



普通高等教育“十一五”国家级规划教材

数据结构——C语言描述

(第三版)

陈慧南 编著

SHUJUEGOU
C YU YAN MIAO SHU



西安电子科技大学出版社
<http://www.xduph.com>

普通高等教育“十一五”国家级规划教材

数据结构——C 语言描述

(第三版)

陈慧南 编著

西安电子科技大学出版社

内 容 简 介

本书第二版及其配套教材为普通高等教育“十一五”国家级规划教材。本次修订除保留上版中的经典数据结构知识外,还增加了伸展树跳表等新内容。本书结构严谨,内容深入浅出,反映了抽象、封装和信息隐蔽等现代软件设计理念,重视算法的时间和空间分析,包括搜索和排序时间的下界分析。本书使用C语言描述。

本书重视实践性和程序设计。书中算法都有完整的C程序,程序代码构思精巧、结构清晰、注释详细,所有程序都已在TC 2.01下编译通过并能正确运行。这些程序既是学习数据结构和算法的很好示例,也是很好的C程序设计示例。本书最后一章是实习指导和实习题,指导学生按软件工程学的方法设计算法、编写程序和书写文档。本书配有大量的实例和图示,并有丰富的习题和实习题,易教易学。本书涵盖计算机学科专业考研大纲数据结构部分的考查内容。

本书可作为计算机类、电子信息类、电气类、自动化类、电子商务、信息管理与信息系统等相关专业数据结构课程的教材和考研参考书,也可供从事计算机软件和应用工作的工程技术人员参考。

本书与西安电子科技大学出版社出版的《〈数据结构——C语言描述〉学习指导和习题解析》一书配合使用,效果更佳。

★本书配有电子教案,需要者可登录出版社网站,免费下载。

图书在版编目(CIP)数据

数据结构: C语言描述 / 陈慧南编著. —3版. —西安: 西安电子科技大学出版社, 2015.7
普通高等教育“十一五”国家级规划教材
ISBN 978-7-5606-3747-1

I. ①数… II. ①陈… III. ①数据结构—高等学校—教材 ② C语言—程序设计—高等学校—教材 IV. ①TP311.12 ②TP312

中国版本图书馆 CIP 数据核字(2015)第 128587 号

策 划 马乐惠

责任编辑 阎 彬 马乐惠

出版发行 西安电子科技大学出版社(西安市太白南路2号)

电 话 (029)88242885 88201467 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfb001@163.com

经 销 新华书店

印刷单位 陕西天意印务有限责任公司

版 次 2015年7月第3版 2015年7月第7次印刷

开 本 787毫米×1092毫米 1/16 印张 21

字 数 490千字

印 数 19 001~22 000册

定 价 37.00元

ISBN 978-7-5606-3747-1/TP

XDUP 4039003-7

*** 如有印装问题可调换 ***

本社图书封面为激光防伪覆膜, 谨防盗版。

序

第三次全国教育工作会议以来,我国高等教育得到空前规模的发展。经过高校布局和调整,各个学校的新专业均有所增加,招生规模也迅速扩大。为了适应社会对“大专业、宽口径”人才的需求,各学校对专业进行了调整和合并,拓宽专业面,相应的教学计划、大纲也都有了较大的变化。特别是进入 21 世纪以来,信息产业发展迅速,技术更新加快。面对这样的发展形势,原有的计算机、信息工程两个专业的传统教材已很难适应高等教育的需要,作为教学改革的重要组成部分,教材的更新和建设迫在眉睫。为此,西安电子科技大学出版社聘请南京邮电学院、西安邮电学院、重庆邮电学院、吉林大学、杭州电子工业学院、桂林电子工业学院、北京信息工程学院、深圳大学、解放军电子工程学院等 10 余所国内电子信息类专业知名院校长期在教学科研第一线工作的专家教授,组成了高等学校计算机、信息工程类专业系列教材编审专家委员会,并且面向全国进行系列教材编写招标。该委员会依据教育部有关文件及规定对这两大类的教学计划和课程大纲,对目前本科教育的发展变化和相应系列教材应具有的特色和定位以及如何适应各类院校的教学需求等进行了反复研究、充分讨论,并对投标教材进行了认真评审,筛选并确定了高等学校计算机、信息工程类专业系列教材的作者及审稿人。

审定并组织出版这套教材的基本指导思想是力求精品、力求创新、好中选优、以质取胜。教材内容要反映 21 世纪信息科学技术的发展,体现专业课内容更新快的要求;编写上要具有一定的弹性和可调性,以适合多数学校使用;体系上要有所创新,突出工程技术型人才培养的特点,面向国民经济对工程技术人才的需求,强调培养学生较系统地掌握本学科专业必需的基础知识和基本理论,有较强的本专业的基本技能、方法和相关知识,培养学生具有从事实际工程的研发能力。在作者的遴选上,强调作者应在教学、科研第一线长期工作,有较高的学术水平和丰富的教材编写经验;教材在体系和篇幅上符合各学校的教学计划要求。

