



普通高等教育“十一五”国家级规划教材
21世纪高等院校计算机教材

数据结构

(第二版)

刘振鹏 张小莉 郑艳娟 编著

所有算法均给出C++语言的描述，并加以详细的注释，以利于读者理解算法的基本思想



强调算法 注重实现

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE



普通高等教育“十一五”国家级规划教材
21 世纪高等院校计算机教材

数 据 结 构

(第二版)

刘振鹏 张小莉 郑艳娟 编著

特等奖内容简介“十一五”国家级规划教材

数据结构/刘振鹏 张小莉 郑艳娟编著 ISBN 978-7-113-07919-2

本书为普通高等教育“十一五”国家级规划教材，是2003年9月出版的《数据结构》的第二版。书中介绍了各种最常用的数据结构，包括线性表、栈、队列、矩阵的压缩存储、树与二叉树、图、查找、排序等。阐明各种数据结构内在的逻辑关系，讨论它们在计算机中的存储表示，以及在这些数据结构下的运算和实现的算法，并对算法的效率进行了简要的分析。本书既注重原理又重视算法的实现，均给出用 Visual C++ 语言描述的算法，并加以详细的注释，分析算法的基本思路，每章都附有大量的习题。

与本书配套的《数据结构习题解答与实验指导》详细给出了书中习题的解答思路和参考答案，并且结合数据结构课堂和实践教学，设计了7项实验内容。它和本书一起构成了一个完整的教学系列。

本书内容丰富、结构清晰、突出算法、注重应用，强调理论与实践的结合。既适合作为高等院校计算机科学与技术、通信工程、电子工程等电子信息类专业的教材，又适合于计算机爱好者自学，对于从事计算机应用和开发的技术人员也具有一定的参考价值。

图书在版编目(CIP)数据 数据结构 刘振鹏 张小莉 郑艳娟 编著

数据结构/刘振鹏，张小莉，郑艳娟编著。—2版。—北

京：中国铁道出版社，2007.4

21世纪高等院校计算机教材

ISBN 978-7-113-07919-2

I. 数… II. ①刘…②张…③郑… 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(2007)第058161号

书 名：数据结构(第二版)

作 者：刘振鹏 张小莉 郑艳娟

出版发行：中国铁道出版社(100054，北京市宣武区右安门西街8号)

策划编辑：严晓舟 秦绪好

责任编辑：翟玉峰 包 宁

特邀编辑：薛秋沛

封面制作：白雪

责任校对：黄园园

印 刷：河北省遵化市胶印厂

开 本：787×1092 1/16 印张：17 字数：397千

版 本：2007年4月第2版 2007年4月第1次印刷

印 数：1~5 000册

书 号：ISBN 978-7-113-07919-2/TP·2342

定 价：24.00元

版权所有 侵权必究

本书封面贴有中国铁道出版社激光防伪标签，无标签者不得销售

凡购买铁道版的图书，如有缺页、倒页、脱页者，请与本社计算机图书批销部调换。

第二版前言

数据结构是计算机科学与技术等电子信息类相关专业的一门重要的基础课程。许多高等院校将它作为理科各系主干基础课列入学校的教学计划之中。通过讲授本课,学生可以较全面地理解算法和数据结构的概念,掌握各种数据结构和算法的实现方式,比较不同的数据结构和算法的特点。这是一门理论与实际紧密结合的课程。通过本课程的学习,学生可以学会分析研究计算机加工的数据结构特性,以便在以后的工作实践中,能够针对具体问题选择和设计出适当的逻辑结构、存储结构及相应的算法,并在此基础上,能编写出结构清晰、正确易读、符合软件工程规范的程序,从而为进一步学习后续专业课程和软件的开发打下坚实的基础。

本书在内容组织和编排上,力求理论与实际应用紧密结合,而且更加突出运用。本书的主要特点有以下3点。

(1) 内容组织上层次分明、结构清晰。在内容的选取上坚持学以致用、学用结合的原则,集先进性、科学性和实用性于一体,尽可能地将最基础、最适用的软件技术写入教材,省略一些纯理论的推导和烦琐的数学证明。

(2) 在内容的深浅程度上,把握理论深度、侧重实用、由浅入深的原则,通过大量翔实的例题、算法和每一章的最后给出的练习题,进一步提高学生对数据的抽象能力和程序设计的能力。

