



浙江省重点教材建设项目

数据结构与算法

瞿有甜 编著



清华大学出版社

21世纪高等学校规划教材 | 计算机类

数据结构与算法

瞿有甜 编著

清华大学出版社

内 容 简 介

本书以数据结构基础和算法设计方法为知识单元,系统地介绍了数据结构与算法的基本知识及应用,简明扼要地阐释了计算机算法的设计与分析方法。本书的主要内容包括线性表、树、图等基础数据结构,同时也包括一些实用性较强的算法及高级数据结构,如并查集、伸展树等。以经典问题算法为例,书中分类介绍了算法设计方法以及查找与排序算法等。编者结合 ACM 国际大学生程序设计竞赛的需求,对各章节知识的灵活应用进行了详细的分析,用丰富的实例帮助读者由浅入深、快速地掌握算法设计的技巧,提升算法设计能力。本书的算法全部采用 C 语言描述,且在 Dev-C++ 中测试通过,习题安排上注重能力培养和实战训练的需求。

本书注重理论与实践相结合,内容深入浅出,可以作为高等院校计算机相关专业的教材或参考书,同时也可供 ACM 竞赛的兴趣爱好者和有关工程技术人员参考。

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 13701121933

图书在版编目(CIP)数据

数据结构与算法/瞿有甜编著.--北京:清华大学出版社,2015

21 世纪高等学校规划教材·计算机科学与技术

ISBN 978-7-302-38396-3

I. ①数… II. ①瞿… III. ①数据结构—高等学校—教材 ②算法分析—高等学校—教材
IV. ①TP311.12

中国版本图书馆 CIP 数据核字(2014)第 256982 号



责任编辑:刘 星 薛 阳

封面设计:傅瑞学

责任校对:时翠兰

责任印制:何 芊

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 刷 者:三河市君旺印务有限公司

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:26.75

字 数:648 千字

版 次:2015 年 4 月第 1 版

印 次:2015 年 4 月第 1 次印刷

印 数:1~2000

定 价:49.50 元

产品编号:061117-01

本书以教育部计算机专业数据结构课程大纲为基础,结合作者对该课程教学近 30 年的体会,以及指导一个省属普通高校学生队参加 ACM 国际大学生程序设计竞赛的经验来组织和安排教材内容。

本书的主要内容包括基础的算法分析(大 O 阶)方法,线性结构中的顺序结构与链式结构,栈与队列的应用,串的 KMP 模式匹配算法;树、二叉树的常见遍历算法,线索二叉树、哈夫曼树;图的深度优先、广度优先遍历算法,最小生成树、最短路径、拓扑排序、关键路径算法;常用的查找和排序算法,散列表技术等。同时,从数据结构课程知识实际应用的角度,用相对比较多的篇幅,以经典算法应用为例重点介绍了常用的一些算法设计方法,例如分治法、贪心法、回溯与分支限界法、动态规划法等。

对于普通高等院校的学生来说,虽然大都能认识到数据结构课程的重要性,并有意要学好它,但事与愿违的事比比皆是。事实上有许多学生都会觉得数据结构比较抽象,上课时似乎听明白了,但实践算法时却总觉得无从下手,从而产生数据结构课程很难、枯燥乏味等不正确的观点和想法。其实数据结构中的很多算法都是人类智慧的结晶,通过它去感受计算机编程的魅力是一种非常愉悦的精神感受,这也正是许多 ACMer 对此乐不思蜀、欲罢不能的原因。

现如今各种数据结构教材、各类获奖教材、十一五、十二五规划教材可以说是层出不穷,但真正适合一个课程初学者自学和独立实践的教材并不多。为此,在本教材编写中,作者尽量摒弃传统数据结构教材“重理论思想而轻代码实现”的做法,关注一个初学数据结构和算法的自学者的感受,在题例中尽量给出完整的算法分析和代码实现。

作者是数据结构与算法课程浙江省省级精品课程负责人,也是教育部微软精品课程负责人,在执掌浙江师范大学 ACM 竞赛队教练期间,曾将一个几乎是零起点的学校 ACM 竞赛队,逐年提升竞技水平,在短短 4 年的时间里,实现了亚洲赛金奖的突破,并顺利取得了参加全球总决赛的资格。期望通过本教材的编写,对长期在这条战线上默默耕耘着的普通高等院校的教练们和 ACM 竞赛的爱好者们能有所启迪、有所帮助。

