

21世纪高等学校规划教材 | 计算机科学与技术



数据结构与算法

张瑞霞 张敬伟 编著



清华大学出版社



数据结构与算法

张瑞霞 张敬伟 编著



内 容 简 介

本书系统完整地介绍了线性表、栈和队列、树和二叉树、图和字符串等数据结构的抽象数据类型、操作实现和应用实例,并讨论了各种查找技术和排序算法。通过数据结构抽象数据类型定义和接口封装,培养读者的抽象思维能力。书中的算法采用规范完整的C语言描述,读者只需添加主程序就能够运行程序,进而在调试层面理解算法,从而跨越抽象和具体之间的鸿沟。本书通过每章开头的两个关键词进行章节主要内容概览,通过大量的图表辅助读者理解复杂的算法过程,通过应用实例和习题中的算法设计题目及应用题目强化、提高读者的应用能力。

本书既可作为高等院校计算机类相关专业的教材,也可作为高职院校计算机类相关专业的教材,还可作为计算机爱好者的自学书籍和计算机软件开发的工程技术人员的参考书。

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 13701121933

图书在版编目(CIP)数据

数据结构与算法/张瑞霞,张敬伟编著. —北京:清华大学出版社,2018

(21世纪高等学校规划教材·计算机科学与技术)

ISBN 978-7-302-50557-0

I. ①数… II. ①张… ②张… 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 ②TP312

中国版本图书馆CIP数据核字(2018)第140941号

责任编辑:郑寅堃 王冰飞

封面设计:傅瑞学

责任校对:焦丽丽

责任印制:丛怀宇

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:三河市国英印务有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:17

字 数:413千字

版 次:2018年6月第1版

印 次:2018年6月第1次印刷

印 数:1~1000

定 价:49.00元

产品编号:075113-01

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: weijsj@tup.tsinghua.edu.cn

“数据结构”是计算机大类相关专业的一门重要的专业基础课,既是计算机大类相关专业考研的必考科目,又是从事 IT 相关工作必须掌握的专业基础素养。

“数据结构”课程的教学目的是使学生能够针对具体问题选择合适的数据结构,并合理地组织数据,有效地存储和处理数据,以培养学生的数据抽象能力;使学生能够将数据结构和应用付诸编程实践,正确地设计、编制高效算法,并对算法进行分析和评价,以培养学生良好的程序设计开发技能;使学生能够应用工程知识和专业背景知识分析复杂工程问题,进行复杂程序设计的训练,解决工程实践问题,以培养学生的工程实践能力。这里围绕教学目的就教学改革、教材内容安排和资源等方面展开说明。

数据结构的教学与改革

在近十年的数据结构教学中,围绕着课程教学目的,编者在以下几个方面体会深刻。

1. 处理好数据结构中抽象与具体的关系

现有的教科书对概念和算法的讲解过于抽象,初学者不容易接受。在数据结构的教学中过程中要注意化抽象为具体,同时也要注意培养学生的抽象思维能力。数据结构的基本操作的算法实现设计更为具体,读者只要添加主程序就可以直接执行,另外每种数据结构都设计了抽象数据类型,能提高读者的抽象思维能力,并且在基本线性表部分通过动态链接库指导学生掌握封装的技术。

2. 通过编程调试理解算法

在课堂教学中,编者通过调试讲解算法,使学生所见即所得,得到了学生的一致认可。例如在学习指针时,教师虽然做了一些比喻,但学生总是觉得和计算机程序真正运行起来相比还是有差距的。在教学开发环境中将调试信息给学生看,观看真正分配的内存地址,指针变量所在的位置,函数调用时栈的分配与回收,链表的插入、删除等,学生切实看到了真实的情况,印象深刻,教学效果好。

3. 描述语言和应用实践

由于 C 语言得到了广泛的应用,对数据结构与算法的初学者来说能够将更多的精力关注在这方面,而不是语言本身。因此教学中的绝大部分实现采用 C 语言,但在有些复杂的应用问题的解决中引导学生使用 STL,并且指导学生有意识地使用 Python、Java、C# 等语言中数据结构的相关接口,其目的是关注应用问题的解决方案,而不是细节的处理。

