

21 世纪 高 职 高 专 规 划 教 材

计 算 机 专 业 基 础 系 列

21

数 据 结 构

(第 2 版)

邓 文 华 主 编

清 华 大 学 出 版 社



21世纪高职高专规划教材

计算机专业基础系列

数据结构 (第2版)

邓文华 主编

21

责任编辑：李红英
封面设计：李红英

出版发行：清华大学出版社
社址：北京清华大学学研大厦A座
邮编：100084
电话：(010) 62770175
http://www.tup.com.cn

印 次：2007年12月第2版
印 数：1-8000
定 价：30.00元
开 本：185×230
印 张：14.25
字 数：358千字

清华大学出版社
北京

内 容 简 介

本书对常用的数据结构做了系统的介绍,力求概念清晰,注重实际应用。全书共分9章,依次介绍了数据结构的基本概念、线性表、栈和队列、串和数组、树和二叉树、图、查找、排序以及综合应用实例。全书用C语言作为算法描述语言,并且每一章后面均列举了典型应用实例,并配有电子课件,以供教师教学使用。

本书配套出版了《数据结构实验与实训教程》,便于教学实验、课程设计及复习应考。本书主要面向高职高专院校计算机类专业的学生,也可以作为大学非计算机专业学生的选修课教材和计算机应用技术人员的自学参考书。

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 13501256678 13801310933

图书在版编目(CIP)数据

数据结构/邓文华主编.—2版.—北京:清华大学出版社,2007.12

21世纪高职高专规划教材.计算机专业基础系列

ISBN 978-7-302-16329-9

I. 数… II. 邓… III. 数据结构—高等学校:技术学校—教材 IV. TP311.12

中国版本图书馆CIP数据核字(2007)第162392号

责任编辑:束传政

责任校对:袁芳

责任印制:李红英

出版发行:清华大学出版社

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

c-service@tup.tsinghua.edu.cn

社总机:010-62770175

投稿咨询:010-62772015

地 址:北京清华大学学研大厦A座

邮 编:100084

邮购热线:010-62786544

客户服务:010-62776969

印刷者:北京四季青印刷厂

装订者:三河市溧源装订厂

经 销:全国新华书店

开 本:185×230 印 张:14.25 字 数:289千字

版 次:2007年12月第2版 印 次:2007年12月第1次印刷

印 数:1~6000

定 价:20.00元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:
010-62770177 转 3103 产品编号:024758-01

出版说明

高职高专教育是我国高等教育的重要组成部分,担负着为国家培养并输送生产、建设、管理、服务第一线高素质技术应用型人才的重任。

进入 21 世纪后,高职高专教育的改革和发展呈现出前所未有的发展势头,学生规模已占我国高等教育的半壁江山,成为我国高等教育的一支重要的生力军;办学理念上,“以就业为导向”成为高等职业教育改革与发展的主旋律。近两年来,教育部召开了三次产学研交流会,并启动四个专业的“国家技能型紧缺人才培养项目”,同时成立了 35 所示范性软件职业技术学院,进行两年制教学改革试点。这些举措都表明国家正在推动高职高专教育进行深层次的重大改革,向培养生产、服务第一线真正需要的应用型人才的方向发展。

为了顺应当前我国高职高专教育的发展形势,配合高职高专院校的教学改革和教材建设,进一步提高我国高职高专教育教材质量,在教育部的指导下,清华大学出版社组织出版了“21 世纪高职高专规划教材”。

为推动规划教材的建设,清华大学出版社组织并成立了“高职高专教育教材编审委员会”,旨在对清华版的全国性高职高专教材及教材选题进行评审,并向清华大学出版社推荐各院校办学特色鲜明、内容质量优秀的教材选题。教材选题由个人或各院校推荐,经编审委员会认真评审,最后由清华大学出版社出版。编审委员会的成员皆来源于教改成效大、办学特色鲜明、师资实力强的高职高专院校、普通高校以及著名企业,教材的编写者和审定者都是从事高职高专教育第一线的骨干教师和专家。