本书的题例和习题均以编程为主。作者认为理解概念还是相对容易的,理解算法思想并付诸于实践以解决实际问题才是本课程学习的真正目的。为增加课程学习的趣味性,提升学习者的实际编程能力或竞赛实战能力,本书的大多数题例来自北京大学和浙江大学的在线测试系统,同时也以这两个测试系统为主选择了课后练习实践题,以下是可供学习者练习的若干在线测试系统地址:

在线测试系统	简称	网 址
北京大学在线评测系统	POJ	http://poj.org/
浙江大学在线评测系统	ZOJ	http://acm.zju.edu.cn/onlinejudge/
UVA 在线评测系统	UVA	http://uva.onlinejudge.org/ http://livearchive.onlinejudge.org/

在线测试系统	简称	网 址
Ural 在线评测系统	Ural	http://acm.timus.ru/
SGU 在线评测系统	SGU	http://acm.sgu.ru/

上述各在线测试系统中均提供了丰富的测试练习题目,一个普通的学习者应该从中选择一些合适的题目进行练习。虽然说编程能力是听不会、看不会,而是靠练会的,但一个人的精力毕竟是有限的,尤其是一个普通高校而又在中学阶段没有经过任何信息学竞赛训练的学习者来说,题海战术显然是不合适,这就要求读者在研读本书时,在具备熟练的代码能力的基础上,还要善于分析总结、善于分类归纳,善于对题目做拓展、变化,真正做到举一反三、触类旁通,以求收到事半功倍的学习效果。作为课程授课的老师,引导学生多做角色换位思考训练,讲解问题点到为止,留出足够的空间给学习者自己去思考、去求解未尝不是一种好的训练方法。

学习数据结构与算法的另外一个目的是要学会分析判断最优化方法,竞赛中出现的题目往往不止一种求解方法,所以学习者首先应该学会分析某问题可以有多少种不同的求解方法,同时针对给定的问题和条件迅速判别出最优的求解方法。当然,此事说起来容易,做起来难,更何况算法和代码的优化是一个高智慧的体力活。限于作者自身的水平和能力,书中给出的算法并不一定能保证是最优的,也不一定能绝对保证是最合适的,从某种程度上说算法优化是无止境的,今天认为是最好的算法,也许明天就不是最好的了。限于篇幅,作者在介绍题例的时候一般也不对多种解法做分析,而是按照常规的理解来求解。关于算法设计的奥秘与乐趣读者可以从绪论中的求最大子序列和问题的求解中慢慢品味。

本书成稿周期颇长,从首次与清华大学出版社签约至今已有整整6年时间,期间也曾几次想放弃过,但每每想起许多挚友十多年来一直为之而默默地奋斗着,又总觉得自己应该来抛砖引玉,本书就是在这样的多次纠结中完稿的。

本书适合想从事编程工作的各类技术人员,以及就读于与计算机编程相关的各类大、中专相关专业的学生,以及工作后需要补学或温习数据结构和算法的各类读者。

本书诸多题例的代码来自众多 ACMer 的博客,而并非选自我的队员的代码,基本的想法也是为了使算法思想和代码实现更具代表性,当然,作者也对一些代码做了适当的修改。正如前面所述,由于成书周期过长,在代码引用中未一一列出各位 ACMer 的博客地址,还请见谅!限于作者的水平,书中的错误也恳请读者批评指正,有兴趣的读者请发送邮件到 liux@tup.tsinghua.edu.cn。

感谢清华大学出版社的多位朋友为本书的出版所做的努力,同时也感谢我的研究生王超男、钟丽丽、盛贤良,曾今的 ACM 弟子陈聪、诸一行等为本书的出版所做出的贡献。

《数据结构与算法》被列为浙江省重点建设教材。

作 者

于杭州下沙高教园区

2014 年 7 月

出版说明

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程”(简称“质量工程”),通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