相信这套精心策划、精心编审、精心出版的系列教材会成为精品教材,得到各院校的认可,对于新世纪高等学校教学改革和教材建设起到积极的推动作用。

系列教材编委会

高等学校计算机、信息工程类专业 规划教材编审专家委员会

主任：杨震（南京邮电大学校长、教授）
副主任：张德民（重庆邮电大学通信与信息工程学院院长、教授）
韩俊刚（西安邮电学院计算机系主任、教授）

计算机组

组长：韩俊刚（兼）
成员：（按姓氏笔画排列）
王小民（深圳大学信息工程学院计算机系主任、副教授）
王小华（杭州电子科技大学计算机学院教授）
孙力娟（南京邮电大学计算机学院副院长、教授）
李秉智（重庆邮电大学计算机学院教授）
孟庆昌（北京信息科技大学教授）
周娅（桂林电子科技大学计算机学院副教授）
张长海（吉林大学计算机科学与技术学院副院长、教授）

信息工程组

组长：张德民（兼）
成员：（按姓氏笔画排列）
方强（西安邮电学院电信系主任、教授）
王晖（深圳大学信息工程学院电子工程系主任、教授）
胡建萍（杭州电子科技大学信息工程学院院长、教授）
徐祎（解放军电子工程学院电子技术教研室主任、副教授）
唐宁（桂林电子科技大学通信与信息工程学院副教授）
章坚武（杭州电子科技大学通信学院副院长、教授）
康健（吉林大学通信工程学院副院长、教授）
蒋国平（南京邮电大学自动化学院院长、教授）

总策划：梁家新
策划：马乐惠 云立实 马武装 马晓娟
电子教案：马武装

前 言

本书第一版于 2003 年出版。本书第二版及其配套教材《〈数据结构——C 语言描述(第二版)〉学习指导与习题解析》于 2009 年出版,均被教育部列为普通高等教育“十一五”国家级规划教材。第二版在保持第一版特色的基础上对书中内容作了较大更新,内容先进、难度适中;对内容安排、图示、示例和表述也作了必要修改,结构更合理,表述更清晰,程序注释更详细,更易为读者理解;增编的第 13 章除对实习要求和实习题给予更明确和具体的要求外,还附上了完整的实习报告样例;附录 A 增加软件文档写作和测试报告书写的内容,指导学生按软件工程原理进行实习,并按软件文档规范撰写实习报告。

教育部自 2009 年起实施计算机科学与技术专业研究生入学初试联合命题统一考试。考研大纲是教育部考试中心对当年考试范围、方法和要求的明确规定,是考试命题和学生准备考试的基本依据。计算机学科专业基础综合考试大纲是为高等院校和科研院所招收计算机科学与技术学科的硕士研究生而设置的具有选拔性质的联考科目的要求。该科目包括计算机专业最重要的数据结构、计算机组成原理、操作系统和计算机网络 4 门基础课程,其中数据结构占比 30%。

虽然本书第二版已涵盖了该考试科目中关于数据结构的考查内容,但为了更突显考研大纲的考查内容要求,本次修订将上版书中未作强调或仅作为习题的个别内容作为知识点单独列出。例如,对半插入排序在第二版中作为习题要求,并在配套教材中给出解答,现单独列为一小节,以提示学生在课程学习中留意。又如上版书中,插值和分块搜索的知识在顺序处理文件的分块插值搜索中有所涉及,修订后两者都分别作为单独小节加以讨论。此外,本次修订继续关注实践能力的培养,新增了上机考核的内容,包括上机考核目的和目标、考核要求和方式,还给出了上机考核试题样例供教学参考之用。本版新增附录 B,给出了 2015 年计算机学科专业基础综合科目考研大纲的数据结构部分,列出了该考研大纲和本书内容的对照表,以方便学生考研复习。原附录 B 改为附录 C。

作 者

2015 年 4 月 15 日于南京

第二版前言

本书为普通高等教育“十一五”国家级规划教材。

本书作者多年来在南京邮电大学讲授“数据结构”和“算法设计与分析”等课程，主持的“数据结构”课程被评为江苏省精品课程，编写出版了《数据结构——C语言描述》、《数据结构——使用C++语言描述》、《算法设计与分析》、《数据结构与算法》、《计算机软件技术基础》等教材，其中前三本已被列为普通高等教育“十一五”国家级规划教材。本书是作者在上述教材的基础上，参考了近年来国内外出版的多种数据结构和算法的优秀教材编写而成的。

