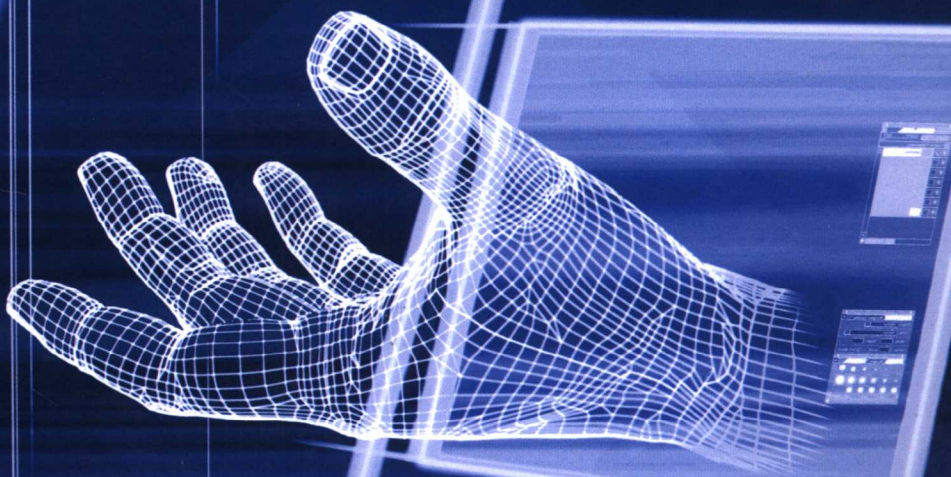


计算机科学与技术系列教材

数据结构

主 编 李春葆 桂 超 周云才 许中元
副主编 陈佛敏 陈建新 张万山 龚伏廷
李国屏



WUHAN UNIVERSITY PRESS

武汉大学出版社

计算机科学与技术系列教材

数据结构

主 编 李春葆 桂 超 周云才 许中元
副主编 陈佛敏 陈建新 张万山 龚伏廷
李国屏
参 编 王 邰 张 俊 刘红梅

武汉大学出版社

图书在版编目(CIP)数据

数据结构/李春葆,桂超,周云才,许中元主编. —武汉:武汉大学出版社,2006.8

(计算机科学与技术系列教材)

ISBN 7-307-05128-1

I. 数… II. ①李… ②桂… ③周… ④许… III. 数据结构—高等学校—教材 IV. TP311.12

中国版本图书馆 CIP 数据核字(2006)第 071726 号

责任编辑:黄金文 史新奎 责任校对:王 建 版式设计:支 笛

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.com.cn)

印刷:湖北恒泰印务有限公司

开本:787×980 1/16 印张:15.875 字数:325千字

版次:2006年8月第1版 2006年8月第1次印刷

ISBN 7-307-05128-1/TP·207 定价:21.00元

版权所有,不得翻印;凡购我社的图书,如有缺页、倒页、脱页等质量问题,请与当地图书销售部门联系调换。

计算机科学与技术系列教材

编 委 会

主 任:何炎祥,武汉大学计算机学院院长,教授

副 主 任:康立山,中国地质大学(武汉)计算机学院院长,教授

陆际光,中南民族大学计算机科学学院院长,教授

编 委:(以姓氏笔画为序)

