

数据结构 及应用 C语言描述

沈 华 等编著

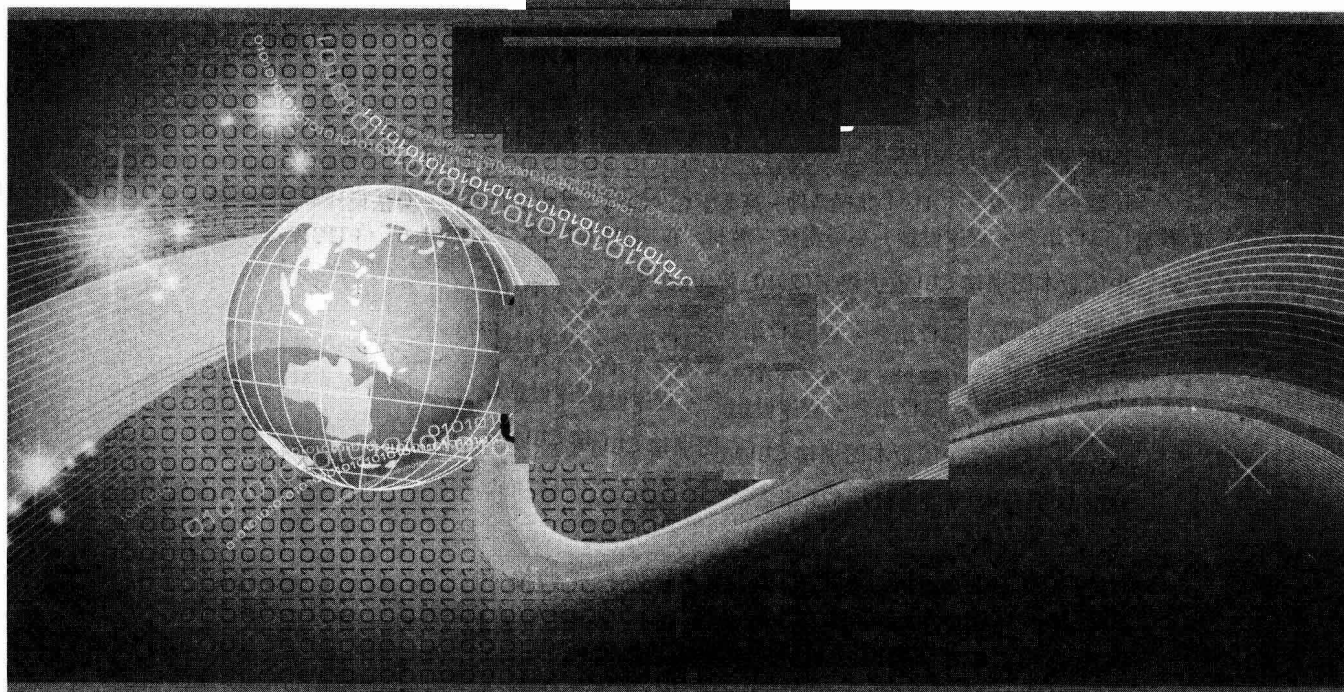


机械工业出版社
China Machine Press

高等院校计算机教材系列

数据结构 及应用 C语言描述

沈华 杨晓艳 马驰 杨华 编著



机械工业出版社
China Machine Press

本书系统地介绍各种常用的数据结构以及排序、查找的各种算法,阐述各种数据结构的逻辑关系、存储表示及运算,涵盖研究生入学考试大纲的所有内容。全书采用C语言作为数据结构和算法的描述语言,并对C语言描述的算法作了详细的注解和简要的性能分析。全书共分为六个部分:第一部分主要介绍什么是数据结构,什么是算法,它们之间有着怎样的联系,如何进行算法分析;第二部分针对后续学习的需要帮助读者温习一些相关知识;第三部分和第四部分分别重点介绍几种常见的线性结构和非线性结构;第五部分介绍在实际应用中最常遇到的两个运算——查找(即搜索)和排序,以及实现这两种运算的各种算法;第六部分则简要介绍文件和外排序的相关内容。

为了帮助读者直观、正确地理解各种数据结构和算法的要旨,本书利用大量的图表进行诠释,并通过典型的思考题、例题和习题来加深读者对相关知识的理解。

本书内容丰富、概念清楚、逻辑推理严谨、通俗易懂,可以作为计算机科学与技术及相关专业本科生的教材,也可以作为高等院校计算机专业硕士研究生入学考试的复习用书,同时还可以作为广大工程技术人员的参考资料。

封底无防伪标均为盗版

版权所有,侵权必究

本书法律顾问 北京市展达律师事务所

图书在版编目(CIP)数据

数据结构及应用:C语言描述/沈华等编著. —北京:机械工业出版社,2010.10
(高等院校计算机教材系列)