(3) 内容叙述深入浅出,文体规范,文字浅显易懂,相互衔接自然,表述严谨,逻辑性强,以利于学生自学和理解。

本书作为2003年版《数据结构》的第二版,保持了前一版的基本框架,概念清楚、论述充实、面向应用,进一步完善了算法与数据结构的体系内容,对所有算法进行了详尽的注释和完善,以利于读者理解算法的基本思想。各章均安排有章节提要 and 课后习题。

本书的第1章和10章由刘振鹏编写修订,第2章~第5章由张小莉编写修订,第6章~第9章由郑艳娟编写修订,最后由刘振鹏、张小莉统一定稿。

本书在写作和修订过程中,得到了许多专家和众多院校数据结构任课教师的大力支持和帮助,他们提出了许多中肯的意见和很好的建议,对本书的编写修订起到了很大的指导作用。对此,作者表示衷心的感谢。也正是他们的认可和支持,使得本书入选普通高等教育“十一五”国家级规划教材。

感谢作者的多位同事和学生,许百成、罗文劼参与了编写大纲的讨论并编写了初稿的部分内容,王苗编写书中习题部分并提供了全部答案,石强、史青宣制作了电子讲义,赵红、苗秀芬等在使用本书的过程中指出了书中的一些不足之外,使得本书更加完善。

尽管我们做了很大的努力,但由于编者水平有限,书中难免有不妥之处,恳请读者予以指正。

编者

2007年2月

第一版前言

数据结构是计算机及相关专业的一门基础课, 计算机科学各个领域及有关的应用软件都要用到各种数据结构。在我国计算机科学与技术专业的教学计划中, 它是核心课程之一。因此, 数据结构知识对所有使用计算机的人都是必需的, 它也是学好计算机专业其他课程的基础和保证。作为计算机科学与技术专业的一门重点课程, 数据结构的教学深度、广度和对实践性环节的要求正逐步增长, 并且随着计算机应用领域的扩大, 数据结构课程已逐步对电子信息类和信息管理类的大部分专业的学生开设。本书介绍各种最常用的数据结构, 阐明各种数据结构内在的逻辑关系, 讨论它们在计算机中的存储表示, 以及在這些数据结构上的运算和实现的算法, 并对算法的效率进行了简要的分析。

本书在介绍数据结构基本理论的基础上, 尽可能多的把近年来数据结构范畴内的新成就奉献给读者。本书是笔者根据多年来教学上的实践和体会, 并参考了国内外较新的有关文献编写而成。为使全书结构避免松散, 本书既详尽地介绍了各种数据结构的理论, 又提出了具体实现的方法。对有关算法的分析, 没有繁复的数学推导和证明, 而是依照由浅入深和循序渐进的原则编排各章节, 力求使全书衔接自然、系统全面。本书共分为 10 章, 第 1 章介绍了数据结构的主要内容和基本概念、算法评价标准和评价方法。第 2 章~第 5 章介绍了几种常用的线性结构, 包括线性表、栈和队列、串以及数组和广义表。着重讨论这些结构的存储表示、运算算法, 并以查找、插入、删除运算为核心, 讨论在各种存储结构上进行相关运算的算法效率。第 6 章和第 7 章介绍了具有广泛应用价值的树形结构(二叉树和树)及其重要应用, 包括二叉树与树的概念和存储结构、二叉树的性质、二叉树和树的遍历及线索二叉树、树与二叉树的转换、哈夫曼树等。第 8 章介绍了复杂的数据结构——图及其应用, 从算法设计的角度(而不是从图论角度)介绍图的存储形式、基本运算算法和几个最优化问题。第 9 章和第 10 章介绍了查找及排序, 较全面地介绍了查找和排序的各种常用方法并简单分析了各种算法的时间复杂度和空间复杂度。

本书既注重原理又重视算法的实现, 分析了大量的实用算法, 均给出用 C++ 语言描述的算法, 并加上较详细的注释, 以利于读者理解算法的基本思想。每章之后附有习题, 以便读者进一步练习并检验学习效果。

