



普通高等教育“十一五”国家级规划教材 计算机系列教材

数据结构与经典算法

李千目 殷新春 李涛 主编

清华大学出版社





普通高等教育“十一五”国家级规划教材 计算机系列教材

李千目 殷新春 李涛 主编

数据结构与经典算法

清华大学出版社
北京

内 容 简 介

本书经过国内著名高校的培优班、精英班的实际教学检验,由浅入深,循序渐进,通过案例来讲解理论,以浅显易懂的文字与图表对各种数据结构和算法的设计进行分析,对问题的解决方法做了详尽的剖析。

本书注重原理和思想,尽量简化模型,强调其背后的基本思想,以基础理论-实验-经典题库为主线进行编写,辅之以相应的类C语言代码,从而增进读者对数据结构的理解与掌握。

全书共分为12章,内容包括基础知识、线性存储结构、栈、队列、串、数组与广义表、树型结构、图状结构、查找、内部排序、经典算法、数据分析与挖掘。其中,第11章是经典算法解析,第12章简要地介绍了数据挖掘的知识,本书安排了大量的实验和练习方便读者学习和使用。

本书各章相对独立,结构清晰、易教易学、实例丰富、可操作性强、学以致用、注重能力,对在学习过程中常见的重点和难点进行立体、详细的讲解,以帮助读者更好地掌握数据结构的基本知识。本书可以作为高等院校计算机专业及相关专业的教材,也适合数据结构的初学者研读,还适合考研复习之用。

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

数据结构与经典算法/李千目等主编. —北京:清华大学出版社,2015

计算机系列教材

ISBN 978-7-302-40083-7

I. ①数… II. ①李… 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(2015)第089628号



责任编辑:谢琛 薛阳

封面设计:常雪影

责任校对:梁毅

责任印制:王静怡

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:北京四季青印刷厂

装订者:三河市吉祥印务有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:24.75

字 数:615千字

版 次:2015年6月第1版

印 次:2015年6月第1次印刷

印 数:1~2000

定 价:49.00元

产品编号:061614-01

程序 = 数据结构 + 算法

——著名的计算机科学家沃斯(Nikiklaus Wirth)

“数据结构”是高等学校计算机科学与技术、软件工程和网络工程专业及其他相关专业的一门核心专业基础课程,也是计算机类专业研究生入学考试全国统考课程。本书正是为了满足人才培养和研究生入学考试发展的需要,经过多年的国内多所著名高校实际教学凝炼而成。

自从著名的计算机科学家沃斯将程序设计形象地用上面的公式表示出来后,这条“黄金定律”便成为了人们学习程序设计、进行程序开发的准则。要想成为一名真正专业的程序设计人员,基本的数据结构基础和常用的算法知识是必须掌握的,脱离了这两点编写出来的程序一定不是健壮的好程序。

然而单纯地掌握了一些数据结构基础和常用的算法知识也是远远不够的。空洞地掌握所谓的数据结构和算法等理论知识只是纸上谈兵,这些知识必须要依托于一门程序设计语言才有生命力,才能转化成真实的程序代码,才能真正地解决实际问题。

本书就是将数据结构基础和常用的算法知识与目前广泛应用、最具有群众基础的C语言相结合。本书的写作思想是理论与实践相结合,以实践为核心,以实例为主要内容。

首先,本书总结归纳了数据结构基础、常用的排序查找算法和经典的算法思想,提纲挈领地阐述了核心理论知识。这样可以使没有系统学习过或者不熟悉数据结构和算法等知识的读者对这部分知识有一个基本了解,并掌握基本的数据结构知识和常用而经典的算法思想,以便更加深入地学习本书的其他内容。

其次,本书列举了大量的编程实例,这些题目都按照知识体系进行了内容上的划分。本书列举的这些编程实例都是一些比较灵活有趣的题目,有些题目渗透了巧妙的算法思想,有些题目则必须借助特殊的数据结构才能更加容易解答。通过这些题目的训练,可以使读者开阔眼界、启迪思维、提高编程的兴趣。最重要的是能够提高读者算法设计,灵活应用各种数据结构,编写程序解决实际问题的能力。