ISBN 978-7-111-32155-2

I. 数… II. 沈… III. ①数据结构—高等学校—教材 ②电子计算机—算法设计—高等学校—教材 ③C语言—程序设计—高等学校—教材 IV. ①TP311.12 ②TP301.6 ③TP312

中国版本图书馆CIP数据核字(2010)第196013号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑:李荣 张少波

北京瑞德印刷有限公司印刷

2011年1月第1版第1次印刷

185mm×260mm·18.25印张

标准书号:ISBN 978-7-111-32155-2

定价:30.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

客服热线:(010)88378991;88361066

购书热线:(010)68326294;88379649;68995259

投稿热线:(010)88379604

读者信箱:hzjsj@hzbook.com

序

在计算机科学中，数据结构是一种在计算机中组织和存储数据，以便高效利用这些数据的有效方式。几乎所有的程序或软件系统都用到了数据结构。数据结构是许多高效算法的基本要素，同时它也使得管理大规模数据成为可能。数据结构课程是计算机科学的一门非常重要的专业基础课，也是 IT 类各专业的核心基础课。

数据结构中的数据和结构是两个紧密联系而又相互关联的概念，数据是数据结构的主要组成部分，它不涉及数据之间的关系；结构才涉及而且只涉及数据之间的关系。从中文构词上来看，数据结构更强调的是结构，即数据之间的关联方式。我们需要将适用于计算机的问题求解策略用计算机能理解的形式输入到计算机中，告诉计算机如何一步一步地处理数据并最终得到问题的解，这就是算法。因此，从抽象层面上讲，数据结构是为算法服务的。不同的应用需求需要不同的算法，不同的算法需要不同的数据结构来支持。从实现层面上讲，数据结构是为算法实现服务的。

本书作者具有多年的教学和科研经验，对各种数据结构及应用算法有深入的了解。本书对什么是数据结构，什么是算法，算法和数据结构之间的关系进行了生动的阐述，并对各种线性数据结构、非线性数据结构、查找算法和排序算法等作了详尽的描述和分析。本书基本概念清晰，重点和难点问题讨论深入，而且循序渐进，为读者深入理解和应用数据结构给出了启示。本书既注重对数据结构经典内容的论述，又强调数据结构的应用，对相关内容进行了启发式讨论，是一本值得阅读和选用的教科书。

何炎祥

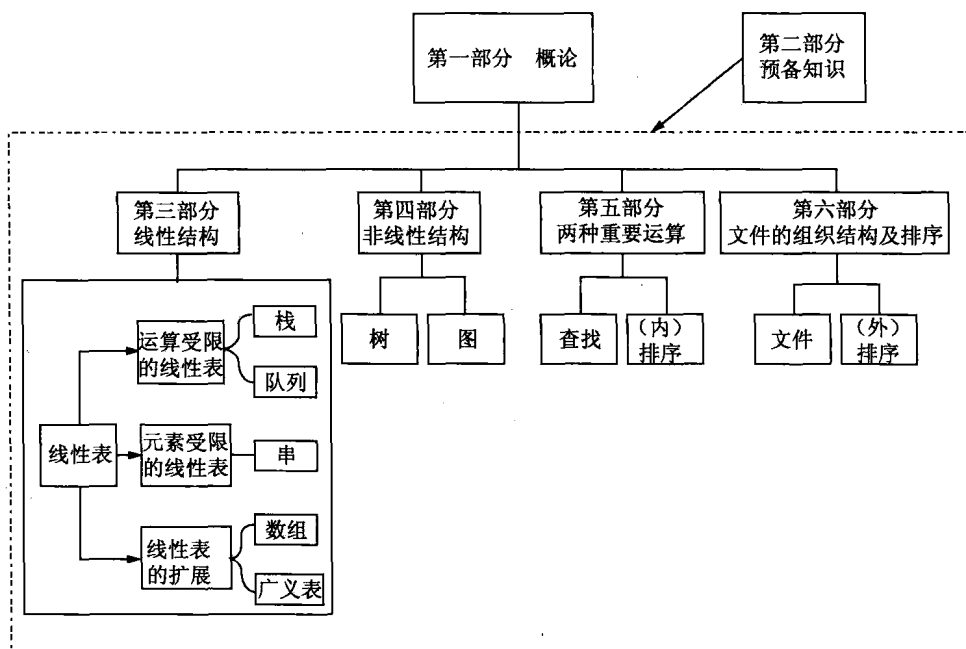
武汉大学计算机学院

2010 年 8 月 18 日于武汉珞珈山

前言

在国外是从 1968 年才开始将“数据结构”作为一门独立的课程的。1968 年，美国唐纳德 E. 克努特教授开创了数据结构的最初体系，他所著的《计算机程序设计艺术：第 1 卷基本算法》是第一本较系统地阐述数据的逻辑结构和存储结构及其操作的著作。“数据结构”是计算机科学与技术、信息类相关专业的一门重要的综合性专业基础课。它是介于数学、计算机硬件和计算机软件三者之间的一门核心课程。当用计算机解决实际问题时，必然会涉及信息表示和信息处理的问题，而这正是“数据结构”课程主要的研究内容。“数据结构”不仅是一般程序设计（特别是非数值性程序设计）的基础，也是编译程序、操作系统、数据库系统等系统程序的重要基础。