编审委员会根据教育部最新文件和政策,规划教材体系,比如部分专业的两年制教材;“以就业为导向”,以“专业技能体系”为主,突出人才培养的实践性、应用性的原则,重新组织系列课程的教材结构,整合课程体系;按照教育部制定的“高职高专教育基础课程教学基本要求”,教材的基础理论以“必要、够用”为度,突出基础理论的应用和实践技能的培养。

本套规划教材的编写原则如下:

- (1) 根据岗位群设置教材系列,并成立系列教材编审委员会;
- (2) 由编审委员会规划教材、评审教材;
- (3) 重点课程进行立体化建设,突出案例式教学体系,加强实训教材的出版,完善教学服务体系;
- (4) 教材编写者由具有丰富教学经验和多年实践经历的教师共同组成,建立“双师型”编者体系。

本套规划教材涵盖了公共基础课、计算机、电子信息、机械、经济管理以及服务等大类的主要课程,包括专业基础课和专业主干课。目前已经规划的教材系列名称如下:

• 公共基础课

公共基础课系列

• 计算机类

计算机基础教育系列

计算机专业基础系列

计算机应用系列

网络专业系列

软件专业系列

电子商务专业系列

• 电子信息类

电子信息基础系列

微电子技术系列

通信技术系列

电气、自动化、应用电子技术系列

• 机械类

机械基础系列

机械设计与制造专业系列

数控技术系列

模具设计与制造系列

• 经济管理类

经济管理基础系列

市场营销系列

财务会计系列

企业管理系列

物流管理系列

财政金融系列

国际商务系列

• 服务类

艺术设计系列

本套规划教材的系列名称根据学科基础和岗位群方向设置,为各高职高专院校提供“自助餐”形式的教材。各院校在选择课程需要的教材时,专业课程可以根据岗位群选择系列;专业基础课程可以根据学科方向选择各类的基础课系列。例如,数控技术方向的专业课程可以在“数控技术系列”选择;数控技术专业需要的基础课程,属于计算机类课程的可以在“计算机基础教育系列”和“计算机应用系列”选择,属于机械类课程的可以在“机械基础系列”选择,属于电子信息类课程的可以在“电子信息基础系列”选择。依此类推。

为方便教师授课和学生学习,清华大学出版社正在建设本套教材的教学服务体系。本套教材先期选择重点课程和专业主干课程,进行立体化教材建设:加强多媒体教学课件或电子教案、素材库、学习盘、学习指导书等形式的制作和出版,开发网络课程。学校在选用教材时,可通过邮件或电话与我们联系获取相关服务,并通过与各院校的密切交流,使其日臻完善。

高职高专教育正处于新一轮改革时期,从专业设置、课程体系建设到教材编写,依然是新课题。希望各高职高专院校在教学实践中积极提出意见和建议,并向我们推荐优秀选题。反馈意见请发送到 E-mail: gzgz@tup.tsinghua.edu.cn。清华大学出版社将对已出版的教材不断地修订、完善,提高教材质量,完善教材服务体系,为我国的高职高专教育出版优秀的高质量的教材。

高职高专教育教材编审委员会

前言(第2版)

数据结构(第2版)

本教材第1版自2004年出版发行以来,受到了广大师生、读者的热烈欢迎,至今已出版发行近3万册,在此对广大师生、读者表示衷心的感谢。为了更好地适应新形势的发展与需要,我们在广泛收集读者意见的基础上对本书进行了修订再版。本版在保留原版特点的基础上主要作了以下修改:

(1) 增加了第9章综合应用实例,给出了五个综合应用实例,如约瑟夫环问题、哈夫曼编译器、用静态栈数据结构实现表达式求值、最小生成树 Kruskal 算法、汉诺塔游戏程序,以供学生实际实习,进一步加强学生应用能力的培养。

(2) 对原书中的某些章节进行了小幅调整与修改,使内容更实用。

(3) 增加和修改了部分例题与习题,更有针对性。

(4) 修改了本书的电子课件,用户可以上清华大学出版社网站(www.tup.com.cn)下载。

本书第1、2、3、4、5、6、9章由邓文华编写,第7、8章由戴大蒙编写。浙江大学计算机学院孔繁胜教授任主审。全书由邓文华最后修改定稿。

本次修订仍然难免存在疏漏,敬请读者批评指正。

编者

2007年4月

前 言

数据结构(第2版)