本书保留经典数据结构知识，并引入伸展树和跳表等新内容，反映抽象、封装和信息隐蔽等现代软件设计理念，重视算法的时间和空间分析，包括搜索和排序时间的下界分析，内容新旧取舍恰当，广度和深度适中。

书中使用C语言描述数据结构。C语言是一种主要的软件开发语言，很多高校将其作为第一门程序设计语言来讲授。本书采用抽象数据类型的观点讨论数据结构。

本书对程序设计和实验给予了足够重视，将上机实习内容专门作为单独的一章编写，指导学生按软件工程学的方法设计和编写程序。书中算法都有完整的C程序，程序代码注释详细、结构清晰、构思精巧。所有程序都已在TC 2.01下编译通过并能正确运行，它们既是学习数据结构和算法的很好示例，也是很好的C程序设计示例。

本书结构严谨，内容深入浅出，配有大量的实例和图示，并有丰富的习题和实习题，易教易学。

全书共13章。

第1章是基础知识，讨论有关数据结构、算法、数据抽象和抽象数据类型的概念，给出数据结构和算法的描述方法，介绍算法分析的基本方法。

第2章介绍数组和链表，它们是实现本书中各种抽象数据类型的基础。本章也包含对C语言的数组、指针和链表的复习。

第3章讨论堆栈和队列、递归和递归过程、表达式的计算以及测试数据结构所需的驱动程序的编写方法。

第4章讨论线性表的逻辑结构、存储表示和运算以及多项式的算术运算；数组作为一种抽象数据类型也在本章中加以讨论。

第5章简要介绍字符串和广义表的定义、存储表示及典型算法，如字符串匹配。

第6章讨论树形结构，内容包括树和森林、二叉树、堆和优先权队列、哈夫曼树和哈夫曼编码、并查集和等价关系。

第7~9章讨论集合的表示和运算。其中，第7章主要讨论顺序搜索和二分搜索，第8章讨论二叉搜索树、二叉平衡树、B树、键树和伸展树，第9章讨论跳表和散列表。

第10章讨论图结构，着重讨论几种图算法，即图的遍历、拓扑排序、关键路径、最小

代价生成树和最短路径算法。

第 11 章介绍多种内排序算法。

第 12 章主要讨论外排序过程和算法。文件作为一种外存的数据结构，也在本章介绍。

数据结构的上机实习是课程教学不可或缺的重要环节，第 13 章和附录 A 是专门为上机实习而编写的。其中，第 13 章为实习指导和实习题，附录 A 介绍软件工程的基本概念，包括软件开发方法、软件文档写作和系统测试方法。

附录 B 是书中出现的专有名词的中英文对照表。

本书满足高等院校计算机科学与技术专业和其他相关专业的“数据结构”课程 80 学时的教学安排。对于学时数少于 80 学时的授课计划，可根据实际学时对内容加以取舍。作者已在目录中将难度较大或非基本的章节标上了*号，供读者参考。

本书的编写得到了南京邮电大学领导的帮助和关心，也得到了西安电子科技大学出版社马乐惠主任的大力推荐和支持，在此表示衷心感谢。书中若有不当之处，敬请读者批评指正。

作 者

2008 年 12 月 28 日于南京

第一版前言

有关数据结构与算法的研究是计算机科学与技术的基础性研究之一。美国计算机协会(ACM)和美国电气和电子工程师学会计算机分会(IEEE-CS)在计算学科教学计划 CC1991 报告中,将算法与数据结构定义为计算学科研究的 9 个主领域之一。该组织在 2001 年 12 月递交的 CC2001 报告中,调整了计算学科研究领域的划分,最终将计算学科划分为 14 个主领域。数据结构与算法方面的知识包含在程序设计基础(PF)、算法与复杂性(AL)和程序设计语言(PL)等领域中。

掌握该领域的知识对于我们利用计算机资源,开发高效的计算机程序是非常必要的。因此,CC2001 把其中的大多数知识单元规定为计算机及相关学科的本科学生必须掌握的核心知识单元,如基本数据结构(PF3),递归技术(PF4),数据类型和数据抽象(PL4 和 PL9),面向对象的程序设计(PL6),算法分析的基本方法(AL1),基本的计算算法(AL3)等。其中,基本数据结构包括数组、字符串、堆栈、队列、树、图和散列表等,基本的计算算法包括查找、排序、哈希表算法、搜索树及图算法。目前,国内外所有的计算机专业都开设一到两门相关的课程讲授这方面的知识。

一般来讲,“数据结构”课程涉及的知识单元有 PF3、PF4、PL4、PL6、PL9、AL1 和 AL3。从课程性质来看,“数据结构”是计算机科学与技术专业的一门核心课程,也是计算机软件和应用工作者必备的专业基础。近年来,“数据结构”课程的内容和讲授体系有了很大的变革,在介绍数据结构与算法知识的同时,普遍重视程序设计方面的训练,如 C/C++ 程序设计、Java 程序设计等。