本书分六部分共 14 章，如下图所示。



第一部分包括第 1 章和第 2 章，主要介绍数据结构的概念、算法的概念、数据结构和算法之间的密切联系以及简单的算法分析。

第二部分包括第 3 章，主要介绍 C 语言中的函数参数传递及函数处理结果返回、结构体、指针等相关内容，两种递归方式，以及操作系统为一个准备运行的程序分配内存空间的情况。

第三部分包括第4章至第8章，这一部分以“线性表”为主线分别详细讨论六种常见的线性结构。

第四部分包括第9章和第10章，重点介绍两种重要的非线性结构——树和图。

第五部分包括第11章和第12章，讨论两种重要的运算——查找和（内部）排序，在介绍各种经典的查找算法和（内部）排序算法的同时，展现各种数据结构的完美应用。

第六部分包括第13章和第14章，讨论数据在外存上的存储形式，以及如何对外存中的文件进行排序。这部分内容为选学内容，章名前以“*”号指示。

本书主要采用面向过程的C语言作为数据结构和算法的描述手段，在保持C语言优点的同时，尽量使算法描述简单清晰。

本书配套的教辅资源包括：电子课件以及课后习题参考答案。

在本书的编写过程中，机械工业出版社负责本书的编辑为本书的出版付出了大量辛勤的劳动。何炎祥教授在百忙之中认真审阅了全书，提出了许多宝贵和中肯的意见。在此，谨向每一位关心和支持本书编写工作的各位朋友、老师表示衷心的感谢！

由于作者的知识水平和写作水平有限，书中难免有错误和不足之处，恳请各位专家、读者批评指正。作者 E-mail: nancy78733@126.com。

沈华

2010年8月于武汉南湖

教学建议

教学内容	教学要点及教学要求	课时安排	
		计算机专业	非计算机专业
第1章 数据结构	- 熟悉并掌握数据、数据元素、数据项的概念 - 熟悉并掌握数据的逻辑结构、数据的物理结构、数据的运算的概念, 以及常见的几种数据的逻辑结构和物理结构, 并充分理解数据结构的涵义 - 熟悉并掌握数据类型、抽象数据类型的概念	2	1 (选讲)
第2章 算 法	- 熟悉并掌握算法的概念和算法的五个特性, 以及算法和程序之间的区别和联系 - 充分理解算法和数据结构之间的关系 - 了解描述算法的方法 - 掌握一般的算法时间复杂度和空间复杂度的分析方法, 能够对给定算法进行时间和空间的复杂度分析	2	1 (选讲)
第3章 C语言、递归及 存储分配方式	- 了解程序设计语言——C语言中的相关内容 - 熟悉并掌握递归的概念 - 了解存储分配的三种方式	0~2	0~2 (选讲)
第4章 线 性 表	- 熟悉并掌握抽象数据类型——线性表 - 熟悉并掌握线性表的几种存储结构和在不同存储结构上线性表的基本运算的实现过程, 应重点掌握的存储结构是顺序表和单链表 - 能够对适合应用线性表的实际问题进行线性表抽象, 并能够根据应用环境和应用要求为其选择合适的存储结构, 同时能够给出实现算法	6~8	4~10 (选讲)
第5章 栈	- 熟悉并掌握抽象数据类型——栈, 重点掌握栈与线性表之间的关系、栈的结构特性 - 熟悉并掌握栈的存储结构——顺序栈和链栈, 以及在不同存储结构上栈的基本运算的实现过程 - 了解两个方向生成的栈 - 能够对适合应用栈的实际问题进行栈抽象, 并能够根据应用环境和应用要求为其选择合适的存储结构, 同时能够给出实现算法	2~4	2 (选讲)

(续)