4. 教学相长

“师者,所以传道授业解惑也。”教学是个不断积累的过程,编者在近十年的教学过程中一方面传授给学生知识,另一方面在解决学生问题的过程中也不断提升自身的教学水平,更新完善教学内容、迭代开发教学资源。

内容组织特色

本书的内容组织有以下特色。

(1) 以基本数据结构为主线,包括线性表、栈和队列、二叉树、图、散列等,以查找为脉络,贯穿其中,包括顺序查找、二分查找、散列查找和模式匹配等,由浅入深,由简及繁,符合学生的认知规律。

(2) 在每章开头总结了两个关键词进行章节内容概览,这样做的目的是“开宗明义”,使学生能够尽快抓住要点展开学习。在章节内容的安排上,按小节组织每章中涉及的知识点,教师既能按照常规组织教学,也能自行组织微课进行教学,学生能够灵活安排进行碎片化学习。

(3) 侧重动手应用实践的主线。每种数据结构基本算法不是伪代码,而是提供完整的可直接运行的程序,之后安排了应用该数据结构的具体应用问题,所有程序按照统一的风格提供。章节间的先后顺序也有侧重实践主线的思想,例如搜索树安排在树和二叉树章节的后面,使学生可以应用已学的二叉树和高级搜索算法进行高效查找算法设计实践。课后习题没有设置填空题和选择题,而是以算法设计和数据结构应用为主,进一步强调对学生的分析问题、解决问题的能力培养。

(4) 点明数据结构应用的多个领域以及课程间的关联。例如 Linux 中用到的双循环链表数据结构,编译原理中的表达式计算,区块链中用到的 Merkle 树,人工智能词汇切分中用到的 Trie 树,让学生感受到数据结构的博大精深和无限魅力。

教学资源特色

本书提供了多样的立体化学习资源,既有多媒体 PPT 课件、完整的工程代码和图表等静态资源,也有跨平台的 H5 演示软件等动态资源,以适应读者的不同需求,同时提供配套的实验教材,强化学生的动手实践能力。

本书由桂林电子科技大学计算机与信息安全学院张瑞霞、张敬伟共同编著而成。张瑞霞负责全书的整体统稿构思,智国建教师为本书的编辑、排版做了大量的工作,课程组组长周娅以及课程组的教师们为本书提出了有益的建议。唐麟老师为部分代码做了调试工作,曾泽宇、苏宣瑞、陈思博、王馨茹、胡星高、李婷、骆志成等同学分别指出了多处纰漏和错误,在此谨向他们表示感谢!

感谢清华大学出版社的各位编辑,正是依靠他们的辛勤工作鼎力支持,本书才得以顺利出版,特别感谢郑寅堃、王冰飞两位编辑高效负责的工作。

由于编者水平有限,书中难免存在不足和错误之处,欢迎读者不遗余力地批评、指正,在此深表感谢。

编 者

2018年3月

目录

第 1 章 绪论	1
1.1 为什么要学习数据结构	1
1.2 抽象数据类型	3
1.3 数据结构	4
1.3.1 数据结构的基本术语	4
1.3.2 数据结构研究的三要素	5
1.4 算法与算法效率	6
1.4.1 算法举例	6
1.4.2 什么是算法	7
1.4.3 算法评价标准	8
1.4.4 算法描述方法	8
1.5 算法分析	10
1.5.1 算法比较举例	10
1.5.2 时间复杂度分析	11
1.5.3 常见循环的时间复杂度举例	12
习题	13
第 2 章 线性表	15
2.1 线性表的概念	15
2.1.1 线性表的定义	15
2.1.2 线性表的抽象数据类型定义	16
2.1.3 顺序表 VS 链表	16
2.2 顺序表的建立与判空	19
2.2.1 创建空的顺序表	19
2.2.2 判断顺序表为空	20
2.2.3 扩展延伸：通过调试理解算法	20
2.3 顺序表的插入和删除	22
2.3.1 插入算法	22
2.3.2 删除算法	24
2.3.3 小白实践：完整示例	25
2.4 顺序表的查找定位	26
2.4.1 查找算法	26

