

21世纪高等学校规划教材 | 计算机科学与技术



# 数据结构教程

(第二版)

王少波 张志 编著

非外借



清华大学出版社

21世纪高等学校规划教材 | 计算机科学与技术

# 数据结构教程

(第二版)



王少波 张志 编著

清华大学出版社  
北京

## 内 容 简 介

“数据结构”是计算机及信息管理专业的必修课程。

本书是作者在总结三十多年数据结构教学经验的基础上编写而成。全书共9章,内容涵盖数据结构的基本概念、线性表和串、栈和队列、树和二叉树、图、数组和矩阵、排序、查找、文件。本书采用C++程序设计语言对算法进行描述。本书不仅介绍了数据结构的相关理论,而且运用大量的实际案例充实教材的内容,力求既有理论深度,又有实用价值。附录A中还给出了数据结构课程实践中用到的VC++6.0编译环境介绍;附录B给出本课程实践内容及要求;附录C给出实践报告范本。每章都提供相关习题并附有部分习题答案。

本书是按高等院校对计算机及信息管理专业本科四年制教学大纲的要求编写的教材,也可以作为其他相关专业的教材,还可以作为计算机科技工作者的参考书。

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

数据结构教程/王少波,张志编著.—2版.—北京:清华大学出版社,2017  
(21世纪高等学校规划教材·计算机科学与技术)  
ISBN 978-7-302-47683-2

I. ①数… II. ①王… ②张… III. ①数据结构—教材 IV. ①TP311.12

中国版本图书馆CIP数据核字(2017)第155285号

责任编辑:闫红梅 战晓雷

封面设计:傅瑞学

责任校对:焦丽丽

责任印制:李红英

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](mailto:c-service@tup.tsinghua.edu.cn)

质 量 反 馈:010-62772015, [zhiliang@tup.tsinghua.edu.cn](mailto:zhiliang@tup.tsinghua.edu.cn)

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

印 装 者:清华大学印刷厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:26.5 字 数:644千字

版 次:2011年7月第1版 2017年10月第2版 印 次:2017年10月第1次印刷

印 数:1~2000

定 价:49.00元

产品编号:074844-01

# 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:weijj@tup.tsinghua.edu.cn

# 前言

本书是按高等院校计算机专业及信息管理专业本科四年制教学计划“数据结构”课程教学大纲要求编写的教材,还可以作为计算机科技工作者及其相关专业人员的参考书。在学习本书知识前,要求读者具备 C++ 程序设计的知识。

“数据结构”已成为一门比较成熟的课程。它是计算机系统软件和应用软件研制者的必修课程。数据结构和算法是计算机基础性研究内容之一,掌握这个领域的知识对于利用计算机资源高效地开发计算机程序是非常必要的。

数据结构理论的应用范围已经深入到编译系统、操作系统、数据库、人工智能、信息科学、系统工程、计算机辅助设计及信息管理领域。数据结构主要解决非数值计算应用问题。

从理论上讲:数据结构的概念严谨、抽象;每种数据结构类型描述层次清晰可见——概念层、逻辑定义层、物理存储层、运算实现层;每种数据结构类型描述反映了实现问题的思想、实现的前提以及不同实现方式的特点和优劣。

数据结构描述的内容看上去如同程序,但不是程序,它是程序设计思想的抽象化、一般化,它不依赖于某种物理设备甚至某种语言系统,学习者通过“数据结构”课程不仅能获得专业知识,而且能学到一种思维方式。

从实践上讲,数据结构是建立在抽象化描述基础之上的实践性理论,这门学科只有赋予实践的内容才具有完备性,具体化是该学科的又一特点。在计算机系统中全面体现着数据结构的作用,系统框架结构的构建、程序实现的精巧化都融入了数据结构的理论思想和技术。

本书叙述了各种基本数据结构的概念,包括数据结构的逻辑定义、物理实现及其相应运算,并举例说明怎样用这些抽象的概念来解决实际问题。通过本书的学习不仅能正确地掌握数据结构的基本理论,并能运用这些理论来解决实际问题。

本书是编者集多年从事计算机软件设计实践及讲授“数据结构”课程的体会,并参考分析了国内外数据结构书籍文献编写而成的。本书采用广泛使用的 C++ 语言描述算法,并进行了适当的算法复杂性分析。