教学内容	教学要点及教学要求	课时安排	
		计算机专业	非计算机专业
第6章 队 列	- 熟悉并掌握抽象数据类型——队列，重点掌握队列与线性表之间的关系、队列的结构特性、队列和栈的共性与个性 - 熟悉并掌握队列的存储结构——链队列、顺序队列和循环队列，以及在不同存储结构上队列的基本运算的实现过程，重点掌握循环队列的产生、实现、判空条件和判满条件 - 了解双端队列 - 能够对适合应用队列的实际问题进行栈抽象，并能够根据应用环境和应用要求为其选择合适的存储结构，同时能够给出实现算法	2~4	2 (选讲)
第7章 串	- 熟悉并掌握抽象数据类型——串，重点掌握串与线性表之间的关系、空串与空格串的区别、串相等的概念、模式匹配的概念、串运算的特点 - 熟悉并掌握串的存储结构——顺序串、串的堆结构、串的块链结构，以及在不同存储结构上串的基本运算的实现过程，重点掌握串的堆结构的存储思想 - 了解表示顺序串串长的显式方式和隐式方式 - 掌握朴素模式匹配算法，了解 KMP 算法	2~4	2 (选讲)
第8章 数组及广义表	- 熟悉并掌握抽象数据类型——数组，重点掌握数组与线性表之间的关系、数组维数的概念，明确一维数组与向量、二维数组与矩阵之间的关系 - 了解将多维数组映射到一维存储空间的方法，重点掌握二维数组映射到一维存储空间的方法，能够将二维数组元素的二维地址转换为一维存储地址，深刻理解数组的“随机访问性” - 了解特殊矩阵（包括特殊形状矩阵和稀疏矩阵）的概念以及对特殊矩阵进行压缩的意义和目标 - 熟悉并掌握几种常见特殊形状矩阵的压缩存储方法，并且能够通过矩阵元素的行列下标得到其在压缩向量中的下标 - 熟悉并掌握稀疏矩阵的压缩存储方法，重点掌握稀疏矩阵的三元组顺序表压缩存储方法，以及基于三元组顺序表实现的矩阵的转置运算的算法 - 熟悉并掌握广义表的概念、广义表与线性表之间的关系、广义表的性质、广义表的基本运算 - 了解广义表的存储结构	6	4 (选讲)
第9章 树	- 熟悉并掌握（根）树的概念及相关术语，掌握（根）树的几种存储结构以及每种存储结构的优缺点 - 熟悉并掌握二叉树的概念，深刻理解二叉树和树的区别，熟悉并能灵活应用二叉树的五个性质，掌握五个性质的证明思路，掌握满二叉树和完全二叉树的概念、特点以及它们之间的关系	8	4~6 (选讲)

(续)

教学内容	教学要点及教学要求	课时安排	
		计算机专业	非计算机专业
第9章 树	- 熟悉并掌握二叉树的几种存储结构, 以及每种存储结构的优缺点 - 熟悉并掌握二叉树的先序、中序、后序遍历的递归和非递归算法以及层次遍历的实现算法, 并能够灵活运用二叉树的遍历算法解决实际问题 - 掌握线索、线索二叉树、二叉树的线索化的概念和给二叉树加线索的方法, 了解对线索二叉树进行遍历的方法以及线索在遍历过程中的作用 - 掌握哈夫曼树的概念、生成方法、特点以及应用(哈夫曼编码), 掌握二叉排序树的概念、特点、生成(插入运算)和删除运算, 了解平衡二叉树的概念和各种失去平衡情况下的调整规则 - 了解树、森林和二叉树的相互转换过程和树转换得到的二叉树的特点, 并深刻体会二叉树的二叉链表和树的左孩子右兄弟表示法之间的联系 - 了解树和森林的遍历算法以及对应二叉树的先序、中序遍历之间的关系 - 了解树的相关应用	8	4 ~ 6 (选讲)
第10章 图	- 熟悉并掌握图的概念、相关术语, 掌握图的两种常用存储结构和每种存储结构的特点以及图的创建算法 - 熟悉并掌握图的深度优先搜索遍历和广度优先搜索遍历, 并能够灵活运用图的两种遍历算法解决实际问题 - 熟悉并掌握连通图的生成树、连通网络的最小生成树的概念, 掌握最小生成树的两种生成算法 - 了解什么是图的最短路径问题, 重点掌握求解网络单源最短路径的算法和求解网络每对顶点间最短路径的算法 - 掌握有向无环图的两种主要应用——AOV网的拓扑排序、逆拓扑排序和AOE网的关键路径, 并深刻理解它们对实际工程和应用问题的指导意义	6	4 ~ 6 (选讲)
第11章 查 找	- 熟悉并掌握查找的概念、分类以及其他相关概念, 明确静态查找和动态查找的区别 - 重点掌握各种查找法和它们的性能评价, 重点掌握二分查找、散列查找、二叉排序树(在第9章给出), 深刻了解散列查找法与其他查找方法之间的区别	4	4 (选讲)
第12章 内 排 序	- 熟悉并掌握排序的概念、分类以及其他相关概念, 明确内排序和外排序的区别 - 重点掌握各种排序法和它们的性能评价及稳定性, 重点掌握直接插入排序法、希尔排序法、冒泡排序法、快速排序法、直接选择排序法、堆排序法、归并排序法、基数排序法	4 ~ 6	4 (选讲)

(续)

教学内容	教学要点及教学要求	课时安排	
		计算机专业	非计算机专业
* 第 13 章 文 件	- 掌握文件的概念以及其他相关概念 - 了解文件的几种存储结构——顺序文件、索引文件、索引顺序文件、散列文件、多关键字文件	2 ~ 4	0 ~ 2 (选讲)
* 第 14 章 外 排 序	- 掌握外排序的概念, 了解磁盘的组成和工作原理 - 了解几种常用的外排序方法——多路平衡归并、置换选择排序、最佳归并树	2 ~ 4	0 ~ 2 (选讲)
教学总学时建议		48 ~ 64	32 ~ 48