数据结构是计算机程序设计的重要理论基础,该课程是计算机及其应用专业的一门重要基础课程和核心课程。它不仅是计算机软件专业课程的先导,而且也逐渐为其他工科类专业所重视。

本教材主要面向高职高专院校计算机专业的学生。教材内容结构力求体现以应用为主体,强调理论知识的理解和运用,实现专科教学以实践体系及技术应用能力培养为主的目标。

本书共分8章。第1章叙述数据、数据结构和算法等基本概念。第2~6章分别讨论了线性表、栈和队列、串和数组、树和二叉树、图等基本数据结构及其应用。第7章和第8章分别讨论了查找和排序的各种实现方法及其应用。

本教材具有以下特点:

(1) 基础理论知识的阐述由浅入深、通俗易懂。内容的组织和编排以应用为主线,省略了一些理论推导和数学证明过程,淡化了算法的设计分析和复杂的时空分析。

(2) 各章(除第1、8章外)都配有应用举例,列举分析了很多实用的例子,且大多数算法都直接给出了相应的C语言程序,以便上机练习和实践。

(3) 为了便于复习和掌握每章的重点,每章的起始处都给出了要点,并在每章结尾处给出了小结。

(4) 本教材配有相应的PowerPoint演示文稿。

本教材推荐课时数约为64学时,上机时数约为20学时(有需要的院校请与我们联系:gzgz@tup.tsinghua.edu.cn)。本书由邓文华副教授主编,编写了该书的第1~4章,第5、6章由李益明老师编写,第7、8章由戴大蒙老师编写。全书由邓文华老师最后修改定稿。浙江大学计算机学院孔繁胜教授审稿。

编写高职高专计算机专业教材是一项新的尝试,因编写时间紧迫,难免存在疏漏,敬请读者批评指正。

编 者

2004年7月

目 录

数据结构(第2版)	1
第1章 绪论	1
1.1 数据结构概念	1
1.1.1 为什么要学习数据结构	1
1.1.2 有关概念和术语	4
1.2 算法描述	7
1.2.1 算法特性	7
1.2.2 算法描述	8
1.3 算法分析	9
小结	10
习题	11
第2章 线性表	13
2.1 线性表的逻辑结构	13
2.1.1 线性表的定义	13
2.1.2 线性表的基本操作	14
2.2 线性表的顺序存储及操作实现	15
2.2.1 顺序表	15
2.2.2 顺序表的基本操作实现	17
2.2.3 顺序表应用举例	21
2.3 线性表的链式存储和运算实现	23
2.3.1 单链表	23
2.3.2 单链表上基本运算的实现	25
2.3.3 循环链表	32
2.3.4 双向链表	33

2.4 应用举例	34
小结	37
习题	37
第3章 栈和队列	40
3.1 栈	40
3.1.1 栈的定义和基本运算	40
3.1.2 栈的存储实现和运算实现	41
3.1.3 栈与递归的实现	44
3.2 栈的应用举例	50
3.3 队列	54
3.3.1 队列的定义和基本运算	54
3.3.2 队列的存储实现和运算实现	55
3.4 队列应用举例	61
小结	62
习题	63
第4章 串和数组	65
4.1 串	65
4.1.1 串的基本概念	65
4.1.2 串的基本运算	66
4.1.3 串的存储和基本运算实现	67
4.2 数组	70
4.2.1 数组的逻辑结构	70
4.2.2 数组的内存映像	71
4.2.3 稀疏矩阵	72
4.3 应用举例	76
小结	77
习题	77
第5章 树和二叉树	79
5.1 树的概念和基本操作	79
5.1.1 树的定义和相关术语	79
5.1.2 树的基本操作	81