本书作者可以提供教材的电子讲义和书中习题的答案, 有需要者可与中国铁道出版社计算机图书中心联系。

本书的第 1 章~第 5 章由张小莉编写, 第 6 章~第 8 章由刘振鹏编写, 第 9、10 章由郝杰编写, 最后由刘振鹏、张小莉统一定稿。

本书在写作过程中, 得到了许多专家的大力支持, 参考了大量的文献资料和国内外优秀的教材, 罗文劼、许百成、王苗、石强等老师在讲义的使用过程中提出了许多宝贵的意见, 陈兰芳、程瑞芬、崔仙翠等同志参与了本书的编排工作, 在此表示诚挚的谢意。

由于时间仓促及编者水平有限, 书中难免有疏漏和不妥之处, 恳请读者予以指正。

编 者

2003 年 8 月

21 世纪高等院校计算机教材

编 委 会

主 任：王凤先

副主任：马胜甫 刘振鹏

委 员：（按姓氏笔画排列）

于 锋	王卫华	王学军	王春红	白乙拉
刘立民	刘清波	李亚平	李继民	张小莉
杨怀卿	孟玉芹	赵英杰	赵新生	赵福来
徐建民	韩宪忠			

目 录

串 章 4 第

算 本 基 其 又 串 1-4

念 册 本 基 的 串 1-1-4

算 本 基 的 串 5-1-4

算 本 基 又 的 算 则 斗 宝 的 串 5-4

前 算 则 斗 宝 的 串 1-5-4

算 本 基 的 串 中 的 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

前 算 则 斗 宝 的 串 5-5-4

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

80

第 1 章 绪论..... 1

1-1 数据结构的概念..... 1

1-1-1 为什么要学习数据结构..... 1

1-1-2 相关概念和术语..... 4

1-1-3 数据结构课程的内容..... 6

1-2 数据类型和抽象数据类型..... 7

1-2-1 数据类型..... 7

1-2-2 抽象数据类型..... 8

1-3 算法和算法分析..... 9

1-3-1 算法特性..... 9

1-3-2 算法描述..... 9

1-3-3 算法性能分析与度量..... 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..... 13

2-2 线性表的顺序存储及运算实现..... 14

2-2-1 顺序表..... 14

2-2-2 顺序表上基本运算的实现..... 16

2-2-3 顺序表应用举例..... 19

2-3 线性表的链式存储和运算实现..... 22

2-3-1 单链表..... 22

2-3-2 单链表上基本运算的实现..... 24

2-3-3 循环链表..... 30

2-3-4 双向链表..... 31

2-3-5 静态链表..... 32

2-3-6 单链表应用举例..... 34

2-4 顺序表和链表的比较..... 36

习题..... 37

第 3 章 栈和队列..... 39

3-1 栈..... 39

3-1-1 栈的定义及基本运算..... 39

3-1-2 栈的存储实现和运算实现..... 40

3-2 栈的应用举例..... 43

3-3 队列..... 53

3-3-1 队列的定义及基本运算..... 53

3-3-2 队列的存储实现及运算实现..... 54

3-4 队列应用举例..... 59

习题..... 62

第4章 串	63
4-1 串及其基本运算	63
4-1-1 串的基本概念	63
4-1-2 串的基本运算	63
4-2 串的定长顺序存储及基本运算	64
4-2-1 串的定长顺序存储	64
4-2-2 定长顺序串的基本运算	65
4-2-3 模式匹配	66
4-3 串的堆存储结构	72
4-3-1 串名的存储映像	72
4-3-2 堆存储结构	74
4-3-3 基于堆结构的串的基本运算实现	74
习题	76
第5章 数组和广义表	77
5-1 多维数组	77
5-1-1 数组的逻辑结构	77
5-1-2 数组的内存映像	77
5-2 特殊矩阵的压缩存储	80
5-2-1 对称矩阵	80
5-2-2 三角矩阵	81
5-2-3 带状矩阵	82
5-3 稀疏矩阵	83
5-3-1 稀疏矩阵的三元组表存储	83
5-3-2 稀疏矩阵的十字链表存储	90
5-4 广义表	95
5-4-1 广义表的定义和基本运算	95
5-4-2 广义表的存储	96
5-4-3 广义表基本操作的实现	99
习题	102
第6章 二叉树	104
6-1 二叉树的定义与性质	104
6-1-1 二叉树的基本概念	104
6-1-2 二叉树的主要性质	106
6-2 二叉树的基本操作与存储实现	107
6-2-1 二叉树的存储	107
6-2-2 二叉树的基本操作及实现	110
6-3 二叉树的遍历	112
6-3-1 二叉树的遍历方法及递归实现	112
6-3-2 二叉树遍历的非递归实现	115
6-3-3 由遍历序列恢复二叉树	118
6-3-4 不用栈的二叉树遍历的非递归方法	120
6-4 线索二叉树	121
6-4-1 线索二叉树的定义及结构	121
6-4-2 线索二叉树的基本操作实现	123