王江晴,中南民族大学计算机科学学院副院长,教授

王春枝,湖北工业大学计算机学院副院长,教授

牛冀平,黄冈师范学院计算机系主任,副教授

石曙东,湖北师范学院计算机科学与技术系主任,教授

朱 英,桂林电子工业学院计算机系副教授

孙扬波,湖北中医学院信息技术系信息管理与信息系统教研室
主任

刘腾红,中南财经政法大学信息学院副院长,教授

陈少平,中南民族大学电信学院副院长,教授

杜友福,长江大学计算机科学学院院长,教授

陆 迟,江汉大学数学与计算机科学学院计算机系主任,副教授

闵华松,武汉科技大学计算机科学与技术学院副院长,副教授

陈佛敏,咸宁学院信息工程学院计算机系主任,副教授

陈建新,孝感学院计算机科学系主任,副教授

李禹生,武汉工业学院计算机与信息工程系副主任,教授

李晓林,武汉工程大学计算机科学与工程学院副院长,副教授

张涣国,武汉大学计算机学院教授

张唯佳,湖北省信息产业厅信息化推进处处长

余敦辉,湖北大学数学与计算机科学学院计算机系副主任

肖 微,湖北警官学院信息技术系副教授

钟 珞,武汉理工大学计算机科学与技术学院院长,教授

钟阿林,三峡大学电气信息学院计算机系主任

姜洪溪,襄樊学院电气信息工程系副主任,副教授

桂 超,湖北经济学院计算机与电子科学系副主任,副教授

黄求根,武汉科技学院计算机科学学院院长,教授

阎 菲,湖北汽车工业学院计算中心主任,副教授

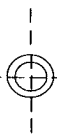
韩元杰,桂林电子工业学院计算机系教授

谢坤武,湖北民族学院信息工程学院计算机系主任,副教授

戴光明,中国地质大学(武汉)计算机学院副院长,教授

魏中海,华中农业大学理学院计算机系副教授

执行编委:黄金文,武汉大学出版社副编审



内 容 提 要

数据结构是计算机科学及相关专业的核心课程。本书是以高等学校数据结构课程教学大纲为基础编写的,内容丰富,讲述清楚,深入浅出,书中提供了大量的学习实例,每章均有练习题。本书适合作为普通高等院校计算机科学本、专科数据结构课程的教材,也适合计算机水平考试者使用。

前 言

数据结构是计算机科学及相关专业的核心课程。学习该课程需要有程序设计和离散数学的基础知识,同时,该课程又为后续课程如操作系统、编译原理、算法设计与分析、软件工程等的学习提供基础知识,所以在本科计算机专业中,数据结构课程起到承上启下的作用。

数据结构主要讲授数据组织、数组分析和数据处理的方法,培养学生用计算机求解实际问题的能力。本书由 11 章组成。第 1 章是绪论;第 2 章是线性表;第 3 章是栈和队列;第 4 章是串;第 5 章是数组和稀疏矩阵;第 6 章是递归和广义表;第 7 章是树和二叉树;第 8 章是图;第 9 章是查找;第 10 章是内排序;第 11 章是文件和外排序。每章都配有相应的习题。

本书由多位长期从事数据结构课程教学的老师编写。第 1 章由湖北经济学院计算机与电子科学系桂超老师编写;第 2 章由湖北大学数学与计算机科学学院张万山老师编写;第 3 章由武汉工程大学计算机科学与工程学院王邯老师编写;第 4 章由武汉工程大学计算机科学与工程学院张俊老师编写;第 5 章由长江大学计算机科学学院周云才老师编写;第 6、9 两章由武汉大学计算机学院李春葆老师和江汉大学计算机系许中元老师编写;第 7 章由华中农业大学计算机系龚伏廷老师编写;第 8 章由长江大学计算机科学学院刘红梅老师编写;第 10 章由咸宁学院计算机系陈佛敏老师编写;第 11 章由孝感学院计算机系陈建新老师和湖北师范学院计算机系李国屏老师编写。最后由李春葆、桂超、许中元、张万山、周云才、龚伏廷和陈佛敏等统稿并修订部分内容。

学习本书需要有 C 语言程序设计和 C++ 中的引用类型(&)的知识,不需要掌握面向对象的程序设计。书中的实例在 VC++6.0 中调试通过。

本书的出版得到武汉大学计算机学院、湖北经济学院计算机与电子科学系、湖北大学数学与计算机科学学院、武汉工程大学计算机科学与工程学院、长江大学计算机科学学院、华中农业大学计算机系、咸宁学院计算机系、江汉大学计算机系、孝感学院计算机系、湖北师范学院计算机系的院系领导及老师的大力支持,在此一并表示衷心的感谢。

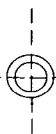
本书适合于作为计算机及相关专业本、专科生“数据结构”课程的教材,也适合

于计算机水平考试人员参考。

由于水平所限,书中可能存在错误和不足之处,敬请读者批评指正。

作 者

2006年5月



目 录

第 1 章 绪论	1
1.1 什么是数据结构	1
1.1.1 数据结构的定义	1
1.1.2 逻辑结构类型	4
1.1.3 存储结构类型	4
1.1.4 数据结构和数据类型	5
1.2 算法和算法分析	5
1.2.1 什么是算法	5
1.2.2 算法描述	7
1.2.3 算法设计的目标	7
1.2.4 算法效率分析	8
练习题 1	12
第 2 章 线性表	14
2.1 线性表的定义	14
2.2 线性表的顺序存储结构	15
2.2.1 顺序表	15
2.2.2 顺序表的应用举例	19
2.3 线性表的链式存储结构	20
2.3.1 单链表	20
2.3.2 循环链表	26
2.3.3 双向链表	27
2.3.4 链表的应用举例	29
2.4 线性表的顺序和链式存储结构的比较	33
2.5 线性表的应用	34
练习题 2	36
第 3 章 栈和队列	38
3.1 栈(Stack)	38