为了深入贯彻落实教育部《关于加强高等学校本科教学工作,提高教学质量的若干意见》精神,紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”,在有关专家、教授的倡议和有关部门的大力支持下,我们组织并成立了“清华大学出版社教材编审委员会”(以下简称“编委会”),旨在配合教育部制定精品课程教材的出版规划,讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师,其中许多教师为各校相关院、系主管教学的院长或系主任。

按照教育部的要求,“编委会”一致认为,精品课程的建设工作从开始就要坚持高标准、严要求,处于一个比较高的起点上。精品课程教材应该能够反映各高校教学改革与课程建设的需要,要有特色风格、有创新性(新体系、新内容、新手段、新思路,教材的内容体系有较高的科学创新、技术创新和理念创新的含量)、先进性(对原有的学科体系有实质性的改革和发展,顺应并符合21世纪教学发展的规律,代表并引领课程发展的趋势和方向)、示范性(教材所体现的课程体系具有较广泛的辐射性和示范性)和一定的前瞻性。教材由个人申报或各校推荐(通过所在高校的“编委会”成员推荐),经“编委会”认真评审,最后由清华大学出版

社审定出版。

目前,针对计算机类和电子信息类相关专业成立了两个“编委会”,即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。推出的特色精品教材包括:

(1) 21 世纪高等学校规划教材·计算机应用——高等学校各类专业,特别是非计算机专业的计算机应用类教材。

(2) 21 世纪高等学校规划教材·计算机科学与技术——高等学校计算机相关专业的教材。

(3) 21 世纪高等学校规划教材·电子信息——高等学校电子信息相关专业的教材。

(4) 21 世纪高等学校规划教材·软件工程——高等学校软件工程相关专业的教材。

(5) 21 世纪高等学校规划教材·信息管理与信息系统。

(6) 21 世纪高等学校规划教材·财经管理与应用。

(7) 21 世纪高等学校规划教材·电子商务。

(8) 21 世纪高等学校规划教材·物联网。

清华大学出版社经过三十多年的努力,在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌,为我国的高等教育事业做出了重要贡献。清华版教材形成了技术准确、内容严谨的独特风格,这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会

联系人:魏江江

E-mail: weijj@tup.tsinghua.edu.cn

目 录

第 1 章 绪论	1
1.1 数据结构简介	1
1.1.1 数据结构课程的内容	1
1.1.2 数据结构的基本概念和术语	1
1.2 抽象数据类型	4
1.3 算法的执行效率及其度量	6
1.3.1 算法	6
1.3.2 算法设计的要求	7
1.3.3 算法效率的度量	8
1.3.4 算法的存储空间需求	11
1.4 算法分析	11
1.4.1 算法设计与分析的重要性	11
1.4.2 一个简单的算法分析设计实例	13
习题	17
ACM/ICPC 实战练习	19
第 2 章 线性结构	20
2.1 线性表	20
2.1.1 线性表的定义	20
2.1.2 线性表的抽象数据类型	20
2.1.3 线性表的存储结构	21
2.2 线性表的顺序存储及运算实现	22
2.2.1 顺序表	22
2.2.2 顺序表上基本运算的实现	23
2.2.3 顺序表应用举例	25
2.3 线性表的链式存储和运算实现	26
2.3.1 单链表	27
2.3.2 单链表上基本运算的实现	28
2.3.3 循环链表	33
2.3.4 双向链表	34
2.3.5 静态链表	35
2.3.6 单链表应用举例	36

2.3.7	线性表实现方法比较	39
2.4	栈	40
2.4.1	顺序栈	41
2.4.2	链式栈	42
2.4.3	栈的应用举例	43
2.4.4	栈与递归	51
2.5	队列	52
2.5.1	顺序队列	52
2.5.2	链式队列	55
2.5.3	基于队列的算法设计实例	56
2.6	数组	59
2.6.1	数组的定义	59
2.6.2	数组的顺序表示和实现	59
2.6.3	特殊矩阵的压缩存储	60
	习题	68
	ACM/ICPC 实战练习	69
第 3 章	字符串	70
3.1	串类型定义	70
3.2	串的实现	71
3.2.1	串的定长顺序存储结构及其基本运算实现	71
3.2.2	串的堆存储结构及其基本运算实现	73
3.2.3	串的链式存储结构及其基本运算实现	75
3.3	串的模式匹配算法	77
3.3.1	朴素匹配算法	78
3.3.2	KMP 算法	79
3.3.3	基于 KMP 算法的应用举例	82
	习题	84
	ACM/ICPC 实战练习	84
第 4 章	树和二叉树	86
4.1	树	86
4.1.1	树的定义和基本术语	86
4.1.2	树的抽象数据类型	87
4.1.3	树的存储结构	88
4.1.4	树的遍历	89
4.1.5	树的应用	89
4.2	二叉树	98
4.2.1	二叉树的定义	98