说明:

- ① 计算机专业本科教学使用本教材时, 建议课堂授课时数为 48~64 (包含习题课、课堂讨论等必要的课堂教学环节, 实验另行安排学时), 不同学校可以根据各自的教学要求和计划学时数酌情对教材内容进行取舍。
- ② 非计算机专业的师生使用本教材时, 可适当降低教学要求。若授课时数少于 48, 建议主要学习几种常见的数据结构——线性表、栈、队列、串、数组、树和图, 有关算法分析、查找、排序的内容可以适当简化, 文件和外排序可以不作要求。

课堂教学建议:

- ① 如果 C 语言基础较好, 则可以跳过第 3 章。
- ② 本书的重点是第 1 章、第 2 章、第 4 章、第 5 章、第 6 章、第 7 章、第 8 章、第 9 章、第 10 章、第 11 章、第 12 章, 这几章对于掌握数据结构的涵义、几种常见的数据结构、灵活应用数据结构是至关重要的。也就是说, 这几章的内容充分说明了 4 个问题: 什么是数据结构? 有哪些数据结构? 为什么需要数据结构? 如何应用数据结构?
- ③ 如果受学时限制, 第 3 章、第 13 章、第 14 章可略去不讲。

实验教学建议:

- ① 实验一: 线性表的应用。
- ② 实验二: 线性结构部分的综合性实验。
- ③ 实验三: 二叉树的创建以及二叉树遍历算法的应用。
- ④ 实验四: 图的创建以及图遍历算法的应用。
- ⑤ 实验五: 查找部分的综合性实验。
- ⑥ 实验六: 内排序部分的综合性实验。

目 录

序	
前言	
教学建议	

第一部分 概 论

第1章 数据结构	3
1.1 什么是数据	3
1.2 什么是数据结构	3
1.2.1 数据的逻辑结构	3
1.2.2 数据的存储结构	4
1.2.3 数据的运算	5
1.3 什么是数据类型	5
1.4 知识点小结	6
习题	6
第2章 算法	7
2.1 什么是算法	7
2.2 算法的描述	7
2.3 算法分析	8
2.3.1 时间复杂度	8
2.3.2 渐近符号	9
2.3.3 空间复杂度	10
2.3.4 复杂度分析举例	10
2.4 知识点小结	13
习题	13

第二部分 预备知识

第3章 C语言、递归及存储分配方式	16
3.1 C语言的相关内容	16

3.1.1 函数的参数传递与结果返回	16
3.1.2 结构体类型	17
3.1.3 指针	18
3.2 递归	18
3.3 存储分配方式	19
3.4 知识点小结	20
习题	20

第三部分 线性结构

第4章 线性表	22
4.1 线性表的类型定义	22
4.1.1 线性表的逻辑结构	22
4.1.2 线性表的基本运算	22
4.2 线性表的顺序存储表示	23
4.2.1 顺序表	23
4.2.2 顺序表中基本运算的实现	24
4.3 线性表的链式存储表示	30
4.3.1 单链表	30
4.3.2 单链表中基本运算的实现	31
4.4 线性表的其他链式存储表示	38
4.4.1 静态单链表	38
4.4.2 双(向)链表	41
4.4.3 循环单(向)链表	43
4.4.4 循环双(向)链表	46
4.5 线性表的应用举例	47
4.6 顺序表和链表的比较	48
4.7 知识点小结	49
习题	49
第5章 栈	51
5.1 栈的类型定义	51
5.1.1 栈的逻辑结构	51