1. 本书阅读指南

本书共分为12章,各章主要内容如下。

第1章介绍数据与结构、数据结构与算法等概念。

第2章介绍线性表的顺序存储方式和链式存储方式以及它们的各种应用。

第3章介绍栈的两种存储形式和基本操作、递归算法、表达式的应用、后缀表达式的计算、迷宫问题等。

第4章介绍队列的两种存储形式和基本操作、循环队列的使用等。

第5章介绍串的基本操作、串的两种存储结构,以及如何进行串的匹配;简述了串操作在文本编辑中的应用,包括输入文本、插入文本和删除文本。

第6章介绍数组的基本操作、各种特殊矩阵的压缩方法,以及稀疏矩阵三元表的构造、基本运算的实现。广义表部分主要介绍了广义表的存储结构和应用。

第7章介绍树的基本概念、二叉树的各种操作、树的两种遍历方法,详细说明了线索二叉树的构造方法和运算二叉树的原理及使用方法,介绍了树和森林与二叉树是如何转换和遍历的,最后介绍了哈夫曼树及哈夫曼编码。

第8章介绍图的存储形式,重点介绍了图的两种遍历方法、最小生成树以及 Dijkstra 算法和 Floyd 算法,最后简单地介绍了拓扑排序和关键路径的基本原理。

第9章介绍静态查找表和动态查找表以及哈希表的构造和冲突解决办法。

第10章介绍各种排序算法的原理,并比较了它们的时间复杂度、空间复杂度、稳定性、待排序记录数 n 的大小和记录本身信息量的大小。

第11章主要在前述章节所授数据结构和算法的基础上,介绍经典算法。

第12章主要给出了复杂数据结构+海量数据条件下,大数据时代的数据挖掘方法和技术简述。

2. 本书特色

(1) 结构清晰,知识完整,内容翔实,剪系统性强,依据高校教学大纲组织内容,并将实际经验融入基本理论之中。

(2) 入门快速,易教易学,突出“上手快、易教学”之特点,以教与学的实际需要取材谋篇。

(3) 学以致用、注重能力,以基础理论-实验-经典题库为主线编写,每一章都设置了实验和经典剖析,以便于读者掌握重点,提高分析问题和解决问题的能力。

(4) 示例丰富,步骤明确,讲解细致,突出了可操作性。

3. 本书读者定位

本书适合对数据结构和算法设计有兴趣的入门人员阅读,也可作为学习过 C 语言程序设计的人士继续深造的理想读物。另外,本书可以作为具有一定经验的程序设计人员巩固和提高编程水平、查阅相关算法实现和数据结构知识的参考资料。

同时,本书可以作为高等院校计算机专业及相关专业的教材,还适合考研复习之用。

本书由李千目、殷新春、李涛主编,徐建、张宏副主编,参加编写的还有侯君、莫璇、林棋、吴小雄。全书框架结构由何光明拟定。另外,卢振侠、王珊珊、石雅琴、陈珍、陈莉萍、杨橙、陈海燕、张居晓、刘菁、吴婷、李佐勇等同志也参与了资料整理、校对等工作,本书还参阅了严蔚敏、李宛洲等学者的研究,在此一并感谢。

限于作者水平,书中难免存在不当之处,恳请广大读者批评指正。

作者

2014 年 10 月

F O R E W O R D

第 1 章 基础知识 /1

- 1.1 数据结构研究什么 /1
- 1.2 基本概念和有关术语 /3
- 1.3 数据类型与抽象数据类型 /4
 - 1.3.1 数据类型 /4
 - 1.3.2 抽象数据类型 /4
- 1.4 算法描述与算法分析 /6
 - 1.4.1 算法的描述 /6
 - 1.4.2 算法的时间复杂度分析 /7
 - 1.4.3 算法的空间复杂度分析 /8
- 1.5 小结 /8
- 1.6 经典题库 /9
 - 1.6.1 要点提醒 /9
 - 1.6.2 经典剖析 /9
- 1.7 练习 /11

第 2 章 线性存储结构 /13

- 2.1 线性表的定义及基本操作 /13
 - 2.1.1 线性表的基本概念 /13
 - 2.1.2 线性表的抽象数据类型 /14
- 2.2 线性表顺序存储结构的定义 /15
 - 2.2.1 线性表顺序存储结构的定义 /15
 - 2.2.2 顺序表的基本操作及实现 /16
 - 2.2.3 顺序表的应用 /18
- 2.3 线性表链式存储结构与实现 /20
 - 2.3.1 线性表链式存储结构 /21
 - 2.3.2 单链表及其基本操作 /21
 - 2.3.3 循环链表的基本操作及实现 /29
 - 2.3.4 双链表的基本操作及实现 /33
 - 2.3.5 循环双链表 /37
- 2.4 实验 /38
 - 实验一：顺序表的操作 /38