2.4.2	二分查找	26
2.5	单链表的建立与判空	29
2.5.1	建立单链表	29
2.5.2	链表的判空	29
2.5.3	用头插法建立单链表	29
2.5.4	用尾插法建立单链表	30
2.6	单链表的查找	31
2.7	单链表的插入	31
2.7.1	后插算法	31
2.7.2	前插算法	32
2.8	单链表的删除	33
2.8.1	按位置删除	33
2.8.2	按值删除	34
2.9	单循环链表	35
2.10	双链表和双循环链表	37
2.10.1	双链表	37
2.10.2	双循环链表	38
2.11	线性表的应用:一元多项式的表示和运算	39
2.12	线性表的应用: Josephus 问题	41
2.13	动态链接库	43
2.13.1	动态链接库的概念	43
2.13.2	动态链接库的优缺点	43
2.13.3	动态链接库的构建与链接	43
习题		46
第3章	栈和队列	48
3.1	栈和队列的概念	48
3.1.1	栈和队列的定义	48
3.1.2	栈的抽象数据类型定义	49
3.1.3	栈混洗	49
3.2	顺序栈	50
3.2.1	创建空栈	51
3.2.2	判断栈空	51
3.2.3	进栈	51
3.2.4	出栈	52
3.2.5	取栈顶元素	52
3.3	链栈	53
3.3.1	创建空栈	53
3.3.2	判断栈空	53

3.3.3	进栈	54
3.3.4	出栈	54
3.3.5	取栈顶元素	54
3.4	栈的应用: 进制转换	55
3.5	栈的应用: 括号匹配	56
3.6	栈的应用: 栈与递归	58
3.7	栈的应用: 迷宫	60
3.8	栈的应用: 表达式求值	64
3.9	循环队列	68
3.9.1	创建空队列	69
3.9.2	判断队列是否为空	70
3.9.3	入队	70
3.9.4	出队	70
3.9.5	取队头元素	71
3.10	链队列	71
3.10.1	创建空队列	71
3.10.2	判断队列是否为空	72
3.10.3	入队	72
3.10.4	出队	73
3.10.5	取队头元素	73
3.11	队列的应用: 迷宫	73
3.12	队列的应用: 农夫过河	76
3.13	双端队列	79
	习题	80
第4章	树和二叉树	82
4.1	二叉树的概念	82
4.1.1	二叉树的基本形态和分类	82
4.1.2	二叉树的抽象数据类型定义	83
4.2	二叉树的数学性质	84
4.3	二叉树的深度优先遍历	85
4.4	二叉树的广度优先遍历	87
4.5	二叉树的重构	88
4.6	二叉树的交叉遍历	90
4.7	二叉树的顺序存储	92
4.8	二叉树的链式存储	93
4.9	二叉树的建立和遍历(递归算法)	94
4.9.1	二叉树的遍历	94
4.9.2	二叉树的建立	95

4.10	二叉树的建立和遍历(非递归算法)	96
4.10.1	二叉树建立的非递归实现	96
4.10.2	先序遍历的非递归实现	97
4.10.3	中序遍历的非递归实现	100
4.10.4	后序遍历的非递归实现	101
4.11	二叉树的其他操作	103
4.11.1	统计二叉树的叶子结点数	103
4.11.2	计算二叉树的深度	103
4.11.3	复制一棵二叉树	104
4.12	线索二叉树	104
4.12.1	线索二叉树的定义	104
4.12.2	建立线索二叉树	105
4.12.3	遍历线索二叉树	106
4.13	二叉树的应用:哈夫曼树与哈夫曼编码	107
4.14	树和森林	113
4.14.1	树和森林的概念	113
4.14.2	树和森林的遍历	113
4.14.3	树的存储表示	114
4.14.4	树、森林与二叉树的转换	115
	习题	117
第5章	搜索树	119
5.1	二分查找判定树	119
5.2	二叉排序树的基本概念	120
5.3	二叉排序树的查找	121
5.4	二叉排序树的插入	122
5.5	二叉排序树的删除	123
5.6	平衡二叉树的概念	127
5.7	平衡二叉树的实例	128
5.8	平衡二叉树的4种调整和两个基本操作	128
5.9	AVL的插入操作	131
5.10	AVL的删除操作	136
5.11	红黑树的基本概念	139
5.12	红黑树的插入	140
5.13	红黑树的删除	144
	习题	147
第6章	图	149
6.1	图的基本概念和抽象数据类型定义	149