“数据结构”课程不但理论性很强,同时实践性也很强。本书在每一章的最后都安排了适量的习题,供读者练习。

本书共分 9 章,介绍了数据结构的基本概念及线性表和串、栈和队列、树和二叉树、图、数组和矩阵、排序、查找、文件的数据结构、算法及其应用案例。

本书由王少波、张志编著。王少波负责编写第 1、2、3、4、7 章及附录,张志负责编写第 5、6、8、9 章,全书由王少波负责统稿。

在成书过程中,编者参考了有关书籍,在此向这些书籍的作者表示感谢。

由于编者水平有限,书中可能存在不妥与疏漏之处,恳请读者不吝指教。

编者

2017 年 6 月

第 1 章 绪论	1
1.1 什么是数据结构	2
1.1.1 数据结构相关事例	2
1.1.2 数据结构的定义	5
1.2 数据结构的相关概念	6
1.2.1 数据和信息	6
1.2.2 数据元素	6
1.2.3 结构类型	7
1.2.4 静态存储空间分配回收和动态存储空间分配回收	10
1.3 数据类型、抽象数据类型和数据结构	11
1.3.1 类和数据类型	11
1.3.2 抽象数据类型	12
1.3.3 数据结构、数据类型和抽象数据类型	13
1.4 算法及算法分析、算法描述	13
1.4.1 算法和程序	13
1.4.2 程序性能和算法效率	15
1.4.3 算法分析	16
1.4.4 算法描述	20
习题 1	23
第 2 章 线性表和串	27
2.1 线性表的定义	27
2.1.1 线性表的逻辑结构	27
2.1.2 线性表的抽象数据类型	28
2.2 线性表的顺序存储及操作	28
2.2.1 线性表顺序存储	28
2.2.2 线性表顺序存储结构下的操作实现	31
2.3 简单链表存储结构及操作	35
2.3.1 简单链表的存储	35
2.3.2 简单链表的操作实现	38
2.4 双向链表	45
2.4.1 双向链表的存储	45

2.4.2	双向链表类定义 .....	46
2.4.3	双向链表的操作 .....	47
2.5	单向循环链表和双向循环链表 .....	52
2.5.1	单向循环链表的存储 .....	52
2.5.2	双向循环链表的存储 .....	53
2.6	模拟指针方式构造简单链表 .....	54
2.6.1	模拟链表的存储空间的构建 .....	54
2.6.2	在模拟链表空间上构建简单链表 .....	57
2.7	多重链表 .....	60
2.8	链表应用 .....	62
2.8.1	结点移至表首运算 .....	62
2.8.2	链表的逆向运算 .....	63
2.8.3	多项式的相加运算 .....	64
2.8.4	十字链表结构的应用 .....	69
2.8.5	一个较复杂的机票售票系统的数据结构方案 .....	71
2.9	串 .....	72
2.9.1	串的定义 .....	73
2.9.2	串的逻辑结构及运算 .....	73
2.9.3	串的顺序存储结构 .....	74
2.9.4	串的链式存储结构 .....	74
2.10	线性表基本算法的程序实现 .....	75
2.10.1	顺序存储结构线性表程序实现 .....	75
2.10.2	带表头结点的简单链表程序实现 .....	80
习题 2	.....	91
<b>第 3 章</b>	<b>堆栈和队列 .....</b>	<b>94</b>
3.1	堆栈的定义 .....	94
3.1.1	堆栈的逻辑结构 .....	94
3.1.2	堆栈的抽象数据类型 .....	94
3.2	堆栈的顺序存储及操作 .....	95
3.2.1	堆栈顺序存储 .....	95
3.2.2	顺序存储结构堆栈的运算实现 .....	96
3.3	堆栈的链式存储及操作 .....	98
3.3.1	堆栈的链式存储 .....	98
3.3.2	链式栈类的定义 .....	99
3.3.3	链式栈类运算的实现 .....	99
3.4	多个栈共享邻接空间 .....	101
3.5	堆栈的应用 .....	102
3.5.1	检验表达式中括号的匹配 .....	102