6-5	二叉树的应用	128
6-5-1	二叉树遍历的应用	128
6-5-2	最优二叉树——哈夫曼树	131
习题		139
第7章	树与森林	141
7-1	树的概念与表示	141
7-1-1	树的定义及相关术语	141
7-1-2	树的表示	142
7-2	树的基本操作与存储	143
7-2-1	树的基本操作	143
7-2-2	树的存储结构	144
7-3	树、森林与二叉树的转换	147
7-3-1	树转换为二叉树	147
7-3-2	森林转换为二叉树	148
7-3-3	二叉树转换为树和森林	149
7-4	树和森林的遍历	150
7-4-1	树的遍历	150
7-4-2	森林的遍历	150
7-5	树的应用	151
7-5-1	判定树	151
7-5-2	集合的表示	152
7-5-3	等价问题	154
习题		155
第8章	图	157
8-1	图的基本概念	157
8-1-1	图的定义和术语	157
8-1-2	图的基本操作	160
8-2	图的存储结构	160
8-2-1	邻接矩阵	160
8-2-2	邻接表	162
8-2-3	十字链表	164
8-2-4	邻接多重表	166
8-3	图的遍历	168
8-3-1	深度优先搜索	168
8-3-2	广度优先搜索	170
8-3-3	应用图的遍历判定图的连通性	172
8-4	生成树与最小生成树	172
8-4-1	生成树和生成森林	172
8-4-2	最小生成树的概念	174
8-4-3	构造最小生成树的 Prim 算法	175
8-4-4	构造最小生成树的 Kruskal 算法	178
8-5	最短路径	180
8-5-1	从一个源点到其他各点的最短路径	180
8-5-2	每一对顶点之间的最短路径	184
8-6	有向无环图及其应用	186

8-6-1	有向无环图的概念	186
8-6-2	AOV 网与拓扑排序	187
8-6-3	AOE 图与关键路径	192
习题		196
第 9 章	查找	199
9-1	基本概念	199
9-2	静态查找表	200
9-2-1	静态查找表结构	200
9-2-2	顺序查找	200
9-2-3	有序表的查找	201
9-2-4	分块查找	205
9-3	动态查找表	206
9-3-1	二叉排序树	206
9-3-2	平衡二叉树	211
9-3-3	B 树和 B ⁺ 树	218
9-4	哈希表查找 (杂凑法)	225
9-4-1	哈希表与哈希方法	225
9-4-2	常用的哈希函数	226
9-4-3	处理冲突的方法	228
9-4-4	哈希表的查找分析	230
习题		231
第 10 章	排序	234
10-1	排序的基本概念	234
10-2	插入排序	235
10-2-1	直接插入排序	235
10-2-2	折半插入排序	236
10-2-3	表插入排序	237
10-2-4	希尔排序	239
10-3	交换排序	240
10-3-1	冒泡排序	240
10-3-2	快速排序	241
10-4	选择排序	244
10-4-1	简单选择排序	244
10-4-2	树形选择排序	244
10-4-3	堆排序	245
10-5	归并排序	248
10-6	基数排序	250
10-6-1	多关键码排序	250
10-6-2	链式基数排序	251
10-7	外排序	253
10-7-1	外部排序的方法	253
10-7-2	多路平衡归并的实现	255
习题		258
参考文献		260

第1章 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，它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或数据处理、过程控制及对文件的存储和检索及数据库技术等计算机应用，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而且效率高的程序，必须研究数据的特性、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。