5.2	二叉树	82
5.2.1	二叉树的基本概念	82
5.2.2	二叉树的主要性质	83
5.2.3	二叉树的基本操作与存储实现	85
5.2.4	二叉树的遍历	88
5.3	树和森林	91
5.3.1	树的存储	91
5.3.2	树、森林与二叉树的转换	94
5.3.3	树和森林的遍历	96
5.4	最优二叉树——哈夫曼树	98
5.4.1	哈夫曼树的基本概念	98
5.4.2	哈夫曼树的构造算法	99
5.4.3	哈夫曼编码	101
5.4.4	哈夫曼编码的算法实现	103
5.5	应用举例	104
	小结	107
	习题	107
第6章	图	110
6.1	图的基本概念	110
6.1.1	图的定义和术语	110
6.1.2	图的基本操作	114
6.2	图的存储表示	114
6.2.1	邻接矩阵	115
6.2.2	邻接表	117
6.3	图的遍历	119
6.3.1	深度优先搜索	119
6.3.2	广度优先搜索	121
6.4	图的应用	122
6.4.1	最小生成树	122
6.4.2	最短路径	126
6.4.3	拓扑排序	129
6.5	应用举例	133
	小结	137

习题	138
第7章 查找	141
7.1 基本概念和术语	141
7.2 静态查找表	143
7.2.1 静态查找表结构	143
7.2.2 顺序查找	143
7.2.3 有序表的折半查找	144
7.2.4 分块查找	147
7.3 动态查找表	148
7.4 哈希表	152
7.4.1 哈希表和哈希方法	152
7.4.2 常用的哈希函数	153
7.4.3 冲突处理方法	155
7.4.4 哈希表的查找及其分析	157
7.5 应用举例	159
小结	166
习题	167
第8章 排序	168
8.1 基本概念	168
8.2 简单排序方法	169
8.2.1 直接插入排序	169
8.2.2 冒泡排序	171
8.2.3 简单选择排序	172
8.3 快速排序	173
8.4 堆排序	176
8.5 归并排序	179
8.6 基数排序	181
8.6.1 多关键码排序	181
8.6.2 链式基数排序	182
8.7 排序方法的比较和讨论	184
小结	185
习题	186

第 9 章 综合应用实例.....	188
9.1 约瑟夫环问题	188
9.2 哈夫曼编译器	190
9.3 用静态栈数据结构实现表达式求值	199
9.4 最小生成树 Kruskal 算法	203
9.5 汉诺塔游戏程序	207
参考文献.....	213

绪 论

计算机科学是一门研究数据表示和数据处理的科学。数据是计算机化的信息,是计算机可以直接处理的最基本和最重要的对象。无论是进行科学计算、数据处理、过程控制,还是对文件的存储和检索及数据库技术的应用,都是对数据进行加工处理的过程。因此,要设计出一个结构好、效率高的程序,必须研究数据的特性及数据间的相互关系及其对应的存储表示,并利用这些特性和关系设计出相应的算法和程序。

本章要点

- (1) 数据、数据元素、数据结构、数据的逻辑结构与物理结构的概念以及逻辑结构与物理结构间的关系。
- (2) 数据类型的概念。
- (3) 算法的定义、特性以及算法的时间复杂度和空间复杂度。
- (4) 用 C 语言描述算法的方法,使用 C 语言编写算法程序。

1.1 数据结构概念

数据结构是计算机科学与技术专业的专业基础课,是十分重要的核心课程。所有的计算机系统软件和应用软件都要用到数据结构。因此,要想更好地运用计算机来解决实际问题,仅掌握几种计算机程序设计语言是难以应对众多复杂的课题的。要有效地使用计算机,充分发挥计算机的性能,还必须学习和掌握好数据结构的有关知识。具备数据结构的扎实基础,对于学习计算机专业的其他课程,如操作系统、编译原理、数据库管理系统、软件工程、人工智能等都是十分有益的。

1.1.1 为什么要学习数据结构

在计算机发展的初期,人们使用计算机的目的主要是处理数值计算问题。使用计算

机解决一个具体问题时,一般需要经过下列步骤。首先要从具体问题中抽象出一个适当的数学模型,然后设计或选择一个求解此数学模型的算法,最后编出程序进行调试、测试,直至得到最终的解答,如图 1-1 所示。例如,求解梁架结构中应力的数学模型的线性方程组,该方程组可以使用迭代算法来求解。