5.1.2 栈的基本运算	52	8.2 数组的顺序存储表示	90
5.2 栈的顺序存储表示	52	8.3 特殊矩阵的压缩存储	92
5.2.1 顺序栈	52	8.3.1 特殊形状矩阵的压缩存储	93
5.2.2 顺序栈中基本运算的实现	53	8.3.2 随机稀疏矩阵的压缩存储 及其运算	95
5.3 栈的链式存储表示	55	8.4 广义表	104
5.3.1 链栈	55	8.4.1 广义表的基本概念	104
5.3.2 链栈中基本运算的实现	55	8.4.2 广义表的基本运算	105
5.4 两个方向生长的栈	56	8.4.3 广义表的存储结构	106
5.5 栈的应用举例	57	8.5 知识点小结	108
5.6 知识点小结	61	习题	108
习题	61		
第6章 队列	63		
6.1 队列的类型定义	63	第四部分 非线性结构	
6.1.1 队列的逻辑结构	63	第9章 树	110
6.1.2 队列的基本运算	63	9.1 概述	110
6.2 队列的链式存储表示	64	9.1.1 树的定义及基本术语	110
6.2.1 链队列	64	9.1.2 树的存储结构	112
6.2.2 链队列中基本运算的实现	65	9.2 二叉树	123
6.3 队列的顺序存储表示	66	9.2.1 二叉树的定义	123
6.3.1 顺序队列	66	9.2.2 二叉树的性质	123
6.3.2 循环队列	69	9.2.3 二叉树的存储结构	128
6.3.3 循环队列中基本运算的实现	71	9.3 二叉树的遍历	132
6.4 双端队列	74	9.3.1 遍历操作	132
6.5 队列的应用举例	74	9.3.2 先序遍历	132
6.6 知识点小结	75	9.3.3 中序遍历	133
习题	76	9.3.4 后序遍历	134
第7章 串	77	9.3.5 层次遍历	135
7.1 串的类型定义	77	9.3.6 二叉树遍历的应用举例	136
7.1.1 串的逻辑结构	77	9.4 线索二叉树	144
7.1.2 串的基本运算	77	9.4.1 二叉树的线索化	145
7.2 串的顺序存储表示	78	9.4.2 线索二叉树上的运算	149
7.3 串的堆分配存储表示	80	9.5 二叉树的应用	156
7.4 串的块链存储表示	81	9.5.1 哈夫曼树及其应用	156
7.5 串的模式匹配	82	9.5.2 二叉排序树	161
7.6 知识点小结	88	9.5.3 平衡二叉树	163
习题	88	9.6 树、森林与二叉树的相互转换	166
第8章 数组及广义表	89	9.6.1 树与二叉树的相互转换	166
8.1 数组的类型定义	89	9.6.2 森林与二叉树的相互转换	168
8.1.1 数组的定义	89	9.7 树、森林的遍历	169
8.1.2 数组的性质	89	9.7.1 树的遍历	169
8.1.3 数组的基本运算	89	9.7.2 森林的遍历	169

9.8 树的应用举例	170
9.9 知识点小结	171
习题	172
第10章 图	173
10.1 概述	173
10.1.1 图的定义及基本术语	173
10.1.2 图的存储结构	179
10.1.3 图的创建	184
10.2 图的遍历	187
10.2.1 深度优先搜索遍历	187
10.2.2 广度优先搜索遍历	191
10.2.3 图遍历的应用举例	193
10.3 生成树	194
10.3.1 连通图的生成树	194
10.3.2 连通网的最小生成树	195
10.4 最短路径	197
10.4.1 单源最短路径	197
10.4.2 每对顶点间的最短路径	201
10.5 有向无环图及其应用	205
10.5.1 AOV网与拓扑排序	205
10.5.2 AOE网与关键路径	207
10.6 知识点小结	209
习题	209

第五部分 两种重要运算

第11章 查找	212
11.1 查找的基本概念	212
11.2 主要查找方法简介	213
11.3 静态查找	213
11.3.1 顺序查找	213
11.3.2 二分查找	215
11.3.3 分块查找	219
11.4 动态查找	220
11.5 散列查找	227
11.5.1 散列表的概念	228
11.5.2 散列函数的构造方法	228
11.5.3 处理冲突的方法	229
11.5.4 散列表的查找	232
11.6 知识点小结	234

习题	235
第12章 内排序	236
12.1 排序的基本概念	236
12.2 插入排序	237
12.2.1 直接插入排序	237
12.2.2 希尔排序	241
12.3 交换排序	242
12.3.1 冒泡排序	242
12.3.2 快速排序	244
12.4 选择排序	247
12.4.1 直接选择排序	247
12.4.2 树形选择排序	248
12.4.3 堆排序	249
12.5 归并排序	254
12.6 分配排序	255
12.6.1 箱排序	255
12.6.2 基数排序	255
12.7 各种内排序法的比较	257
12.8 知识点小结	257
习题	258

第六部分 文件的组织结构及排序

*第13章 文件	260
13.1 文件的基本概念	260
13.2 顺序文件	260
13.3 索引文件	261
13.4 索引顺序文件	262
13.5 散列文件	267
13.6 多关键字文件	268
13.7 知识点小结	268
习题	268
*第14章 外排序	269
14.1 多路平衡归并	269
14.2 置换选择排序	272
14.3 归并树及最佳归并树	274
14.4 知识点小结	275
习题	275
参考文献	276

第一部分 概 论

有用就会有需求，有了需求就会有动力，因此在开始本课程学习之前，我们必须弄清楚一个问题：**我们为什么需要学习数据结构？**

计算机领域分两个大的方向：一个是硬件方向，另一个是软件方向。硬件方向又包括两个大的方面：一个是硬件技术的发展，另一个是如何在现有硬件条件下设计出高性能的计算机系统。软件是程序、数据以及文件的集合，它大致可分为系统软件和应用软件两大类。系统软件追求的是如何在方便用户使用的前提下将计算机系统的性能发挥到极致；应用软件则是直接面对终端用户，它寻求的是更高的服务质量和更好的用户体验。我们将从应用软件开发者的视角来探寻数据结构的重要性。