本章知识要点：

- 数据结构的概念
- 数据类型和抽象数据类型
- 算法和算法分析

1-1 数据结构的概念

计算机在发展的初期，其应用范围是数值计算，所处理的数据都是整型、实型、布尔型等简单数据，以此为加工、处理对象的程序设计称为数值型程序设计。随着计算技术的发展，计算机逐渐进入到商业、制造业等其他领域，广泛地应用于数据处理和过程控制中。与此相对应，计算机所处理的数据也不再是简单的数值，而是字符串、图形、图像、语音、视频等复杂的数据。这些复杂的数据不仅量大，而且具有一定的结构。例如，一幅图像是一个由简单数值组成的矩阵，一个图形中的几何坐标可以组成表。此外，语言编译过程中所使用的栈、符号表和语法树，操作系统中用到的队列、磁盘目录树等，都是有结构的数据。数据结构所研究的就是这些有结构的数据，因此，数据结构的知识不论对研制系统软件还是开发应用软件都非常重要，它是学习软件知识和提高软件设计水平的重要基础。

数据结构是计算机科学与技术专业的基础课，是核心课程。所有的计算机系统软件和应用软件都要用到各种类型的数据结构。因此，要想更好地运用计算机来解决问题，仅掌握几种计算机程序设计语言是难以应付众多复杂课题的。要想有效地使用计算机并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。打好数据结构这门课程的扎实基础，对于学习计算机专业的其他课程，如操作系统、编译原理、数据库原理、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。当使用计算机来解决一个具体问题时，一般需要经过如下几个步骤：首先要从该具体问题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，得到最终的解答。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。

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是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问题占用了 90% 以上的机器时间。这类问题涉及到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来描述。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才能有效地解决问题。下面所列举的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时，或者想查询某个专业或年级的学生的有关情况时，只要建立了相关的数据结构，按照某种算法编写了相关程序，就可以实现计算机自动检索。为此，可以在学生信息检索系统中建立一张按学号顺序排列的学生信息表和若干张分别按姓名、专业、年级顺序排列的索引表，如表 1-1~表 1-4 所示。由这 4 张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。

表 1-1 学生信息表

学 号	姓 名	性 别	专 业	年 级
030101	崔永志	男	计算机科学与技术	2003 级
030205	刘淑芳	女	信息与计算科学	2003 级
040310	李 丽	女	数学与应用数学	2004 级
040212	张志会	男	信息与计算科学	2004 级
050123	贾宝国	男	计算机科学与技术	2005 级
050101	陆文颖	女	计算机科学与技术	2005 级
050332	石胜利	男	数学与应用数学	2005 级
050231	崔文靖	男	信息与计算科学	2005 级
060135	刘淑芳	女	计算机科学与技术	2006 级
060314	史文斌	男	数学与应用数学	2006 级

表 1-2 姓名索引表

姓 名	索 引 号	姓 名	索 引 号	姓 名	索 引 号
崔文靖	8	崔永志	1	李丽	3
刘淑芳	2, 9	陆文颖	6	贾宝国	5
石胜利	7	史文斌	10	张志会	4

表 1-3 专业索引表

专 业	索 引 号
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

表 1-4 年级检索表

年 级	索 引 号	年 级	索 引 号
2003 级	1, 2	2004 级	3, 4
2005 级	5, 6, 7, 8	2006 级	9, 10

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的是一种简单的线性关系，这类数学模型可称为线性的数据结构。

【例 1-2】八皇后问题。在八皇后问题中，处理过程不是根据某种确定的计算法则，而是利用试探和回溯的探索技术求解。为了求得合理布局，在计算机中要存储布局的当前状态。从最初的布局状态开始，一步步地进行试探，每试探一步形成一个新的状态，整个试探过程形成了一棵隐含的状态树，如图 1-1 所示（为了描述方便，将八皇后问题简化为四皇后问题）。回溯法求解过程实质上就是一个遍历状态树的过程。在这个问题中所出现的树也是一种数据结构，它可以应用在许多非数值计算问题中。