本书在内容上注意兼顾广度和深度,不仅系统地介绍了各种传统的数据结构和各种搜索、内外排序方法,还引入了一些比较高级的数据结构,如伸展树和跳表。本书重视算法的时间和空间分析,包括对搜索和排序时间的下界分析。

本书既注重数据结构和算法原理,又十分强调程序设计训练。书中各算法都配有完整的 C 程序,程序结构清晰,构思精巧。所有程序都已在 TC 2.01 下编译通过并能正确运行,它们对 C++ 环境(如 Borland C++、Visual C++)同样适用。这些程序既是很好的学习数据结构和算法的示例,也是很好的程序设计示例。

书中使用 C 语言描述数据结构。C 语言是一种主要的软件开发语言,目前很多高校将其作为第一门程序设计语言讲授。选择 C 语言描述数据结构,使学习过 C 语言、C++ 语言的读者能够很容易地学习本书。对于只学过 Pascal 语言的读者来说,也可将本书作为学习教材或参考资料。

本书采用抽象数据类型的观点讨论数据结构。面向对象方法是在抽象数据类型的基础上发展起来的软件方法。采用抽象数据类型的观点讨论数据结构与用面向对象的观点讨论数据结构本质上是一致的。学生通过本书的学习,就可以掌握数据抽象的原理。因此,学习本书的读者如果具有 C++ 语言的知识就能方便地阅读用 C++ 语言描述的数据结构书籍。

全书条理清晰,内容详实且深入浅出,书中对算法都做了较详细的解释,尽可能做到可读易懂,并配有大量的实例和图示,有利于读者理解算法的实质和编程思想。各章结尾处的小结可帮助读者了解本章的要点。各章配有丰富的习题,适于自学。

作者多年在南京邮电学院从事“数据结构”和“算法设计与分析”课程教学，曾使用多种教材讲授数据结构和算法，其中包括 C 语言和 C++ 语言描述的《数据结构》英语原版教材。作者与他人合作编写了《数据结构》(PASCAL 语言描述)教材，1994 年由人民邮电出版社出版。2001 年作者主编了《数据结构——使用 C++ 语言描述》一书，由东南大学出版社出版。此外，作者还于 1998 年主编了《计算机软件技术基础》一书，由人民邮电出版社出版。本书是在作者上述“数据结构”课程教学和教材的基础上，参考了近年来国外出版的多种数据结构和算法的优秀教材编写而成的。

全书共分 12 章。

第 1 章是预备知识，首先给出传统的数据结构的概念，继而介绍算法规范和数据抽象，最后讨论算法的效率和算法分析的基本方法。

第 2 章介绍两种实现数据的顺序和链接存储的基本数据结构：数组和链表，它们是实现本书中各种抽象数据类型的基础。

第 3、4 章我们定义了几种线性数据结构：堆栈、队列、线性表、数组和矩阵，讨论它们的顺序表示和链接表示，并给出若干应用实例，如表达式计算、多项式的算术运算和稀疏矩阵算法。第 3 章中还讨论了递归和递归过程以及测试数据结构所必需的演示驱动程序的编写方法。

第 5 章简要介绍字符串和广义表的定义、存储表示及典型算法，如字符串匹配等内容。

第 6 章讨论树形数据结构，包括树和森林、二叉树、堆和优先权队列、哈夫曼树和哈夫曼编码、并查集和等价关系等内容。

第 7 章讨论集合和表的搜索，如顺序搜索、二分搜索(包括对半搜索和斐波那契搜索)，搜索算法的二叉判定树以及搜索算法的时间下界等内容。

第 8 章内容包括二叉搜索树、二叉平衡树、伸展树、B 树、键树等。

第 9 章讨论跳表和散列表。

第 10 章讨论图数据结构，着重讨论几种图算法：图的遍历、拓扑排序、关键路径、最小代价生成树和最短路径算法。

第 11 章介绍若干内排序算法。

第 12 章讨论文件和外排序。

附录 A 介绍软件工程的基本概念，包括软件开发方法和系统测试方法。

附录 B 对实习目的、实习要求、实习步骤、实习报告和实习题作了说明和规定。

附录 C 是书中出现的专用名词的中英文对照表。

本书可作为高等院校计算机科学与技术专业和其他相关专业的“数据结构”课程 80 学时的教材。对于学时数少于 80 学时的教学计划，可根据实际学时对本书内容加以剪裁。作者已将难度较大，或非基本的章节标上了*号，供读者参考。