3.5.2	表达式的求值	104
3.5.3	背包问题求解	106
3.5.4	地图四染色问题求解	109
3.6	队列的定义	114
3.6.1	队列的逻辑结构	114
3.6.2	队列的抽象数据类型	115
3.7	队列的顺序存储及操作	116
3.7.1	队列的顺序存储	116
3.7.2	顺序存储结构下队列的运算实现	119
3.8	队列的链式存储及操作	121
3.8.1	队列的链式存储	121
3.8.2	链式队列模板类的定义	122
3.8.3	链式队列的操作	122
3.9	队列的应用	124
3.9.1	列车重排	124
3.9.2	投资组合问题	129
3.10	堆栈和队列基本算法的程序实现	134
3.10.1	堆栈顺序存储结构程序实现	134
3.10.2	队列顺序存储结构程序实现	139
	习题 3	144
<b>第 4 章</b>	<b>树和二叉树</b>	<b>146</b>
4.1	树、森林的概念	146
4.1.1	树的定义	146
4.1.2	树的术语	147
4.2	二叉树定义及性质	148
4.2.1	二叉树的定义	148
4.2.2	二叉树的性质	150
4.2.3	二叉树的抽象数据类型	152
4.3	二叉树的存储结构	152
4.3.1	二叉树的顺序存储	152
4.3.2	二叉树的链式存储	153
4.4	二叉树链式存储结构下的操作	154
4.4.1	二叉树的操作概念	154
4.4.2	二叉树的前序、中序、后序遍历操作	157
4.4.3	二叉树的层次遍历运算	165
4.5	线索树	168
4.5.1	线索树的概念	168
4.5.2	二叉线索树的操作	172

4.6	一般树的表示和遍历 .....	181
4.6.1	一般树的二叉链表示及其与二叉树的关系 .....	181
4.6.2	二叉树、一般树及森林的关系 .....	182
4.6.3	一般树的遍历概念 .....	183
4.6.4	一般树的运算 .....	184
4.7	树的应用 .....	186
4.7.1	分类二叉树 .....	186
4.7.2	堆树 .....	192
4.7.3	树的路径长度和赫夫曼树 .....	203
4.8	二叉树基本算法的程序实现 .....	214
	习题 4 .....	219
<b>第 5 章</b>	<b>图</b> .....	<b>223</b>
5.1	图的概念 .....	223
5.1.1	图的定义 .....	223
5.1.2	图的术语 .....	223
5.1.3	图的抽象数据类型 .....	226
5.2	图的存储结构 .....	227
5.2.1	邻接矩阵表示法 .....	227
5.2.2	邻接表表示法 .....	230
5.2.3	十字链表 .....	234
5.2.4	邻接多重表 .....	235
5.3	图的遍历 .....	237
5.3.1	深度优先搜索遍历 .....	237
5.3.2	宽度优先搜索遍历 .....	240
5.3.3	图的连通性 .....	242
5.4	最小生成树 .....	244
5.4.1	生成树 .....	244
5.4.2	最小代价生成树 .....	245
5.5	最短路径 .....	248
5.5.1	单源最短路径 .....	248
5.5.2	任意两个顶点之间的路径 .....	251
5.6	拓扑排序 .....	254
5.6.1	有向无环图 .....	254
5.6.2	AOV 网的概念 .....	255
5.6.3	AOV 网的算法 .....	256
5.7	关键路径 .....	258
5.7.1	AOE 的概念 .....	258
5.7.2	关键路径的概念 .....	258

5.7.3 关键路径的算法	259
习题 5	262
<b>第 6 章 数组、矩阵和广义表</b>	<b>266</b>
6.1 数组的定义	266
6.1.1 数组的逻辑结构	267
6.1.2 数组的抽象数据类型	268
6.2 数组的顺序表示及运算	269
6.2.1 数组的顺序存储结构	269
6.2.2 数组顺序存储结构描述	271
6.2.3 数组顺序存储结构下的操作	273
6.3 矩阵的存储及操作	274
6.3.1 矩阵的定义及操作	274
6.3.2 矩阵的顺序存储	274
6.3.3 特殊矩阵的压缩存储及操作	275
6.3.4 稀疏矩阵的压缩存储及操作	277
习题 6	291
<b>第 7 章 排序</b>	<b>294</b>
7.1 排序的基本概念	294
7.2 待排序数据对象的存储结构	296
7.3 插入排序	297
7.3.1 直接插入排序	297
7.3.2 折半插入算法	299
7.3.3 希尔排序	300
7.4 交换排序	302
7.4.1 冒泡排序	302
7.4.2 快速排序	304
7.5 选择排序	308
7.5.1 直接选择排序	308
7.5.2 堆排序	309
7.5.3 树形选择排序	310
7.6 归并排序	310
7.7 基数排序	314
7.7.1 用二维数组表示桶	316
7.7.2 用链式存储结构实现桶	317
7.8 内部排序方法比较	321
7.9 外排序	322
7.9.1 外部排序	322