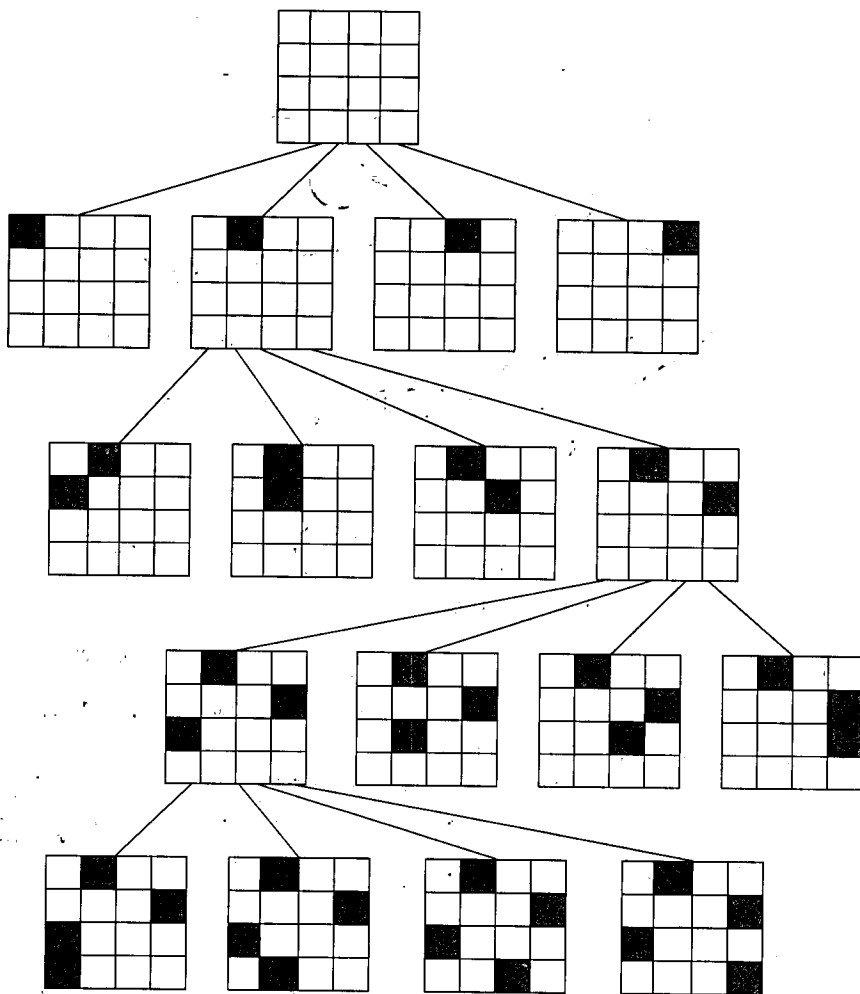


图 1-1 四皇后问题中隐含的状态树

【例 1-3】教学计划编排问题。一个教学计划包含许多课程，在这些课程之间，有些必须按规定的先后次序进行，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，如表 1-5 所示。这种各个课程之间的次序关系可用一个称为图的数据结构来表示，如图 1-2 所示。有向图中的每个顶点表示一门课程，如果从顶点 C_i 到 C_j 之间存在有向边 $\langle C_i, C_j \rangle$ ，则表示课程 i 必须先于课程 j 进行。

表 1-5 计算机专业的课程设置

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

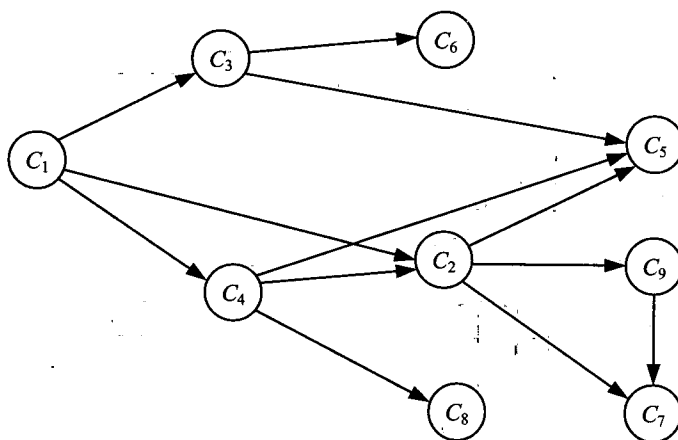


图 1-2 教学计划编排问题的数据结构

由以上 3 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。

学习数据结构的目的是为了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，通过算法训练来提高学生的思维能力，通过程序设计的技能训练来提高学生的综合应用能力和专业素质。

1-1-2 相关概念和术语

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含义。

(1) 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, 应用程序处理各种各样的数据。计算机科学中, 所谓数据就是计算机加工处理的对象, 它可以是数值数据, 也可以是非数值数据。数值数据是一些整数、实数或复数, 主要用于工程计算、科学计算和商务处理等; 非数值数据包括字符、文字、图形、图像、语音等。