本书的编写得到南京邮电学院和计算机科学与技术系领导的推荐和关心，并得到了西安电子科技大学出版社的支持，在此表示衷心感谢。

书中若有不当之处，敬请读者批评指正。

作 者

2003 年 3 月于南京

目 录

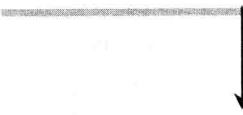
第 1 章 概论	1	2.3 链表	20
1.1 什么是数据结构	1	2.3.1 指针	20
1.1.1 基本概念	1	2.3.2 单链表	24
1.1.2 数据的逻辑结构	2	2.3.3 带表头结点的单链表	30
1.1.3 数据的存储结构	3	2.3.4 循环链表	31
1.1.4 数据结构的运算	4	2.3.5 双向链表	31
1.2 数据抽象和抽象数据类型	5	小结	33
1.2.1 抽象、数据抽象和过程抽象	5	习题 2	33
1.2.2 封装与信息隐蔽	5	第 3 章 堆栈和队列	35
1.2.3 数据类型和抽象数据类型	5	3.1 堆栈	35
1.2.4 数据结构与抽象数据类型	6	3.1.1 堆栈 ADT	35
1.3 描述数据结构	7	3.1.2 堆栈的顺序表示	36
1.3.1 数据结构的规范	7	3.1.3 堆栈的链接表示	38
1.3.2 实现数据结构	7	3.2 队列	39
1.4 算法和算法分析	9	3.2.1 队列 ADT	39
1.4.1 算法及其性能标准	9	3.2.2 队列的顺序表示	40
1.4.2 算法的时间复杂度	10	3.2.3 队列的链接表示	43
1.4.3 渐近时间复杂度	12	*3.3 表达式的计算	43
1.4.4 最坏、最好和平均情况时间 复杂度	13	3.3.1 表达式	43
1.4.5 算法的空间复杂度	13	3.3.2 中缀表达式转换为后缀表达式	44
小结	13	3.3.3 计算后缀表达式的值	48
习题 1	14	*3.4 递归和递归过程	50
第 2 章 数组和链表	16	3.4.1 递归的概念	50
2.1 结构与联合	16	3.4.2 递归的实现	51
2.1.1 结构	16	*3.5 演示和测试	53
2.1.2 联合	17	小结	55
2.2 数组	18	习题 3	55
2.2.1 一维数组	18	第 4 章 线性表和数组 ADT	57
2.2.2 二维数组	18	4.1 线性表	57
2.2.3 多维数组	20	4.1.1 线性表 ADT	57
		4.1.2 线性表的顺序表示	58

4.1.3 线性表的链接表示	61	6.2 二叉树	103
4.1.4 两种存储表示的比较	65	6.2.1 二叉树的定义和性质	103
*4.2 多项式的算术运算	65	6.2.2 二叉树 ADT	105
4.2.1 多项式 ADT	65	6.2.3 二叉树的存储表示	106
4.2.2 多项式的链接表示	66	6.2.4 二叉树的遍历	109
4.2.3 多项式的输入和输出	67	*6.2.5 二叉树遍历的非递归算法	113
4.2.4 多项式相加	69	*6.2.6 二叉树遍历的应用实例	115
4.3 数组作为抽象数据类型	70	*6.2.7 线索二叉树	117
4.4 特殊矩阵	71	6.3 树和森林	121
4.4.1 对称矩阵	71	6.3.1 森林与二叉树的转换	121
*4.4.2 带状矩阵	72	6.3.2 树和森林的存储表示	122
4.5 稀疏矩阵	73	6.3.3 树和森林的遍历	124
4.5.1 稀疏矩阵 ADT	73	*6.4 堆和优先权队列	125
4.5.2 稀疏矩阵的顺序表示	74	6.4.1 堆	125
4.5.3 稀疏矩阵转置	75	6.4.2 优先权队列	128
*4.5.4 稀疏矩阵相乘	78	6.5 哈夫曼树和哈夫曼编码	131
4.5.5 稀疏矩阵的正交链表表示	81	6.5.1 树的路径长度	131
*4.5.6 建立正交链表	83	6.5.2 哈夫曼树和哈夫曼算法	132
*4.5.7 打印正交链表	84	6.5.3 哈夫曼编码	134
小结	85	*6.6 并查集和等价关系	136
习题 4	85	6.6.1 并查集	136
第 5 章 字符串和广义表	87	6.6.2 并查集的实现	136
5.1 字符串	87	6.6.3 集合按等价关系分组	139
5.1.1 字符串 ADT	87	小结	140
5.1.2 字符串的存储表示	88	习题 6	140
5.1.3 简单模式匹配算法	89	第 7 章 集合和搜索	142
*5.1.4 模式匹配的 KMP 算法	92	7.1 集合及其表示	142
*5.2 广义表	96	7.1.1 集合和搜索	142
5.2.1 广义表的概念	96	7.1.2 集合 ADT	143
5.2.2 广义表 ADT	97	7.1.3 集合的表示	144
5.2.3 广义表的存储表示	98	7.2 顺序搜索	144
5.2.4 广义表的算法	99	7.3 二分搜索	146
小结	100	7.3.1 对半搜索	147
习题 5	100	7.3.2 二叉判定树	148
第 6 章 树	101	*7.3.3 斐波那契搜索	150
6.1 树的基本概念	101	7.3.4 插值搜索	151
6.1.1 树的定义	101	7.4 分块搜索	152
6.1.2 基本术语	102	*7.5 搜索算法的时间下界	153

