从第一台电子计算机问世的半个世纪以来,计算机的飞速发展已远远超出人们的预料。它的应用范围也在不断扩大,不仅被用于科学计算,而且更多地被用于数据处理、实时控制、辅助设计等。与此同时,计算机加工处理的对象也从纯粹的数值发展到字符、声音、图像等各种复杂的具有一定结构的数据。因此,要设计出一个好的程序,必须研究数据的特性和数据间存在的关系,这样就形成了数据结构这门学科。

1.1 数据结构概述

应用计算机来解决一个实际问题,一般需要经过下列步骤:首先从实际问题中抽象出一个数学模型,然后设计一个解此数学模型的算法,再编出程序,进行测试,最后运行得到最终结果(图 1-1)。对于许多物理现象或化学现象,经过分析后抽象出的数学模型可以是代数方程或微分方程,然而,更多的问题无法用这些数学方程来描述,而是属于非数值计算的问题。让我们来看几个例子。



图 1-1 计算机解决实际问题的步骤



[例 1.1] 利用计算机管理一个单位的人事档案。

表 1-1 职工人事档案信息表

工号	姓名	性别	出生年月	民族	职称	...	简历
0001	张林丽	女	1962.10	汉	工程师
0002	李大伟	男	1953.6	汉	高工
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1256	赵小明	男	1983.12	回	工人

一家单位的所有职工的人事情况可以用一张表格来表示(表 1-1),每个职工的信息占一行(也称一条记录),根据实际情况,经常需要到该表中查找职工人事信息,或是增加或是删除职工情况,即计算机操作的对象是这张表格。因此,计算机管理职工人事档案问题抽象出的数学模型就是包含所有职工人事信息的表以及对表进行的各种操作(如查找、增加、删除职工情况等)。

[例 1.2] 高校的组织结构。

对一个高校组织体系结构抽象出的数学模型就是包含所有部门及隶属关系的树型结构(图 1-2),以及对该树进行的各种操作(如查找、增加、删除部门情况等)。大学下属的部门有机械学院、电子学院、管理学院、化工学院等;各学院下属有各专业系,如管理学院下属有工商管理系、国际贸易系、信息管理系等。

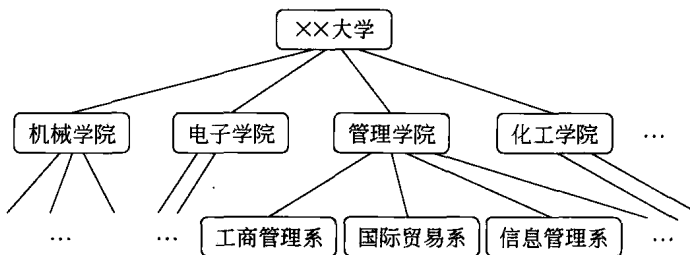


图 1-2 高校组织体系结构

[例 1.3] 城市间的交通网。

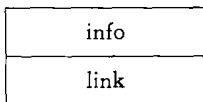
如果有 7 个城市分别用 A、B、C、D、E、F、G 表示,有些城市之间有公路,抽象出的数学模型就是包括这些城市及公路的网状图结构(图 1-3)和对图进行的各种操作(如遍历等)。

由前面例子可见,描述这些问题的数学模型不再是数学方程,而是表格、树、网状图等非数值性的数据结构。因此,数据结构就是一门研究非数值性程序设计中计算机操作的对象以及它们相互之间关系的学科。数据结构需要考虑的是数据在计算机存储器内的存



储方式和对这种存储方式的逻辑描述,以及针对这些数据的各种操作。

任何一种数据结构都是由基本元素组成的,这种基本元素在不同的场合,有时又称为节点、顶点、记录等。一个基本元素在存储器中总是占用一片连续的空间;基本元素可以包含多种不同类型的内容,如下面的基本元素包含两种不同的内容:



其中,info 称为数据项,用来存放该基本元素的数据值,根据具体情况,有些基本元素可能有多个数据项;link 称为指针项(也称链域),用来保存指向与其处于同一数据结构中的前一个或后一个基本元素的指针值,一个基本元素既可以有多个指针项,也可以没有指针项。

1.2 算法及其复杂度

1.2.1 什么是算法

算法是对特定问题求解步骤的描述,即能被机械地执行的动作(或称规则、指令)的有穷集合,一个动作的一次执行称为一步。算法具有以下特点:

- (1) 输入。一个算法可以有零个或多个输入量,这些输入量是算法所要求的初始信息。
- (2) 输出。一个算法有一个或多个输出信息,它们通常是与输入信息有特定联系的数量。
- (3) 确定性。算法的每一步骤必须有确定的意义,不能模棱两可。
- (4) 有穷性。一个算法对任一合法输入必须在执行有穷步后终止。
- (5) 有效性。算法中所有动作必须是相当基本的,也就是说,每一步至少在原理上能由人在有限时间内用笔和纸来完成。

算法与程序有密切的联系,但不是等同的概念。首先,程序也要求有输入、输出、确定性和有效性,但不要求有穷性,即可能为无限的循环。例如,操作系统的作业调度是一个永远不会停止的运行程序。其次,程序必须是用计算机可以接受并执行的计算机语言来编写的,而算法既可以用计算机语言描述,也可以用自然语言、流程图等来描述。

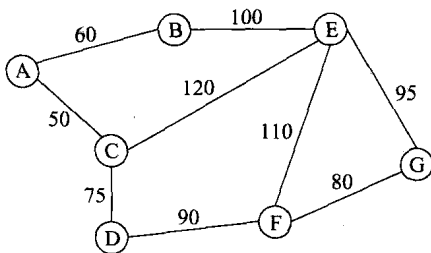


图 1-3 城市交通图