4.2.2	二叉树的性质	99
4.2.3	二叉树的存储结构	100
4.2.4	表达式树	102
4.2.5	二叉树的基本操作及实现	104
4.3	遍历二叉树和线索二叉树	106
4.3.1	遍历二叉树	106
4.3.2	二叉树遍历的非递归实现	108
4.3.3	线索二叉树	112
4.4	树、森林和二叉树的转换	118
4.4.1	树转换为二叉树	118
4.4.2	森林转换为二叉树	119
4.4.3	二叉树转换为树和森林	120
4.5	Huffman 编码树	121
4.5.1	最优二叉树(Huffman 树)	121
4.5.2	Huffman 编码	124
4.6	二叉搜索树	126
4.6.1	二叉搜索树的基本操作	126
4.6.2	平衡二叉树(AVL 树)	131
4.7	伸展树	136
4.7.1	伸展树的基本操作	136
4.7.2	伸展树的参考例程	138
4.8	堆与优先队列	141
4.8.1	堆的逻辑定义	141
4.8.2	堆的性质	142
4.8.3	堆的基本操作	142
4.8.4	堆的实现例程	142
4.9	B-树和 B ⁺ 树	144
4.9.1	B-树及其查找	144
4.9.2	B-树的插入和删除	146
4.9.3	B ⁺ 树	149
4.10	树型结构搜索算法应用案例	149
4.10.1	基于二叉树遍历的应用	149
4.10.2	ACM/ICPC 竞赛题例分析	152
	习题	158
	ACM/ICPC 实战练习	159
第 5 章	图论算法	161
5.1	图	161
5.1.1	图的定义和术语	161

5.1.2	图的抽象数据类型	164
5.1.3	图的存储结构	164
5.2	图的遍历算法	171
5.2.1	深度优先搜索	172
5.2.2	广度优先搜索	173
5.2.3	深度与广度优先搜索应用	175
5.3	图的连通性	182
5.3.1	无向图的连通性	183
5.3.2	有向图的连通性	183
5.3.3	生成树和生成森林	184
5.3.4	关节点和重连通分量	186
5.3.5	有向图的强连通分量	188
5.4	有向无环图及其应用	194
5.4.1	有向无环图的概念	194
5.4.2	AOV 网与拓扑排序	195
5.4.3	AOE 网与关键路径	199
5.5	最短路径算法	203
5.5.1	无权最短路径	204
5.5.2	Dijkstra 算法	205
5.5.3	具有负边值的图	209
5.5.4	所有点对的最短路径	212
5.6	最小支撑树	216
5.6.1	Prim 算法	217
5.6.2	Kruskal 算法	219
5.6.3	最小生成树算法应用	222
5.7	网络流问题	223
5.7.1	网络流的最大流问题	224
5.7.2	网络流应用	229
	习题	232
	ACM/ICPC 实战练习	235
第 6 章	内部排序	236
6.1	概述	236
6.2	基于顺序比较的简单排序算法	236
6.2.1	插入排序(Insertion Sort)	236
6.2.2	冒泡排序(Bubble Sort)	241
6.2.3	直接选择排序(Directly Selection Sort)	242
6.2.4	简单排序算法的时间代价对比	242
6.3	缩小增量排序方法——Shell 排序(Shell Sort)	243