小结	154	9.3.2 散列函数	197
习题 7	154	9.3.3 解决冲突的拉链法	198
第 8 章 搜索树	155	9.3.4 解决冲突的线性探查法	199
8.1 二叉搜索树	155	9.3.5 解决冲突的其他开地址法	203
8.1.1 二叉搜索树的定义	155	9.3.6 性能分析	205
8.1.2 二叉搜索树的搜索	155	小结	206
8.1.3 二叉搜索树的插入	156	习题 9	206
8.1.4 二叉搜索树的删除	158	第 10 章 图	207
*8.1.5 二叉搜索树的高度	160	10.1 图的基本概念	207
8.2 二叉平衡树	161	10.1.1 图的定义与术语	207
8.2.1 二叉平衡树的定义	161	10.1.2 图 ADT	209
8.2.2 二叉平衡树的平衡旋转	162	10.2 图的存储结构	210
8.2.3 二叉平衡树的插入	168	10.2.1 矩阵表示法	210
8.2.4 二叉平衡树的删除	171	10.2.2 邻接表表示法	214
8.2.5 二叉平衡树的高度	174	*10.2.3 正交链表和多重表表示法	217
8.3 B-树	174	10.3 图的遍历	219
8.3.1 m 叉搜索树	174	10.3.1 深度优先遍历	219
8.3.2 B-树的定义	176	10.3.2 宽度优先遍历	221
8.3.3 B-树的高度	176	10.4 拓扑排序和关键路径	223
8.3.4 B-树的搜索	177	10.4.1 拓扑排序	223
8.3.5 B-树的插入	177	*10.4.2 关键路径	227
8.3.6 B-树的删除	180	10.5 最小代价生成树	230
*8.4 键树	182	10.5.1 普里姆算法	231
8.4.1 键树的定义	182	*10.5.2 克鲁斯卡尔算法	232
8.4.2 双链树	183	*10.6 最短路径	234
8.4.3 Trie 树	184	10.6.1 单源最短路径	235
*8.5 伸展树	185	10.6.2 所有顶点之间的最短路径	238
小结	187	小结	240
习题 8	188	习题 10	241
第 9 章 跳表和散列表	189	第 11 章 内排序	243
9.1 字典	189	11.1 排序的基本概念	243
*9.2 跳表	189	11.2 插入排序	244
9.2.1 什么是跳表	190	11.2.1 直接插入排序	244
9.2.2 跳表的搜索	193	*11.2.2 希尔排序	248
9.2.3 跳表的插入	194	11.2.3 对半插入排序	249
9.2.4 跳表的删除	195	11.3 交换排序	249
9.3 散列表	196	11.3.1 冒泡排序	250
9.3.1 散列技术	196	11.3.2 快速排序	251

11.4 合并排序	256	13.4 实习题	289
11.4.1 两路合并排序	256	实习 1 数组操作	289
11.4.2 合并排序的迭代算法	256	实习 2 链表操作	290
*11.4.3 链表上的合并排序	258	实习 3 表达式计算	290
11.5 选择排序	261	实习 4 队列运算和用户界面设计 ...	291
11.5.1 简单选择排序	262	实习 5 线性表运算及应用	291
*11.5.2 堆排序	263	实习 6 一元多项式的相加和相乘 ...	291
*11.6 排序算法的时间下界	264	实习 7 对称矩阵的压缩存储	292
*11.7 基数排序	265	实习 8 稀疏矩阵的三元组表	292
小结	269	实习 9 稀疏矩阵的正交链表	292
习题 11	269	实习 10 字符串运算和文本处理	293
第 12 章 文件和外排序	271	实习 11 二叉树的基本运算和	
*12.1 辅助存储器简介	271	应用	293
12.1.1 主存储器和辅助存储器	271	实习 12 哈夫曼编码和译码系统	294
12.1.2 磁盘存储器	271	实习 13 B-树检索	294
12.2 文件	273	实习 14 散列表检索	295
12.2.1 文件的基本概念	273	实习 15 图运算及其应用	295
12.2.2 文件的组织方式	273	实习 16 内排序算法及其	
12.2.3 C 语言文件	277	性能比较	296
12.3 文件的索引结构	278	实习 17 外排序	296
12.3.1 静态索引结构	278	13.5 实习报告范例	297
12.3.2 动态索引结构	279	13.5.1 实习题: 表达式计算	297
*12.4 外排序	280	13.5.2 实习报告	297
12.4.1 外排序的基本过程	280	13.6 上机考核	303
12.4.2 初始游程的生成	280	13.6.1 考核目的	303
12.4.3 多路合并	282	13.6.2 考核目标	303
12.4.4 最佳合并树	284	13.6.3 考核要求	303
小结	285	13.6.4 软件环境	303
习题 12	286	13.6.5 考核方式	303
第 13 章 实习指导和实习题	287	13.6.6 试题样例	303
13.1 实习目的和要求	287	附录 A 软件工程概述	305
13.1.1 实习目的	287	附录 B 考研大纲和教材内容	312
13.1.2 实习要求	287	附录 C 专用名词中英文对照表	316
13.2 实习步骤	288	参考文献	322
13.3 实习报告	288		