3.1.1	栈的定义	38
3.1.2	栈的顺序存储结构及其基本运算的实现	39
3.1.3	栈的链式存储结构及其基本运算的实现	41
3.1.4	栈的应用举例	44
3.2	队列 (Queue)	48
3.2.1	队列的定义	48
3.2.2	队列的顺序存储结构及其基本运算的实现	50
3.2.3	队列的链式存储结构及其基本运算的实现	53
3.2.4	队列的应用举例	56
	练习题 3	59
第 4 章	串	60
4.1	串的基本概念	60
4.1.1	串的基本概念	60
4.1.2	串的基本运算	61
4.2	串的存储结构	61
4.2.1	串的顺序存储结构——顺序串	61
4.2.2	串的链式存储结构——链式串	64
4.3	串的模式匹配	65
4.3.1	Brute-Force 算法	65
4.3.2	KMP 算法	67
	练习题 4	73
第 5 章	数组和稀疏矩阵	75
5.1	数组	75
5.1.1	数组的定义	75
5.1.2	数组存储的排列顺序	76
5.1.3	数组基本运算的实现	76
5.2	特殊矩阵的压缩存储	78
5.2.1	对称矩阵的压缩存储	78
5.2.2	对角矩阵的压缩存储	79
5.3	稀疏矩阵	80
5.3.1	稀疏矩阵的三元组表示	80
5.3.2	稀疏矩阵的十字链表表示	85
	练习题 5	86



第 6 章 递归和广义表	88
6.1 递归.....	88
6.1.1 什么是递归.....	88
6.1.2 如何设计递归算法.....	89
6.2 广义表的定义.....	90
6.3 广义表的存储表示.....	91
6.4 广义表的基本运算算法.....	92
练习题 6.....	96
第 7 章 树与二叉树	97
7.1 树的基本概念.....	97
7.1.1 树的定义.....	97
7.1.2 树的逻辑表示.....	98
7.1.3 树的基本术语.....	99
7.2 二叉树.....	101
7.2.1 二叉树的定义.....	101
7.2.2 二叉树的性质.....	102
7.3 二叉树的基本运算.....	105
7.3.1 二叉树的存储结构.....	105
7.3.2 二叉树的基本运算及其实现算法.....	107
7.4 遍历二叉树.....	113
7.4.1 二叉树的遍历算法.....	114
7.4.2 二叉树遍历算法的应用.....	117
7.5 线索二叉树.....	119
7.5.1 线索二叉树的概念.....	119
7.5.2 线索二叉树的逻辑表示图.....	120
7.5.3 二叉树的线索化算法.....	120
7.5.4 遍历线索化二叉树.....	122
7.5.5 线索二叉树的应用.....	123
7.6 哈夫曼树.....	124
7.6.1 哈夫曼树的基本概念.....	124
7.6.2 哈夫曼树的构造及其算法.....	126
7.6.3 哈夫曼编码.....	129
7.6.4 哈夫曼树的应用——判定树.....	131
7.7 树与森林.....	133
7.7.1 树的存储结构.....	133



7.7.2 树与二叉树之间的转换	136
7.7.3 森林与二叉树之间的转换	138
7.7.4 一般树的遍历	139
练习题 7	140
第 8 章 图	142
8.1 图的基本概念	142
8.1.1 图的定义	142
8.1.2 图的基本术语	142
8.2 图的存储结构	145
8.2.1 邻接矩阵存储方法	145
8.2.2 邻接表存储方法	146
8.2.3 十字链表存储方法	148
8.2.4 邻接多重表存储方法	149
8.3 图的遍历	150
8.3.1 图的遍历的概念	150
8.3.2 深度优先搜索遍历	150
8.3.3 广度优先搜索遍历	152
8.3.4 非连通图的遍历	153
8.4 生成树和最小生成树	153
8.4.1 生成树的概念	153
8.4.2 无向图的连通分量和生成树	154
8.4.3 有向图的强连通分量	154
8.4.4 普里姆算法	155
8.4.5 克鲁斯卡尔算法	157
8.5 最短路径	159
8.5.1 路径的概念	159
8.5.2 从一个顶点到其余各顶点的最短路径	160
8.5.3 每对顶点之间的最短路径	163
8.6 AOV 网与拓扑排序	166
8.6.1 什么是 AOV 网	166
8.6.2 拓扑排序	167
8.7 AOE 网与关键路径	169
8.7.1 什么是 AOE 网	169
8.7.2 查找关键路径的算法	170
练习题 8	174