6.4	基于分治策略的排序	244
6.4.1	快速排序(Quick Sort)	244
6.4.2	归并排序(Merge Sort)	247
6.5	树形排序方法	249
6.5.1	堆排序(Heap Sort)	249
6.5.2	树形选择排序(Tree Select Sort)	252
6.6	分配排序和基数排序	253
6.6.1	桶式排序(Bucket Sort)	253
6.6.2	基数排序(Radix Sort)	253
6.7	内部排序问题讨论与分析	257
6.7.1	常用排序算法性能简要分析	257
6.7.2	排序问题的下限	257
6.8	排序应用举例	258
	习题	262
	ACM/ICPC 实战练习	264
第 7 章	文件管理和外排序	265
7.1	外存储器	265
7.1.1	磁盘	265
7.1.2	磁盘访问时间估算	267
7.2	外存文件的组织	268
7.2.1	文件组织	268
7.2.2	文件上的操作	270
7.2.3	C 语言中的文件流操作	270
7.3	缓冲区和缓冲池	272
7.4	外排序(External Sort)	273
7.4.1	二路外排序	274
7.4.2	多路平衡归并的实现	275
7.5	置换-选择排序	279
7.6	最佳归并树	283
	习题	284
第 8 章	检索与散列表	286
8.1	检索的基本概念	286
8.2	基于线性表的检索	288
8.2.1	顺序检索	288
8.2.2	有序表的二分检索	289
8.2.3	有序表的插值查找和斐波那契查找	293
8.2.4	分块检索	294

8.3	集合的检索	295
8.3.1	集合的数学特性	295
8.3.2	计算机中的集合	296
8.4	键树	298
8.4.1	基本概念	298
8.4.2	键树的存储表示	298
8.4.3	键树相关算法实现	299
8.5	散列方法及其检索	301
8.5.1	散列函数	303
8.5.2	开散列方法(分离链接法)	307
8.5.3	开放定址法	308
8.5.4	散列方法的效率分析	310
8.6	散列表及检索的应用	312
	习题	318
	ACM/ICPC 实战练习	319
第9章	算法设计方法	321
9.1	算法设计方法概论	321
9.1.1	常用的算法设计方法	321
9.1.2	算法设计中的最优化问题	322
9.2	分治法	324
9.2.1	分治策略的思想	324
9.2.2	高精度计算	326
9.2.3	矩阵相乘的 Strassen 算法	334
9.2.4	选择问题的分治算法	335
9.2.5	分治法求最近点对问题	336
9.2.6	分治法的应用	338
9.3	贪心法	347
9.3.1	贪心算法的思想	347
9.3.2	活动安排问题	348
9.3.3	背包问题	350
9.3.4	多机调度问题的近似算法	352
9.3.5	任务调度问题	353
9.3.6	贪心算法的应用	355
9.4	回溯与分支限界	362
9.4.1	回溯与分支限界简介	362
9.4.2	八皇后问题	364
9.4.3	0-1 背包问题的回溯算法	366
9.4.4	分支限界算法	368

9.4.5 回溯与分支限界应用	370
9.5 动态规划法	377
9.5.1 Fibonacci 数的计算	377
9.5.2 矩阵连乘的顺序问题	379
9.5.3 适合动态规划算法的两个条件	384
9.5.4 最长不降子序列问题	390
9.5.5 最长公共子序列问题	393
9.5.6 石子合并问题	399
9.5.7 凸多边形最优三角剖分	405
习题	408
ACM/ICPC 实战练习	409
参考文献	411

第1章

绪论

信息的表示和处理是计算机科学的基础。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基本对象。无论是用计算机进行数值计算还是非数值计算,都是对数据进行加工处理的过程。因此,要设计出一个结构良好且高效率的程序,必须研究数据的特性、数据间的相互关系及数据在计算机中的存储表示,并利用这些特性和关系设计出相应的算法和程序。研究数据结构与算法以高效的支持程序的实现是计算机科学中的一个核心基础理论和技术问题。