6.1.1 图的基本概念	149
6.1.2 图的抽象数据类型定义	151
6.2 图的存储表示	152
6.2.1 邻接矩阵	152
6.2.2 邻接表	154
6.3 图的遍历	156
6.3.1 深度优先搜索	156
6.3.2 广度优先搜索	158
6.3.3 图的连通分支	160
6.3.4 图的层数	160
6.4 Prim 算法	162
6.5 Kruskal 算法	165
6.6 Dijkstra 算法	169
6.7 拓扑排序	171
6.7.1 AOV 网	171
6.7.2 拓扑排序算法	173
6.8 关键路径	175
6.8.1 AOE 网	175
6.8.2 关键路径算法	176
6.9 六度空间问题	181
6.10 中国邮递员问题	183
6.10.1 问题的引入	183
6.10.2 相关知识点	184
6.10.3 算法流程	184
6.10.4 核心算法设计	185
6.10.5 具体实现	186
习题	191
第 7 章 字典	194
7.1 字典的基本概念	194
7.2 跳跃链表的基本概念	195
7.3 跳跃链表的建立和查找	196
7.3.1 空跳跃链表的建立	196
7.3.2 跳跃链表的查找	197
7.4 跳跃链表的插入和删除	198
7.4.1 跳跃链表的插入	198
7.4.2 跳跃链表的删除	199
7.5 散列表的基本概念	200
7.6 散列函数和冲突	201

7.6.1	散列函数	201
7.6.2	生日悖论	203
7.6.3	解决冲突的方法	204
7.7	散列表的建立、查找、插入和删除	207
7.7.1	散列表的建立	207
7.7.2	散列表的查找	208
7.7.3	散列表的插入	210
7.7.4	散列表的删除	211
7.8	Merkle 树的基本概念	211
7.9	Merkle 树的建立和查找比较	213
7.9.1	Merkle 树的建立	213
7.9.2	Merkle 树的查找比较	215
	习题	217
第 8 章	排序	218
8.1	排序的基本概念	218
8.2	插入排序	219
8.2.1	直接插入排序	219
8.2.2	二分插入排序	221
8.2.3	Shell 排序	223
8.3	选择排序	225
8.3.1	直接选择排序	225
8.3.2	堆排序	226
8.4	交换排序	230
8.4.1	冒泡排序	230
8.4.2	快速排序	231
8.5	基数排序	234
8.6	归并排序	237
8.7	排序算法的比较	240
	习题	241
第 9 章	字符串	243
9.1	字符串的基本知识	243
9.1.1	字符串的基本概念	243
9.1.2	串的抽象数据类型定义	243
9.1.3	C 库接口	244
9.1.4	正则表达式	245
9.2	朴素的模式匹配算法	245
9.3	KMP 算法	247

9.3.1 KMP 算法的思想	247
9.3.2 next 表的存在性分析	248
9.3.3 构造 next 表	249
9.3.4 改进 next 表	250
9.4 Trie 树	251
9.4.1 Trie 树的基本概念	251
9.4.2 Trie 树的基本操作	252
习题	254
参考文献	255

第1章

绪论

本章关键词：脉络和算法。

关键词——脉络。脉络在这里指数据结构研究三方面的问题,包括逻辑结构、物理存储和操作实现(本书采用C语言描述),在后续各章数据结构的具体学习过程中主要把握这三方面的问题。逻辑结构反映的是数据内部之间的构成方法,而物理存储是指数据在内存中的存储方法,操作实现是某种逻辑结构的数据在确定的存储形式下的具体算法。不同数据结构有着各自不同的逻辑特点,因此有着不同的应用场景。