(2) 数据项 (Data Item, 也称项或字段): 是具有独立含义的标识单位, 是数据不可分割的最小单位, 如表 1-1 中的“学号”、“姓名”、“年级”等。数据项有名和值之分, 数据项名是一个数据项的标识; 用变量定义, 而数据项值是它的一个可能取值, 表 1-1 中的 050332 是数据项“学号”的一个取值。数据项具有一定的类型, 依数据项的取值类型而定。

(3) 数据元素 (Data Element): 是数据的基本单位。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、八皇后问题中状态树的一个状态、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。

有时, 一个数据元素可由若干个数据项组成。例如, 学生信息检索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、专业和年级数据项。这些数据项可以分为两种: 一种叫做初等项, 如学生的性别、年级等, 这些数据项是在数据处理时不能再分割的最小单位; 另一种叫做组合项, 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, 在解决实际问题时是把每个学生记录当作一个基本单位进行访问和处理的。

(4) 数据对象 (Data Object) 或数据元素类 (Data Element Class): 具有相同性质的数据元素的集合。在某个具体问题中, 数据元素都具有相同的性质 (元素值不一定相等), 属于同一数据对象 (数据元素类), 数据元素是数据元素类的一个实例。例如, 在交通咨询系统的交通网中, 所有的顶点是一个数据元素类, 顶点 A 和顶点 B 各自代表一个城市, 是该数据元素类中的两个实例, 其数据元素的值分别为 A 和 B 。

(5) 数据结构 (Data Structure): 是指互相之间存在着一种或多种特定关系的数据元素的集合。在任何问题中, 数据元素都不会是孤立的, 它们之间都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。根据数据元素间关系的不同特性, 通常有下列 4 种基本的结构, 如图 1-3 所示。

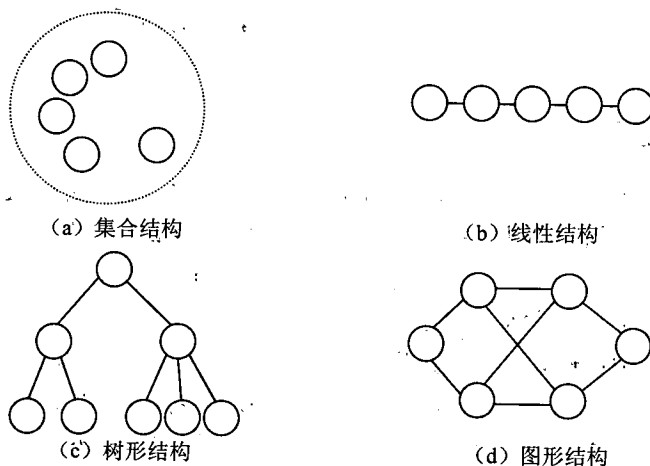


图 1-3 4 类基本结构的示意图

- 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”。集合是元素关系极为松散的一种结构。
- 线性结构。该结构的数据元素之间存在着一对一的关系。
- 树形结构。该结构的数据元素之间存在着一对多的关系。
- 图形结构。该结构的数据元素之间存在多对多的关系;图形结构也称为网状结构。

由于集合是数据元素之间关系极为松散的一种结构,因此也可用其他结构来表示它。从上面所介绍的数据结构的概念中可以知道,一个数据结构包含两个要素:一个是数据元素的集合,另一个是关系的集合。在形式上,数据结构通常可以采用一个二元组来表示。

数据结构的定义为一个二元组:

$\text{Data_Structure} = (D, R)$

其中, D 是数据元素的有限集, R 是 D 上关系的有限集。

数据结构包括数据的逻辑结构和数据的物理结构。数据的逻辑结构可以看作是从具体问题抽象出来的数学模型,它与数据的存储无关。研究数据结构的目的是为了在计算机中实现对它的操作;为此还需要研究如何在计算机中表示一个数据结构。数据结构在计算机中的标识(又称为映像)称为数据的物理结构,或称为存储结构。它所研究的是数据结构在计算机中的实现方法,包括数据结构中元素的表示及元素间关系的表示。