1.1 数据结构简介

1.1.1 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。数据结构与算法是计算机科学与技术专业的一门核心基础课程,同时也是许多非计算机专业的重要选修课程。数据结构与算法是计算机专业相关后续课程学习的基础,它不仅直接影响到许多后续专业课程(如编译原理、操作系统、数据库、人工智能等)的学习,而且直接影响到学习者计算机应用开发能力和水平的提高。数据结构与算法及其相关技术已被广泛地应用于信息科学、系统工程、应用数学以及各种工程技术应用领域。在国际大学生程序设计竞赛(ACM International Collegiate Programming Contest, ACM/ICPC)、全国或国际青少年信息学奥林匹克(NOI/IOI)竞赛中,数据结构与算法是该类竞赛考核的核心基础内容。

用计算机进行实际应用问题求解的一般过程是分析问题、建立数学模型(或计算模型)、设计或选择算法、编程与调试、测试验证。即通过对具体问题的分析,抽象出能反映问题本质的、恰当的可求解数学模型,设计或选择解此数学模型的相关算法,完成代码程序的编写和测试验证。数据结构与算法课程集中讨论软件开发全过程的诸如分析、设计、设计实现等阶段的若干基本问题,讨论问题相关数据对象的逻辑结构及关系,研究其在计算机中的存储表示,定义相关的操作及其实现。

1.1.2 数据结构的基本概念和术语

(1) **数据(Data)**是信息的载体,是对客观事物的符号化表示,在计算机科学中是指所有能输入到计算机中并被计算机程序处理的符号的总称,它能够被计算机识别、存储和加工处

理。数据是计算机加工处理的对象,它可以是数值型数据,也可以是非数值型数据。数据的含义极为广泛,它包含了人们日常生活中方方面面的各种信息在计算机内的表示。例如整数、实数或复数等用于工程计算、科学计算和商务处理的数值型数据,字符、文字、图形、图像、语音等用于事务处理、多媒体信息处理的非数值型数据等。

(2) **数据元素**(Data Element)是数据的基本单位。在数据结构课程中,数据元素习惯上又常被称为元素、结点、顶点、记录等。例如,学籍管理系统中的一个学生基本信息记录、八皇后问题求解状态树中的一个状态、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。

(3) **数据项**(Data Item)或**域**(Domain)是数据的最小单位。一个数据元素通常可由若干个数据项组成,例如,组成学籍管理系统中学生基本信息的学号、姓名、性别、籍贯、出生年月、就读专业、年级、成绩等均为数据项。可见,数据的基本单位是数据元素,数据的最小单位即具有独立含义的最小标识单位是数据项。数据项是数据元素的成员,数据项又可分为简单数据项(Simple Data Item)和复杂数据项(Aggregate Data Item)。简单数据项是不包含子结构的数据项,例如学号、姓名等,而复杂数据项是包含子结构的数据项,例如成绩可能包括英语、高等数学、线性代数、概率论、程序设计等多门课程的成绩。

(4) **数据对象**(Data Object)或**数据元素类**(Data Element Class)是同质的数据元素的集合,是数据的一个子集。同质是指数据元素都具有相同的性质,但数据元素的值不一定相等。数据元素是数据对象或数据元素类的一个实例。例如,城市交通网是一个数据对象或数据元素类,它是由组成该交通网的所有城市(顶点)的集合组成。顶点 A 和顶点 B 分别代表一个城市,是该数据对象的两个实例,其值分别为 A 和 B。

(5) **数据结构**(Data Structure)通常包括数据元素之间的逻辑关系(逻辑结构)、数据元素在计算机内的存储表示(存储结构)和定义在这种结构之上的一组操作(运算)三个方面。数据的逻辑结构是从具体应用问题中抽象出来的数学模型,是对现实世界中某个特定领域知识或概念的抽象。数据的存储结构是指充分利用计算机主存储器的顺序及随机访问特点完成从逻辑结构到存储空间的一种映射。一般而言,数据元素之间都不会是孤立的,它们之间总是存在着这样或那样的关系,数据元素之间存在的这种关系称为结构。为方便起见,人们常将数据元素间的逻辑关系结构简称为数据结构。根据数据元素间关系的不同特性,通常有下列 4 类基本的数据结构。