关键词——算法。数据结构和算法是程序的两大要素,二者相辅相成。同一个问题可以有多个不同的算法,因此在设计算法时除了需要考虑算法的正确性、可读性和健壮性以外,还需要考虑算法的时间效率和空间效率这两个重要评价指标。其中,大 O 记号表示的渐近时间复杂度是最基本、最常用的度量标准,也是本书采用的记号方法。在本章中读者需要掌握最常见的复杂度等级的分析方法和技巧。值得注意的是,分析算法效率的目的是不断地改进算法。

“不积跬步,无以至千里;不积小流,无以成江海。”

——《荀子》

1.1 为什么要学习数据结构

用计算机解决问题一般需要几个步骤,首先是将具体问题抽象出一个适当的数学模型,然后选择描述此模型的数据结构和算法,最后实现程序的调试和测试。计算机在最初出现的时候,其目的是为了进行科学和工程的计算。这时处理的纯数值计算问题往往可以用数学方程描述,例如微积分方程的计算、线性方程组、矩阵的计算等。随着计算机的迅速发展,计算机的应用领域不断扩大,计算机处理的数据不再局限于纯数值的计算,符号、声音、图像等信息均可通过编码存储到计算机中进行处理,针对这样的非数值数据难以用数学方程描述,需要考虑数据之间的相互关系。下面通过3个例子说明数据之间的相互关系以及表示。

例 1-1 职工信息管理系统。

职工信息管理系统能够实现对职工的增加、删除、查询、更改等工作。在职工信息管理系统中建立一张简单的职工信息登记表,如表1-1所示。在该表中,每一行对应一个职工信息,称为信息表项,各个表项之间的关系是线性的,可以用线性表数据结构来描述。

表 1-1 职工信息表

工 号	姓 名	性 别	工 龄	基 本 工 资
0001	张三	女	15	4100
0002	李四	男	6	3500
0003	王五	男	23	5600
...	...	...	...	...

例 1-2 人机对弈问题。

考虑“井”字棋的人机对弈问题,前提是需要将对弈的策略事先存入计算机。在对弈过程中,计算机的操作对象是对弈过程中可能出现的棋盘布局,而棋盘布局之间的关系是由对弈规则决定的,这个关系通常不是线性的。这是因为从一个布局可以演变成多个布局,将对弈过程中从开始到结束的所有可能出现的布局描述出来,如图 1-1 所示。这些布局可以用树形结构来描述。“树根”是起始布局,所有的“叶子”是最后可能出现的布局,而对弈过程就是从树根沿着“树枝”到叶子的过程。

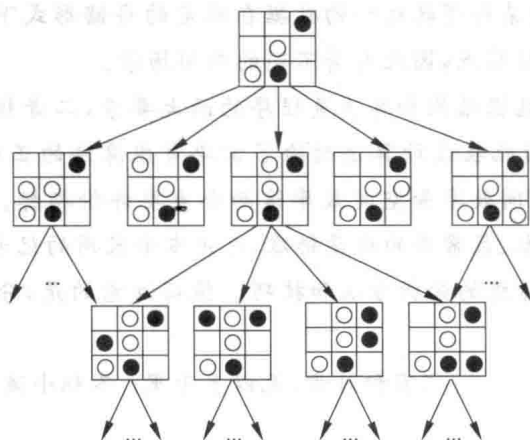


图 1-1 对弈局部

例 1-3 大厦维修活动安排问题。

某商业大厦要进行维修,会涉及多个活动,而这些活动中有些活动之间是有先后次序关系的,有些没有先后次序关系,如表 1-2 所示。具有先后次序关系的活动必须按照先后次序进行,例如门窗的维修需要在外墙维修完成之后进行。这些先后次序关系可以用图的数据结构来描述,如图 1-2 所示。

表 1-2 活动计划表

活 动 名 称	符 号	活动时间/天	依 赖 活 动
框架	C_0	14	
屋面	C_1	22	C_0
外墙	C_2	25	C_0
门窗	C_3	17	C_2
卫生管道	C_4	34	C_2
各种电气	C_5	35	C_1