数据的存储结构可采用顺序存储或链式存储的方法。

顺序存储方法是把逻辑上相邻的元素存储在物理位置相邻的存储单元中,由此得到的存储表示称为顺序存储结构。顺序存储结构是一种最基本的存储表示方法,通常借助于程序设计语言中的数组来实现。

链式存储方法对逻辑上相邻的元素不要求其物理位置相邻,元素间的逻辑关系通过附设的指针字段来表示,由此得到的存储表示称为链式存储结构。链式存储结构通常借助于程序设计语言中的指针来实现。

除了通常采用的顺序存储方法和链式存储方法外,有时为了查找的方便还采用索引存储方法和散列存储方法。

1-1-3 数据结构课程的内容

数据结构与数学、计算机硬件和软件有十分密切的关系。数据结构是介于数学、计算机硬件和计算机软件之间的一门计算机科学与技术专业的核心课程,是高级程序设计语言、编译原理、操作系统、数据库、人工智能等课程的基础。同时,数据结构技术也广泛应用于信息科学、系统工程、应用数学及各种工程技术领域。

数据结构课程集中讨论软件开发过程中的设计阶段,同时涉及编码和分析阶段的若干基本问题。此外,为了构造出好的数据结构及其实现,还需考虑数据结构及其实现的评价与选择。因此,数据结构的内容包括3个层次和5个“要素”,如表 1-6 所示。

表 1-6 数据结构课程内容体系

方面 层次	数 据 表 示	数 据 处 理
抽象	逻辑结构	基本运算
实现	存储结构	算法
评价	不同数据结构的比较及算法分析	

数据结构的核心技术是分解与抽象。通过对问题的抽象,舍弃数据元素的具体内容,就得到逻辑结构;类似地,通过分解将处理要求划分成各种功能,再通过抽象舍弃实现细节,就得到运算的定义。上述两个方面的结合使人们将问题转换为数据结构,这是一个从具体(即具体问题)到抽象(即数据结构)的过程。然后,通过增加对实现细节的考虑进一步得到存储结构和实现运算,从而完成设计任务。这是一个从抽象(即数据结构)到具体(即具体实现)的过程。熟练地掌握这两个过程是数据结构课程在专业技能培养方面的基本目标。

数据结构作为一门独立的课程在国外是从1968年才开始的,但在此之前其有关内容已放在编译原理及操作系统课程之中。20世纪60年代中期,美国的一些大学开始设立有关课程,但当时的课程名称并不叫数据结构。1968年美国唐·欧·克努特(Donald E.knuth)教授开创了数据结构的最初体系,他所著的《计算机程序设计技巧》(Art of Computer Programming,现翻译为《计算机程序设计艺术》)第一卷《基本算法》是第一部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。从20世纪60年代末到70年代初,出现了大型程序,软件也相对独立,结构程序设计成为程序设计方法学的主要内容,人们越来越重视数据结构。从20世纪70年代中期到80年代,各种版本的数据结构著作相继问世。目前,数据结构的发展并未终结,一方面,面向各专门领域中特殊问题的数据结构得到研究和发展,例如,多维图形数据结构等;另一方面,从抽象数据类型和面向对象的观点来讨论数据结构已成为一种新的趋势,越来越受到人们的重视。

1-2 数据类型和抽象数据类型

对于一个复杂问题,往往会涉及到许多因素。为了使问题得到简化,以便建立相应的数学模型进行深入研究,并进而加以解决,通常采用“抽象”这一思想方法,抽取反映问题本质的东西,舍去其非本质的细节。

在计算机软件的发展过程中,“抽象”这一思想方法将到了充分的应用。回顾一下程序设计发展的几个阶段:二进制的机器指令;符号化的汇编语句;高级语言的执行语句;高级语言的过程(或函数)模块;面向对象的软件开发系统。在这些阶段中,后一阶段都是在前一阶段的基础上经过进一步抽象后得以建立起来的。

通过一步步抽象,不断地突出“做什么?”,而将“怎么做?”隐藏起来,即将一切用户不必了解的细节封装起来,从而简化了问题。所以,抽象是程序设计最基本的思想方法。

下面回顾一下在程序设计语言中出现的各种数据类型。

1-2-1 数据类型

数据类型是和数据结构密切相关的一个概念。它最早出现在高级程序设计语言中,用以刻画程序中操作对象的特性。在用高级语言编写的程序中,每个变量、常量或表达式都有一