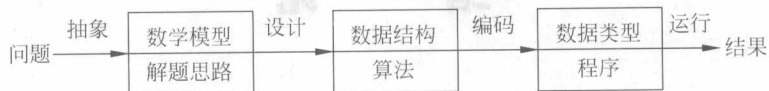


图 1-1 计算机解决问题的一般过程

由于当时所涉及的运算对象是简单的整型、实型或布尔型数据,所以程序设计者的主要精力集中于程序设计的技巧上,而无需重视数据结构。随着计算机应用领域的扩大和软、硬件的发展,非数值计算问题显得越来越重要。据统计,当今处理非数值计算性问题占用了 90% 以上的机器时间。这类问题涉及的数据结构更为复杂,数据元素之间的相互关系一般无法用数学方程式描述。因此,解决这类问题的关键不再是数学分析和计算方法,而是要设计出合适的数据结构。下面所列举的实例就是属于这一类的具体问题。

例 1.1 学生信息检索系统。当我们需要查找某个学生的有关情况,或者想查询某个专业或年级的学生的有关情况时,只要建立相关的数据结构,按照某种算法编写相关程序,就可以实现计算机自动检索。由此,可以在学生信息检索系统中建立一个按学号顺序排列的学生信息表,以及分别按姓名、专业、年级顺序排列的索引表,如图 1-2 所示。由这 4 个表构成的文件便是学生信息检索的数学模型,计算机的主要操作便是按照某个特定要求(如给定姓名)对学生信息文件进行查询。

诸如此类的还有电话自动查号系统、考试查分系统、仓库库存管理系统等。在这类文档管理的数学模型中,计算机处理的对象之间通常存在着一种简单的线性关系,因此,这类数学模型称为线性数据结构。

例 1.2 教学计划编排问题。一个教学计划包含许多课程,在这些课程中,有些课程必须按规定的先后次序进行学习,有些则没有次序要求。即有些课程之间有先修和后续的关系,有些课程可以任意安排次序。这种各个课程之间的次序关系可用称之为图的数据结构来表示,如图 1-3 所示。有向图中的每一个顶点表示一门课程,如果从顶点 v_i 到 v_j 之间存在有向边 $\langle v_i, v_j \rangle$,则表示课程 i 必须先于课程 j 进行。

由以上两个例子可见,描述这类非数值计算问题的数学模型不再是数学方程,而是诸如表、树、图之类的数据结构。因此,可以说数据结构课程主要是研究非数值计算的程序设计问题中所出现的计算机操作对象,以及它们之间的关系和操作的学科。

学号	姓名	性别	专业	年级
980001	吴承志	男	计算机科学与技术	1998级
980002	李淑芳	女	信息与计算科学	1998级
990301	刘丽	女	数学与应用数学	1999级
990302	张会友	男	信息与计算科学	1999级
990303	石宝国	男	计算机科学与技术	1999级
000801	何文颖	女	计算机科学与技术	2000级
000802	赵胜利	男	数学与应用数学	2000级
000803	崔文靖	男	信息与计算科学	2000级
010601	刘丽	女	计算机科学与技术	2001级
010602	魏永鸣	男	数学与应用数学	2001级

(a) 学生信息表

崔文靖	8
何文颖	6
李淑芳	2
刘丽	3,9
石宝国	5
魏永鸣	10
吴承志	1
赵胜利	7
张会有	4

(b) 姓名索引表

计算机科学与技术	1,5,6,9
信息与计算科学	2,4,8
数学与应用数学	3,7,10

(c) 专业索引表

2000级	6,7,8
2001级	9,10
1998级	1,2,3
1999级	4,5

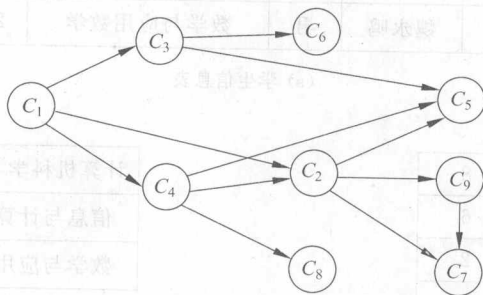
(d) 年级索引表

图 1-2 学生信息查询系统中的数据结构

学习数据结构的目的是了解计算机处理对象的特性,将实际问题中所涉及的处理对象在计算机中表示出来,并对它们进行处理。与此同时,通过算法训练提高学生的思维能力,通过程序设计的技能训练促进学生的综合应用能力和专业素质的提高。