① 线性结构(1:1): 结构中的数据元素之间存在着一对一的关系。

② 树形结构(1:N): 结构中的数据元素之间存在着一对多的关系。

③ 图形结构(M:N): 结构中的数据元素之间存在着多对多的关系,图形结构也称为网状结构。

④ 集合结构: 集合结构中的数据元素之间的关系仅仅是“属于同一个集合”。由于集合中的元素关系极为松散,该结构在数据结构课程中基本不讨论。

图 1.1 为表示上述 4 类基本结构的示意图。

从上述数据结构的简单定义可知,数据结构有两个基本组成要素:数据元素的集合和关系的集合。因此,在形式上数据结构是一个二元组,其形式定义是:

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

其中, D 是数据元素的有限集, R 是 D 上关系的有限集。

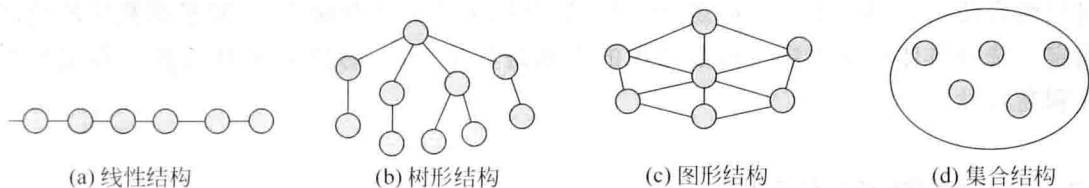


图 1.1 4 种基本结构示意图

依据数据元素之间存在的不同关系,数据结构可分为线性结构和非线性结构。线性结构中各个数据元素依次排列在一个线性序列中;非线性结构中各个数据元素不再保持在一个线性序列中,每个数据元素可能与零个或多个数据元素发生联系。此外,数据结构还可依据视点的不同,分为数据的逻辑结构和数据的物理结构(又称存储结构)。

数据的逻辑结构与数据的存储无关,研究数据结构的目的是为了在计算机中实现对数据的操作,而对数据的操作则依赖于数据的存储。

数据在计算机中的存储表示即数据的存储结构可分为顺序存储和非顺序存储,即顺序存储结构和链式存储结构,这是两种最基本的存储结构。此外,针对一些特殊的应用,还有可能会使用索引存储结构、散列存储结构等其他存储表示方法,以满足一些特殊的应用需求。

(1) 顺序存储方法是把逻辑上相邻的数据元素存储在物理位置上相邻的存储单元中,这种存储表示方法称为顺序存储结构。顺序存储结构是一种最基本的存储表示方法,算法设计时常用的数组就是用顺序存储结构来实现的。

(2) 链式存储方法对逻辑上相邻的数据元素不要求其物理位置上相邻,数据元素间的逻辑关系通过附设的指针字段值来指示,这种存储表示方法称为链式存储结构,链式存储结构是数据结构中最常见的一种存储表示方法,算法设计时常用的指针就是用链式存储结构来实现的。

(3) 索引存储方法在采用上述基本方法存储数据元素的同时,还需建立一个附加的索引表,其主要目的是为了提高检索的速度和性能。索引表中的每一项称为一个索引项。索引项的一般形式是(关键字,地址),关键字是指能唯一标识数据元素的数据项。例如,一本教材的目录就是一张索引表,它的一般形式是(章节标题和内容,页码),通过对索引表中关键字(标题和内容)的搜索,可迅速定位到所关心的主题内容所在的页码。若每个数据元素在索引表中均有一个索引项,则该索引表称为稠密索引(Dense Index);若一个索引项对应一组数据元素,则该索引表称为稀疏索引(Sparse Index)。

(4) 散列存储方法(Hash 方法)是依据数据元素的关键字值,采用一个事先已经设计好的散列函数计算出该数据元素的对应的存储地址,然后依据该地址及冲突处理办法实施对数据元素的存储和检索。这个地址称为散列地址,这种存储方法称为散列存储方法。显然,散列存储方法的关键在于设计一个合适的散列函数及冲突的处理办法,其优点是处理速度快。

在上述 4 种存储结构中,顺序、链式、索引存储结构在完成其数据元素值的机内表示的同时,还能完成其数据元素间逻辑关系的机内表示。散列存储方法只能存储数据元素的值,而不能存储数据元素间的逻辑关系。在实际应用中,各种不同的存储方法既可以单独使用,