第 I 章



概 论

这一章中，我们首先给出传统的数据结构概念，继而介绍数据抽象和抽象数据类型概念，阐明其与数据结构的关系，然后给出基于抽象数据类型的数据结构描述方法，最后介绍算法的效率和算法分析的基本方法。

1.1 什么是数据结构

1.1.1 基本概念

数据结构是计算机科学与技术领域中广泛使用的术语，然而，究竟什么是数据结构，在计算机科学界至今没有标准的定义。

计算机由硬件和软件组成。硬件是躯体，软件是灵魂。硬件通过软件发挥效用。软件的核心是程序。学习程序设计需要掌握一门程序设计语言，它是学习计算机后续课程所必需的技能。但程序设计不等于编码，为了充分利用计算机资源，开发高效的程序，计算机技术人员还必须掌握计算机学科多方面的知识，如数据的组织、算法的设计和分析、软件工程技术等。

随着计算机科学与技术的发展，计算机应用已远远超出了单纯进行科学计算的范围。从传统的应用领域，如工业控制、情报检索、企业管理、商务处理、图形图像、人工智能等诸多的数据处理领域，发展到电子政务、电子商务、办公自动化、企业资源管理、电子图书馆、远程教育、远程医疗等更广泛的领域。计算机技术已渗透到国民经济的各行各业和人们日常生活的方方面面。今天，信息技术作为现代技术的标志已成为世界各国经济增长的主要动力。

现实世界各领域中的大量信息都必须转换成数据才能在计算机中存储、处理。数据是信息的载体。应用程序可以处理各种各样的数据。笼统地说，所谓**数据(data)**，就是计算机加工处理的对象。数据一般分为两类：**数值数据(numerical data)**和**非数值数据(non-numerical data)**。数值数据是一些整数、实数或复数，主要用于工程计算、科学计算、商务处理等。非数值数据包括字符、文字、图形、图像、语音、表格等。这类数据的特点是量大，而且往往有着复杂的内在联系。如果单纯依靠改进程序设计技巧，已无法编制出高效可靠的程序，而必须对数据本身的结构加以研究。数据的组织和表示方法直接影响使用计算机求解问题的效率。算法设计通常建立在所处理数据的一定组织形式之上。在许多应用中，对于相同数据的同样的处理要求，如果选择不同的数据结构，会有不同的处理效率，即运算时间和存储

空间都会不同。

数据结构主要是为研究和解决如何使用计算机处理这些非数值问题而产生的理论、技术和方法。

对计算机学科来说，数据结构与算法的概念是至关重要的，它们是计算机学科的基础之一，更是软件技术的基础。数据结构与算法之间有着本质的联系。当谈论一种算法时，自然要涉及算法所处理的数据问题；而讨论数据的组织或结构，离开了对处理此类数据的运算及其算法的研究也是无意义的。因此有人概括出一个公式，即

$$\text{程序} = \text{算法} + \text{数据结构}$$

ACM 和 IEEE-CS 的计算学科教学计划 1991 将计算学科划分为 9 个主领域，“算法与数据结构”是其中之一。计算学科教学计划 2001 调整为 14 个主领域，数据结构和算法的基本内容主要涵盖在“程序设计基础”和“算法与复杂性”两个领域中。

如前所述，计算机处理的对象称为数据。一个数据可以由若干成分数据构成，并具有某种结构。在这里，我们称组成数据的成分数据为**数据元素(data element)**。数据元素可以是简单类型的，如整数、实数、字符等，也可以是结构类型的，如记录。如果把每个学生的记录看做一个数据元素，那么该数据元素包括学号、姓名、性别等**数据项(data item)**。一个班学生的记录组成了表 1-1 所示的学生情况表。

表 1-1 学生情况表

学号	姓名	性别	其他信息
B02040101	王小红	女	...
B02040102	林悦	女	...
B02040103	陈菁菁	女	...
B02040104	张可可	男	...
⋮	⋮	⋮	⋮