课程编号	课程名称	先修课程
C ₁	计算机导论	无
C ₂	数据结构	C ₁ , C ₄
C ₃	汇编语言	C ₁
C ₄	C 程序设计语言	C ₁
C ₅	计算机图形学	C ₂ , C ₃ , C ₄
C ₆	接口技术	C ₃
C ₇	数据库原理	C ₂ , C ₉
C ₈	编译原理	C ₄
C ₉	操作系统	C ₂

(a) 计算机专业的课程设置



(b) 课程之间优先关系的有向图

图 1-3 教学计划编排问题的数据结构

1.1.2 有关概念和术语

在系统地学习数据结构知识之前,先对一些基本概念和术语赋予确切的定义。

数据(data)是信息的载体,它能够被计算机识别、存储和加工处理。它是计算机程序加工的原料,应用程序可以处理各种各样的数据。它可以是数值数据,也可以是非数值数据。数值数据包括整数、实数或复数,主要用于工程计算、科学计算和商务处理等;非数值数据包括字符、文字、图形、图像、语音等。

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。一个数据元素可由若干个数据项组成。在不同的条件下,数据元素又可称为元素、结点、顶点、记录等。例如,学生信息检索系统中学生信息表中的一个记录,教学

计划编排问题中的一个顶点等,都被称为一个数据元素。

数据项(data item)是不可分割的、含有独立意义的最小数据单位,有时也称为字段(field)或域。例如,学籍管理系统中学生信息表的每一个数据元素就是一个学生记录,它包括学生的学号、姓名、性别、籍贯、出生年月、成绩等数据项。这些数据项可以分为两种,一种是初等项,如学生的性别、籍贯等,这些数据项是在数据处理时不能再分割的最小单位;另一种是组合项,如学生的成绩,它可以再划分为数学、物理、化学等更小的项。通常,在解决实际应用问题时把每个学生记录当做一个基本单位进行访问和处理。

数据结构(data structure)是指相互之间存在一种或多种关系的数据元素的集合。在任何问题中,数据元素之间都不是孤立的,都会存在着这样或那样的关系,这种数据元素之间的关系称为结构。根据数据元素之间关系的不同特性,通常有以下4类基本的结构(数据的逻辑结构)。

(1) 集合结构

在集合结构中,数据元素间的关系是“属于同一个集合”。数据元素之间除了同属一个集合外,不存在其他关系。集合是元素关系中极为松散的一种结构。

(2) 线性结构

线性结构的数据元素之间存在一对一的关系。

(3) 树型结构

树型结构的数据元素之间存在一对多的关系。

(4) 图状结构

图状结构的数据元素之间存在多对多的关系。图状结构也称为网状结构。图1-4为上述4类基本结构的示意图。

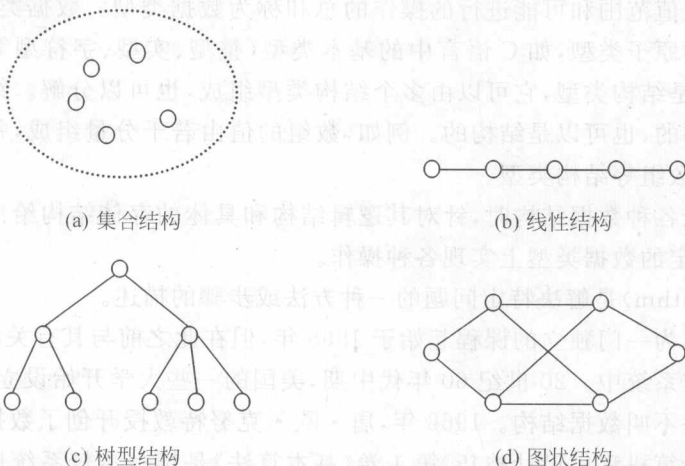


图 1-4 4类基本结构示意图