实验二：单链表连接	/40
实验三：循环链表连接	/43
实验四：循环双链表操作	/47
2.5 小结	/52
2.6 经典题库	/52
2.6.1 要点提醒	/52
2.6.2 经典剖析	/53
2.7 练习	/61
第3章 栈	/65
3.1 栈	/65
3.1.1 栈的基本概念	/65
3.1.2 栈的抽象数据类型	/65
3.1.3 栈的顺序存储结构及实现	/66
3.1.4 栈的链式存储结构与实现	/70
3.2 栈的应用	/74
3.2.1 表达式求值	/74
3.2.2 数制转换	/76
3.2.3 括号匹配检验	/77
3.2.4 栈与递归的实现	/78
3.3 实验	/79
实验：利用栈寻找迷宫路径	/79
3.4 小结	/82
3.5 经典题库	/82
3.5.1 要点提醒	/82
3.5.2 经典剖析	/83
3.6 练习	/86
第4章 队列	/90
4.1 队列	/90
4.1.1 什么是队列	/90
4.1.2 队列的抽象数据类型	/90

4.1.3	队列的顺序存储结构与实现	/91
4.1.4	队列的链式存储结构与实现	/97
4.2	队列的应用	/101
4.3	实验	/105
	实验：利用队列模拟病人看病	/105
4.4	小结	/109
4.5	经典题库	/109
4.5.1	要点提醒	/109
4.5.2	经典剖析	/109
4.6	练习	/112
第5章	串	/116
5.1	串的定义	/116
5.1.1	串的定义	/116
5.1.2	串的抽象数据类型	/116
5.2	串的存储结构与实现	/118
5.2.1	串的顺序存储结构与实现	/118
5.2.2	串的堆存储结构与实现	/121
5.2.3	串的块链存储结构与实现	/122
5.3	串的模式匹配算法	/123
5.3.1	简单的模式匹配算法——BF 算法	/123
5.3.2	改进的模式匹配算法——KMP 算法	/125
5.4	串的应用	/127
5.5	实验	/128
	实验一：显示多位数数字字符	/128
	实验二：块链的基本操作	/130
	实验三：统计串中最长的重复子串	/135
5.6	小结	/137
5.7	经典题库	/138
5.7.1	要点提醒	/138
5.7.2	经典剖析	/138

5.8 练习 /140

第6章 数组与广义表 /142

6.1 数组的定义 /142

6.1.1 数组的概念与性质 /142

6.1.2 抽象数据类型 /143

6.2 数组相关结构的实现 /143

6.2.1 数组的顺序存储 /143

6.2.2 数组的基本操作实现 /144

6.3 矩阵的压缩存储探究 /147

6.3.1 特殊矩阵 /147

6.3.2 稀疏矩阵 /149

6.4 广义表 /153

6.4.1 广义表相关概念 /153

6.4.2 抽象数据类型 /154

6.4.3 广义表存储结构详解 /155

6.4.4 广义表的相关应用 /156

6.5 实验 /157

实验一：矩阵乘法 /157

实验二：三元组实现两个矩阵的乘法 /160

实验三：广义表的基本运算 /163

6.6 小结 /168

6.7 经典题库 /169

6.7.1 要点提醒 /169

6.7.2 经典剖析 /169

6.8 练习 /172

第7章 树型结构 /175

7.1 树的基本概念和术语 /175

7.1.1 树的基本概念 /175

7.1.2 基本术语 /175

7.2 二叉树 /176

7.2.1	二叉树的基本概念	/176
7.2.2	二叉树性质的探究	/178
7.2.3	抽象数据类型	/179
7.2.4	存储结构	/181
7.3	遍历二叉树的方法	/185
7.3.1	遍历的定义	/185
7.3.2	遍历算法探究	/186
7.4	线索二叉树	/189
7.5	树和森林	/192
7.5.1	树的存储结构	/192
7.5.2	二叉树与森林	/194
7.5.3	树和森林的遍历探究	/195
7.6	哈夫曼树	/196
7.6.1	哈夫曼树的基本概念	/196
7.6.2	哈夫曼树的构造算法	/197
7.6.3	哈夫曼树和哈夫曼编码	/199
7.7	实验	/201
	实验一：运算二叉树	/201
	实验二：统计二叉树结点的个数	/203
	实验三：统计二叉树的宽度	/205
	实验四：按层遍历二叉树	/207
7.8	小结	/210
7.9	经典题库	/210
7.9.1	要点提醒	/210
7.9.2	经典剖析	/210
7.10	练习	/217