从概念上讲，一个**数据结构(data structure)**是由数据元素依据某种逻辑联系组织起来的。对数据元素间的逻辑关系的描述称为数据的**逻辑结构(logical structure)**。数据必须在计算机内存储，数据的**存储结构(storage structure)**是数据结构的实现形式，是其在计算机内的表示。此外，讨论一个数据结构必须同时讨论在该类数据上执行的运算(operation)才有意义。表是一个数据结构。

1.1.2 数据的逻辑结构

根据数据结构中数据元素之间的结构关系的不同特征，可将其分为四类基本的逻辑结构。图 1-1 给出了它们的示意图。图中用小圆圈表示数据元素，用带箭头的线表示元素间的次序关系。两个不同元素的连线称为**边**，边的起点称为**前驱(predecessor)**元素，终点称为**后继(successor)**元素。这四种基本逻辑结构是：

(1) **集合结构(set)**。集合结构中，元素间的次序是随意的。元素之间除了“属于同一个集合”的联系之外没有其他关系。由于集合结构的元素间没有固有的关系，因此往往需要借助其他结构才能在计算机中实际表示此结构。

(2) **线性结构(linear)**。线性结构是数据元素的有序序列，其中，第一个元素没有前驱只

有后继, 最后一个元素只有前驱没有后继, 其余元素有且仅有一个前驱和一个后继。数据元素之间形成一对一的关系。

(3) **树形结构(tree)**。树中, 除一个称为**根**的特殊元素没有前驱只有后继外, 其余元素都有且仅有一个前驱, 但后继的数目不限。对于非根元素, 都存在着一从根到该元素的路径。树中的数据元素之间存在一对多的关系。树是层次数据结构。

(4) **图状结构(graph)**。图是最一般的数据结构, 图中每个元素的前驱和后继的数目都不限。图中数据元素之间的关系是多对多的关系。

上述四种基本结构关系可分为两类:**线性结构(linear structure)**和**非线性结构(non-linear structure)**。我们把除了线性结构以外的几种结构关系——树、图和集合都归入非线性结构一类。

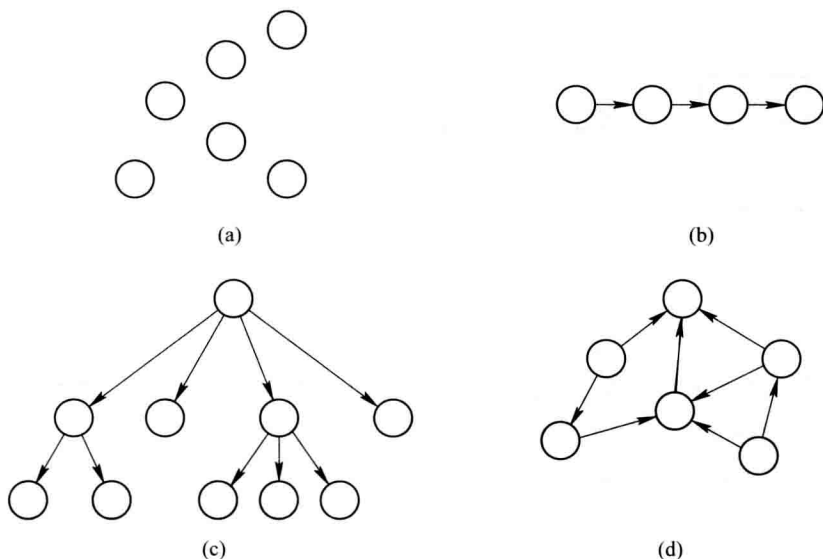


图 1-1 四种基本结构示意图

(a) 集合结构; (b) 线性结构; (c) 树形结构; (d) 图状结构

1.1.3 数据的存储结构

数据的逻辑结构是面向应用问题的, 是从用户角度看到的数据结构。数据必须在计算机内存储, 数据的存储结构是数据在计算机内的组织方式, 是逻辑数据的存储映像, 它是面向计算机的。

我们知道, 计算机内存是由有限个存储单元组成的一个连续存储空间, 这些存储单元或者是字节编址的, 或者是字编址的。从存储器角度看, 内存中是一堆二进制数据, 它们可以被机器指令解释为指令、整数、字符、布尔数等, 也可以被数据结构的算法解释为具有某种结构的数据。

为一个数据结构找到一种有效的表示方法, 使它适于计算机存储和处理是十分重要的。顺序(sequential)存储结构和链接(linked)存储结构是两种最基本的存储表示方法。

顺序(或称连续(contiguous))表示方法需要一块连续的存储空间, 并把逻辑上相关的数据元素依次存储在该连续的存储区中。例如由 4 个元素组成的线性数据结构(a_0, a_1, a_2, a_3)存储在某个连续的存储区内, 设存储区的起始地址是 102, 假定每个元素占 2 个存储单元,