第 9 章 查找	176
9.1 顺序查找	176
9.2 二分查找	177
9.3 分块查找	180
9.4 二叉排序树	182
9.4.1 二叉排序树的定义	182
9.4.2 二叉排序树的基本运算	183
9.5 二叉平衡树	188
9.6 哈希表查找	193
9.6.1 哈希表查找的基本概念	193
9.6.2 构造哈希函数的方法	193
9.6.3 哈希冲突解决方法	195
练习题 9	199
第 10 章 内排序	201
10.1 排序的基本概念	201
10.2 插入排序	201
10.2.1 直接插入排序	201
10.2.2 希尔排序	203
10.3 选择排序	205
10.3.1 直接选择排序	205
10.3.2 堆排序	207
10.4 交换排序	209
10.4.1 冒泡排序	210
10.4.2 快速排序	211
10.5 归并排序	214
10.6 基数排序	217
练习题 10	220
第 11 章 文件和外排序	221
11.1 文件的基本概念	221
11.1.1 什么是文件	221
11.1.2 文件的逻辑结构及操作	222
11.1.3 文件的存储结构	222
11.1.4 外排序概述	222



11.2	顺序文件	225
11.3	索引文件	226
11.3.1	ISAM 文件	226
11.3.2	VSAM 文件	227
11.4	哈希文件	229
11.5	多关键字文件	230
11.5.1	多重表文件	230
11.5.2	倒排文件	231
11.6	磁盘排序	233
11.6.1	磁盘排序过程	233
11.6.2	初始归并段的生成	233
11.6.3	多路平衡归并	235
11.6.4	最佳归并树	236
	练习题 11	237
	参考文献	240



第1章 绪论

数据结构是伴随着计算机软件技术的发展而分化出的一个领域,产生于20世纪60年代末期,在70年代得到极大的发展,形成了较为完善的研究内容和教学体系。1968年,美国的唐·欧·克努特教授在其撰写的《计算机程序设计技巧》第一卷《基本算法》中系统地阐述了数据结构的重要思想及基本方法,美国的一些大学也开设了有关内容的课程,主要是介绍和讲述图论、表、树等非线性结构,并扩展到网络、集合代数、格、关系等方面,形成了当今计算机科学领域的分支——“离散结构”。

IEEE(Institute of Electrical and Electronics Engineers,美国电气与电子工程师协会)和ACM(Association for Computing Machinery,美国计算机协会)于2001年发布的全美国计算机科学教学计划CC2001把算法与数据结构列入计算机及信息技术相关学科专业的本科必修基础课程,并以此科目作为评估学生对计算机科学掌握、了解的重要指标。最新的CC2004对此也作出了重要的陈述。

学习数据结构只要具有高级语言程序设计的基础,并具备一定的数学知识,就可以了。数据结构是操作系统、编译原理、软件工程、数据库原理的前导课程。此外,它也是算法设计与分析、计算机图形学的前导课程。

操作系统中用到的队列、存储管理表、目录树等,数据库原理中用到的线性表、多链表及索引树等,人工智能中用到的广义表、集合、搜索树及各种有向图等,均是数据结构的基本内容。在我国,数据结构已经成为理工科非计算机专业必修的信息技术基础课程之一,世界上许多科技人员对学习、研究数据结构都非常重视,将其作为计算机基础技术加以掌握。

1.1 什么是数据结构

1.1.1 数据结构的定义

1946年,第一台电子计算机在美国诞生,它的主要用途是帮助人们进行数值计算。随着计算机的不断发展,计算机深入社会的各个层面,它的使用功能发生了极大变化,从单纯的数值计算,发展演化为非数值计算,诸如处理各种事务型数据——数据处理、自动控制复杂机器运行——实时控制、辅助人们设计新的工业产品——辅助设计、智能化识别与推理——人工智能、网络通信数据共享——网络通信、多媒体声