第8章 图状结构 /220

8.1	图的相关定义和术语	/220
8.1.1	图的基本概念	/220
8.1.2	图的基本术语	/220
8.1.3	抽象数据类型	/222

8.2	图的存储结构探究	/223
8.2.1	邻接矩阵	/223
8.2.2	邻接表	/225
8.2.3	十字链表	/227
8.2.4	多重邻接表	/228
8.3	图的遍历	/229
8.3.1	深度优先遍历	/229
8.3.2	广度优先遍历	/230
8.4	最小生成树	/231
8.4.1	生成树的概念	/231
8.4.2	最小生成树	/231
8.4.3	Prim 算法	/235
8.5	最短路径探究	/238
8.5.1	单源点最短路径问题分析	/238
8.5.2	所有顶点对应最短路径问题分析	/240
8.6	拓扑排序探究	/241
8.7	关键路径	/244
8.8	实验	/244
	实验一：遍历算法	/244
	实验二：Prim 算法	/248
8.9	小结	/253
8.10	经典题库	/253
8.10.1	要点提醒	/253
8.10.2	经典剖析	/253
8.11	练习	/264

第 9 章 查找 /269

9.1	查找的基本概念	/269
9.2	静态查找表	/269
9.2.1	顺序表的查找	/269
9.2.2	有序表的查找	/270
9.2.3	索引顺序表的查找	/271

9.2.4	静态树表的查找	/273
9.3	动态查找表	/274
9.3.1	二叉排序树	/274
9.3.2	平衡二叉树	/276
9.3.3	B-和 B+ 树	/280
9.3.4	键树	/288
9.4	哈希表	/289
9.4.1	哈希表的概念	/289
9.4.2	哈希表的构造方法	/289
9.4.3	处理冲突的方法	/290
9.4.4	哈希表的查找及分析	/293
9.5	实验	/294
	实验一：二叉排序树的查找	/294
	实验二：哈希查找	/298
9.6	小结	/300
9.7	经典题库	/301
9.7.1	要点提醒	/301
9.7.2	经典剖析	/301
9.8	练习	/308
第 10 章	内部排序	/310
10.1	插入排序	/310
10.1.1	直接插入排序	/310
10.1.2	希尔排序	/311
10.2	交换排序	/313
10.2.1	冒泡排序	/313
10.2.2	快速排序	/315
10.3	选择排序	/316
10.3.1	简单选择排序	/316
10.3.2	堆排序	/317
10.4	归并排序	/320
10.5	基数排序	/321

10.6	各种内排序方法的比较	/324
10.7	实验	/325
	实验一: 双向冒泡排序	/325
	实验二: 数组归并排序	/328
	实验三: 计数排序	/331
	实验四: 字符串排序	/332
	实验五: 最高位关键字排序 MSD	/335
10.8	小结	/338
10.9	经典题库	/338
	10.9.1 要点提醒	/338
	10.9.2 经典剖析	/339
10.10	练习	/343

第 11 章 经典算法 /345

11.1	基础算法	/345
	11.1.1 河内之塔	/345
	11.1.2 八皇后	/346
	11.1.3 背包问题	/348
11.2	数、运算	/352
	11.2.1 筛选求质数	/352
	11.2.2 最大公因数、最小公倍数、因式分解	/353
	11.2.3 最大访客数	/356
11.3	关于博弈	/358
	11.3.1 洗扑克牌(乱数排列)	/358
	11.3.2 Craps 赌博游戏	/360
11.4	集合问题	/362
	11.4.1 格雷码(Gray Code)	/362
	11.4.2 m 元素集合的 n 元子集	/364
11.5	矩阵	/365
	11.5.1 稀疏矩阵存储	/365
	11.5.2 上三角、下三角、对称矩阵存储	/367

11.5.3 奇数魔方阵 /369

第 12 章 数据分析与挖掘 /371

12.1 大数据时代的挖掘 /371

12.1.1 数据挖掘 /371

12.1.2 从数据挖掘应用的角度看大
数据 /372

12.2 挖掘技术发展和历史 /373

12.3 十大挖掘算法简介 /374

参考文献 /378

第1章 基础知识

人们的日常生活和工作越来越离不开计算机,它的应用越来越广泛,计算机已不再是仅仅用于科学计算这样的高深领域,而是更多地用于控制、管理以及日常的休闲娱乐等领域。目前,计算机应用可以分为两大类,一类是科学计算和工业控制,另一类是商业数据处理。面向工业控制与科学计算的内容主要涉及它的计算方法、效率与速度等因素,某一特定的测控对象有特定的算法,本书主要侧重于解决问题的方法研究,例如高次方程的迭代算法、快速傅氏变换的蝶型算法等。面向商业管理是要解决海量数据的管理与关联分析,即使是一个特定的对象也有通用的数据管理形式,例如商业数据库系统,无论何种具体应用,它都是大量的表格一类的数据处理形式,在海量数据中检索与查询是一类至关重要的操作工具,于是,数据的逻辑结构与物理组织形式是本书要解决的主要问题,例如表数据的存储形式、索引结构等,也就是数据结构问题。