实际问题往往很复杂，当我们求解实际问题的时候，往往是对实际问题的模型进行求解。所谓模型就是对实际问题的简化，是反映问题本质的数据集合以及数据之间关系的集合。对模型的求解则是找到对给定的输入数据的一系列处理步骤，使得能够得到预期的输出。为了让计算机实现我们对模型的求解思路，则必须将模型映射到存储器中，这样计算机才能根据人的意图对操作对象（即数据）进行处理。

通过上面的分析可知，计算机的各种应用实际上都是对数据的处理，因此对数据以及它们之间的关系进行分析、表示是必不可少的。

什么是结构？我们来看下面的例子：一个家族成员之间的双亲与孩子的关系构成了一个树形的家谱图；城市之间的互通关系构成了一张网状的交通图；操作系统各模块之间的单向依赖关系（即第 $i+1$ 层中的模块可以调用第 i 层中的模块，但第 i 层中的模块不能调用第 $i+1$ 层中的模块）构成了一个层次结构；长、宽、高之间的关系构成了三维空间结构；自然数集中的“ $<$ ”关系构成了一个有向无环图；人与人之间的不同社会关系构成了不同的社会结构。通过这些例子我们可以发现，结构是关系的一种表现形式，可以说结构即关系。数据结构就是数据以及数据之间的关系，它对更好地使用计算机来解决实际问题至关重要。

我们已经认识到了学习数据结构的重要性，那么我们应该从这门课程中学到什么？如何学？

我们将会从这门课程中学习几种典型的数据结构（这些数据结构基本涵盖了现实中的绝大多数关系类型），以及两种在实际问题中经常要涉及的运算（查找和排序）的各种实现算法。我们要能清晰地知道每种数据结构描述的是数据之间的一种怎样的关系，有哪些典型应用，它们在存储器中有哪几种存储表示方式，这些存储表示方式各有什么优缺点，不同数据结构之间有哪些联系和区别，两种运算有哪几种不同的实现算法，这些算法在时间开销和空间开销方面的比较，等等。学习的目的是应用，使得所学知识内化为自身的能力。因此，在学习数据结构的过程中，我们应该尝试着用所学内容去解决实际问题。首先通过对问题的观察理解分析出问题求解的对象、涉及的主要操作，然后寻找该问题与我们已有知识之间的联系，选用适当的数据结构描述问题的处理对象，最终达到解决问题的目的。

弄清楚上述问题将有助于端正学习动机，明确学习目标和找到一种有效的学习方法。

一个问题可以用不同的数据结构来描述，因此它有多种不同的处理方法，有些方法效率高，有些则不然。也就是说，我们需要针对问题找到合适的数据结构，这样才能得出求解问题的高效方法。可见，数据结构和问题的处理方法（即算法）是密不可分的。瑞士著名科学家 N. Wirth 用下述著名公式生动地描述了数据结构与算法之间的这种密切关系：程序 = 数据结构 + 算法。

下面就让我们来详细了解一下这两大主角吧！

第1章 数据结构

通过前面的介绍，我们已经对数据结构有了一个初步的印象：它是数据之间关系的描述。我们需要进一步了解的是：什么是数据，数据之间常见的关系有哪些，这些关系在存储器中如何表示，“数据结构”和我们通常所说的“数据类型”、“抽象数据类型”之间有什么区别和联系。这些问题将在本章中得到一一解答。

1.1 什么是数据

简单地说，**数据 (Data)** 是描述客观事物且能被计算机识别并加工处理的对象，它可以是数值、字符、声音、图像等。**数据元素 (Data Element)** 是数据的基本单位，是数据集合中的个体，在计算机程序中通常作为一个整体来考虑和处理。数据元素在不同的数据结构中有不同的称谓，如在树结构中称为结点，在图结构中称为顶点，在数据库表中称为记录等。它可以由一个或多个数据项组成。**数据项 (Data Item)** 是数据的最小单位，是不可再分的。因此仅由一个数据项构成的数据元素称为原子元素，由多个数据项构成的数据元素称为结构元素。**数据对象 (Data Object)** 是性质相同的数据元素的集合，它是数据的一个子集。

例如，某班级的成绩表就是一个数据对象，其中某个同学的成绩（对应于表中的一行）就是一个数据元素，构成该同学成绩的各个字段就相当于数据项。再如，整数集是一个数据对象，其中的每个整数是一个数据元素且为原子元素。

1.2 什么是数据结构

结构即关系，**数据结构 (Data Structure)** 是相互之间存在一种或多种特定关系的数据元素的集合。数据结构可以表示成一个二元组，记为： $\text{data_structure} = (D, S)$ ，其中 D 为数据元素的集合， S 是 D 上关系的集合。数据结构包括数据元素的逻辑结构、存储结构和相适应的运算三个方面的内容。