音图形技术的处理——多媒体技术,等等。数值计算的基本数学模型是由数学方程、公式或计算方法理论来确定的,基本对象是数值,而非数值计算问题大量存在于实际工作中,它们的数学模型相对比较复杂,基本对象除了数值以外,更多的是字符、表格、记录、语音、图像等类型。要处理这些非数值型信息,就必须有效地将这些信息按照一定的特点进行组织和描述,形成特有的数据结构,并设计正确可行的算法,然后提交计算机处理,得到问题的求解。因此,数据结构这门课程主要是研究非数值计算问题中信息的具体对象的组织描述、相互关系、运算等。以下通过三个实例说明非数值计算问题的基本特点。

【例 1.1】 一个学生成绩表。

在学生成绩管理系统中,学生的成绩表如表 1.1 所示,存放的是某班学生的成绩,共有学号、学期、姓名、性别、英语、数学、语文、总分等 8 个数据项,其中每一行内容称为一个记录,描述的是某位学生的各项数据。对这些数据,系统设计了录入、查询、修改、删除、统计、打印等运算。整个表格是一个数据结构,每个记录的前面只有 1 个或 0 个记录,后面只有 1 个或 0 个记录,各个记录顺序排列,属于“一个对一个”的简单线性关系,这种类型的数据结构称为线性的数据结构。

表 1.1

第 1 学期学生成绩表

学号	学期	姓名	性别	英语	数学	语文	总分
95003	1	张三	男	80	85	90	255
95005	1	李四	男	80	90	70	240
...
...

【例 1.2】 一个大学的部门分布图。

图 1.1 描述的是“湖北经济学院”机构设置的一个部分,在“湖北经济学院”的下面分为“行政部门”和“教学部门”,在“行政部门”下面又分为“教务处”、“学工处”等,“学工处”下又分为“招生办”、“毕业办”、“学生科”等。“教学部门”分为“会计系”、“旅游系”、“计算机系”等。图上的每一个部门的前面最多只有一个部门,后面有多个或 0 个部门,这种数据结构具有层次分明的特点,排列就像一棵倒置过来的树,“湖北经济学院”是树的根,“行政部门”、“教学部门”是第一级的分支……我们称这种数据结构为“一个对多个”的树形结构。

【例 1.3】 点的到达问题。

图 1.2 是一个简单的有向图形,从点 1 可以通过路径 1→2→3 或 1→4→3 或 1→2→4→3 等方式到达点 3,如果知道每一个→代表的长度,则可计算出从 1 到 3 的最短路径。其实,可以计算出任意两个点之间的最短路径。这种类型的数据结构与例

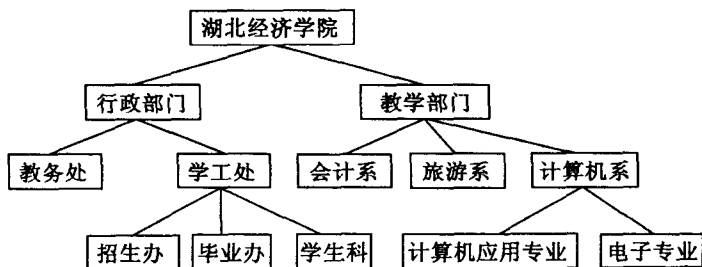


图 1.1 大学机构设置图

1.1、例 1.2 有明显的不同。复杂情况下,每一个点前面可能有多个点,它的后面也可能有多个点,属于“多对多”的非线性图形数据结构。

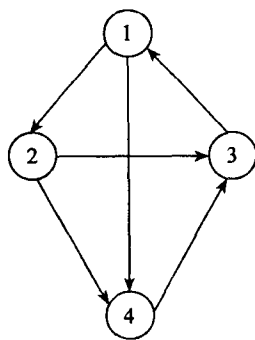


图 1.2 具有方向性的点的到达

这三类具体的问题,其特点与数值计算完全不同,必须根据问题的特征,引入新的描述方法和分析方法加以解决。

为了便于叙述,我们介绍以下概念。

数据(Data) 描述客观事物的数字、字符、图像、语音等可以用符号表示,能输入到计算机内并被计算机处理的信息的总称。

数据元素(Data Element) 是数据的基本单位,在计算机中通常被作为一个整体来进行考虑和处理,也称为结点或记录。例如,在 C 语言中,我们用结构变量来处理学生成绩表的问题,每一位学生的学号、姓名、性别、英语、语文、数学等各项信息可以用一个具体的结构变量表示,这个代表学生各项信息的结构变量就是一个数据元素,或者说是一个记录,其中的学号、姓名、性别、英语、语文、数学各个项我们称做数据项。

类型(Type) 一组值的集合。例如,在 C 语言中,int 代表一个整数类型,通常由