什么是数据结构?简单来说,数据结构的研究对象是数据元素,其目的是建立数据元素在计算机中的表达方法。即在一群有限的数据元素集合里,元素与元素之间相互关系的描述,称为它的数据结构。

例如,有一个字典的数据结构如下:

$D = \{(\text{able}, \text{能干的}), (\text{apple}, \text{苹果}), (\text{bug}, \text{虫}), (\text{code}, \text{代码}), (\text{cool}, \text{酷}), \dots, (\text{x-ray}, \text{X光}), (\text{year}, \text{年}), (\text{zoo}, \text{动物园})\}$

这里,单词是数据元素检索关键字,单词与注释构成数据元素(结点),元素结点之间所表达的关系是按字母的顺序排列,这就是字典这一特定对象的数据结构。

1.1 数据结构研究什么

数据结构研究的是客观事物个体属性在计算机中表达及描述的方法。学习数据结构,主要的学习内容包括:

(1) 数据结构的逻辑结构。根据应用对象设计有限元素集合中结点之间的逻辑关系,如线性表、树、图等。

(2) 逻辑结构在计算机中的物理实现。数据结构在计算机内存中的表示方法,称为数据结构的物理结构,以区别于前者的逻辑结构形式,如顺序表、链表、二叉树等。

(3) 数据结构中结点的操作运算。如插入、删除、检索等。

数据结构有线性与非线性之分。一个数据结构的关系里,除去端点外,每一个结点有且仅有一个前趋和后继时,这个数据结构是线性的,如数组、链表。如果数据结构关系中,其结点有一个以上的前驱或后继,则称为非线性的数据结构,如树、图。

表 1-1 为一份学生成绩单。可以把这张表看成由记录组成,表中的每个记录由 5 个数据项组成。记录之间是一种前后有序的线性关系。此外还有诸如账目管理、图书馆书

表 1-1 学生成绩单

姓名	网络安全	Visual Basic	计算机理论基础	数据库
张三	80	90	70	75
李四	85	75	80	60
王五	62	86	95	100
赵六	70	74	85	80

籍管理等也与成绩单具有类似的数学模型。这类数学模型可称为线性的数据结构。

在体育比赛中，一般都是通过逐层选拔，产生最后的冠军。因此可以把最初的参赛者看成树叶，冠军看成树根，其他各级看成是树的枝干，这样就构成了一个树型结构，如图 1-1 所示。树中的结点之间不再是顺序关系，而是分层、分叉的非线性结构。

还有一类问题可以用通信网络来模拟，如图 1-2 所示。在通信网络中，如果把路由器看成若干顶点，把路由器之间的链路看成边，它们可以构成一个网状图。

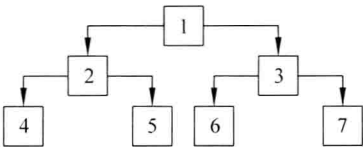


图 1-1 树型图

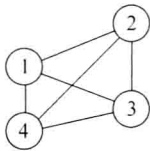


图 1-2 网状图

图 1-2 中顶点之间的关系纵横交错、错综复杂，这种关系称为图状结构。在实际应用中，如果一个数据包要在图中传递，则需要仔细考虑它的传送路径，以使传送时间最短。网络图的存储、管理以及结点的插入、删除等已不再是一个单纯的数值计算问题，而是一个非数值的信息处理问题。

综上所述，描述这类非数值问题的数学模型是诸如树、线性表、图之类的结构，而数据结构则是研究这些结构的一门学科。因此，它的重要性是不言自明的。

1968 年，D. E. Knuth 教授开创了数据结构的最初体系。数据结构是一门研究怎样合理地组织数据、建立合适的数据结构、提高计算机执行程序所用的时间和空间效率的学科。在计算机科学中，“数据结构”是一门综合性的专业基础课，不仅涉及计算机硬件，而且和计算机软件的研究也有着密切的关系，无论是编译程序还是操作系统，都涉及数据元素在存储器中的分配问题。另外，在研究信息检索时必须考虑如何组织数据，以便查找和存取数据元素更为方便。因此，可以认为“数据结构”是介于数学、计算机硬件和软件三者之间的一门核心课程。数据结构不仅是一般程序设计的基础，而且是设计和实现编译程序、操作系统、数据库系统及其他系统程序和大型应用程序的重要基础。

在结束有关数据结构的概念讨论之前，必须明确数据结构内容的三要素是：

数据结构 = 数据逻辑结构 + 物理结构 + 数据运算