1.2.1 数据的逻辑结构

数据的逻辑结构 (Logical Structure of Data) 是指数据之间的逻辑关系。它与计算机无关，它可以作为从具体问题中抽象出来的数据模型。数据的逻辑结构通常有以下四种：集合结构、线性结构、树形结构和网状结构（也称为图状结构），如图 1-1 所示。

从图 1-1 可以看出，在集合结构中数据元素彼此之间没有直接关系，只有“属于同一集合”的联系；在线性结构中数据元素之间存在着一对一的关系；在树形结构中数据元素之间存在着一对多的关系；在网状结构中数据元素之间存在着多对多的关系。我们把描述数据元素之间一对一关系的逻辑结构称为线性结构，把描述数据元素之间一对多或多对多关系的逻辑结构称为非线性结构。

一般概念的数据结构指的是数据的逻辑结构，在不引起混淆的情况下，我们将数据的逻辑结构简称为数据结构。

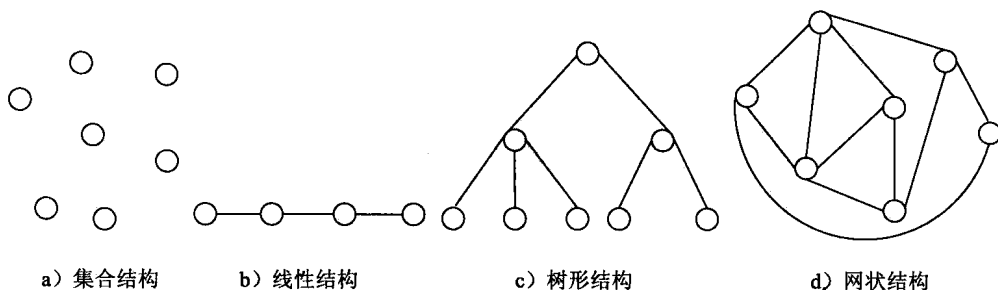


图 1-1 四种常见的数据（逻辑）结构

1.2.2 数据的存储结构

数据的存储结构（Storage Structure of Data，也称为数据的物理结构）是指数据元素及其关系在计算机存储器中的表示（映像）。其主要内容是指在存储空间中使用一个存储结点来存储一个数据元素，通过在存储空间中建立各存储结点间的关联来表示数据元素之间的逻辑关系。数据的存储结构有以下四种基本方式：

1) 顺序存储方式：在这种存储方式中所有的存储结点相继存储在连续的存储区内，用存储结点间的位置关系表示数据元素之间的逻辑关系。按照这种方法得到的存储表示称为**顺序存储结构**（Sequential Storage Structure）。

2) 链式存储方式：在这种存储方式中每一个存储结点不仅存储一个数据元素，还需要存储一个指针。该指针指向与本存储结点有逻辑关系的存储结点，即用指针来表示数据元素之间的逻辑关系。采用这种方法得到的存储表示称为**链式存储结构**（Linked Storage Structure）。

3) 索引存储方式：这种存储方式通常在存储结点信息的同时，另附设索引表。索引表中的每一项称为索引项，其一般形式为：（关键字，地址），用于指示一个存储结点或一组存储结点的存储地址。如果每一个存储结点均在索引表中有索引项，则将该索引表称为**稠密索引**（Dense Index），此时索引项记录的是相应存储结点的存储地址；如果一组存储结点在索引表中对应一个索引项，则该索引表称为**稀疏索引**（Sparse Index），此时索引项记录的是该组存储结点的存储首地址。采用这种方法得到的存储表示称为**索引存储结构**（Index Storage Structure）。

4) 散列存储方式：这种存储方法的基本思想是将存储结点关键字作为选定散列（Hash）函数的输入，得到的函数值作为存储结点的存储地址。按照这种方法得到的存储表示称为**散列存储结构**（Hash Storage Structure）。

一种逻辑结构可映像成不同的存储结构。例如，某班级的同学要到 2 号教学楼 310 教室去上数据结构这门课，该班级的同学按照学号的大小就具有一种线性关系（逻辑结构）。将 310 教室的座位从前往后、从左往右自“0”开始依次编号，从而得到一个一维的存储空间。

如果该班级的同学按照学号大小从 310 教室的第 10 号座位开始依次就座，那么就得到了该班级同学在 310 教室这个空间中的一种顺序存储映像。老师只要知道学号为 1 的同学的座位号，就可以很方便地在教室中找到任意合法学号的本班同学。

如果该班级的同学在 310 教室是任意就座的，那么逻辑上相邻（即学号相邻）的同学的座位号不一定相邻。为了存储同学之间的逻辑关系，此时每个同学需要准备一张卡片，卡片上记录了学号相邻的下一个同学的座位号，学号最大的同学的卡片上可以记录一个特殊的值（如 0 或 -1）表示他是最后一个同学。这样就得到了该班级同学在 310 教室这个空间中的一种链式存储映像。对于老师而言，他只需要知道学号为 1 的同学的座位号，通过每个同学手上的卡片就可以访问到本班的其他同学。