7.9.2 多路平衡归并	324
习题 7	328
<b>第 8 章 查找</b>	<b>331</b>
8.1 查找的概念	331
8.2 静态查找技术	332
8.2.1 顺序查找	333
8.2.2 二分查找	334
8.2.3 分块查找	337
8.3 动态查找技术	340
8.3.1 平衡二叉树	340
8.3.2 B 树	351
8.3.3 B+树	358
8.4 哈希表的查找	359
8.4.1 基本概念	359
8.4.2 构造哈希函数的方法	360
8.4.3 哈希冲突的解决方法	362
8.4.4 哈希表的查找	364
8.4.5 哈希算法	365
8.4.6 哈希表的查找分析	368
习题 8	369
<b>第 9 章 文件</b>	<b>372</b>
9.1 外部存储设备	372
9.1.1 磁带	372
9.1.2 磁盘	373
9.1.3 光盘	374
9.1.4 闪存	374
9.2 基本概念	375
9.3 顺序文件	376
9.4 索引文件	377
9.5 索引顺序文件	378
9.6 直接存取文件	380
9.7 倒排文件	380
习题 9	381
<b>附录 A VC++6.0 编译环境介绍</b>	<b>383</b>
<b>附录 B 实践内容及要求</b>	<b>398</b>
<b>附录 C 数据结构课程实验报告格式范本</b>	<b>402</b>
<b>参考文献</b>	<b>410</b>

计算机是一门研究信息表示和处理的科学。信息表示包括组成信息的元素之间的相互关系(逻辑顺序)和信息元素在计算机中的存储方式(物理顺序)。信息处理是根据解决实际问题的需要对信息加工计算的过程。

在计算机领域中,一般所讨论的“计算”有别于通常数学概念中的计算。在计算机领域中,通常将计算又称为“运算”或“操作”。运算或操作的内涵不仅包括传统意义上的四则运算和各种函数运算(公式化运算),而且包括数据存取、插入、删除、查找、排序和遍历等运算。

从第一台电子计算机问世以来,计算机的应用主要包括两个方面:数值计算和非数值计算。计算机发展的初期,计算机主要为数值计算服务,其特点是计算过程复杂,数据类型相对简单,数据量相对较少。

随着计算机的应用深入到各个领域,大数据时代的应用方面也不再限于数值计算,更多地表现为非数值计算应用,非数值计算应用的特点表现为计算过程相对简单,数据类型相对复杂,数据的组织排列结构从某种意义上决定着非数值计算应用的有效性,数据的组织排列结构成为处理和解决数据处理问题的核心。不同的数据组织排列结构对问题的解决产生很大的影响,甚至是决定性的影响。

数值计算主要是指对一个或一组数据进行较复杂的四则运算、函数运算或迭代运算等。运算过程表现为数据处理的深入性,一般运算的原始数据对象较少。数据类型一般是数值型数据(整型、浮点型)。数值计算的程序设计主要围绕程序设计技巧,是典型地以程序为中心的设计过程。

非数值计算主要是指对类型较为复杂的大量数据进行内在联系的分析,根据处理的需要,合理地将数据按一定结构顺序进行组织存储,并完成对数据处理的程序描述。在非数值计算中,一般所处理的数据对象的类型较为复杂,通常是描述一个实体的若干个属性值的集合(结构类型或记录类型)。另外,数据具备大量性。由于处理的数据具备大量性特征,所以移动全部或部分数据将消耗大量的时间或空间。还有,对数据进行较复杂四则运算、函数运算等相对较少,较多地是对数据进行“管理”的运算,如存取、查找、排序、插入、删除、更新、统计分析等。

解决非数值计算问题,仅仅依赖程序设计的技巧已经无法达到目的,必须对这些被加工的大量数据的组织形式加以研究,针对要解决的问题,找出最佳的数据组织形式,并与合理、优秀的程序设计技巧相配合,才能达到高效处理大量数据的目的。所以,非数值计算问题是以复杂的数据为中心,研究数据的合理组织形式,并设计出基于合理数据组织结构下的高效程序。

## 1.1 什么是数据结构

数据结构是随着计算机科学的发展而建立起来的围绕非数值计算问题的一门科学。在处理非数值计算问题时,首先要建立问题的数据模型,然后设计相应的算法。数据模型包含数据的组成结构,数据间的关联方式,以对数据实施相应运算后,数据组成结构的完整性。数据组成结构的完整性是指不因对数据运算而改变数据模型的性质,运算方法本身是在保证数据组成结构的完整性的前提下,以相同规律进行的。

### 1.1.1 数据结构相关事例

为了说明什么是数据结构,先讨论现实生活中的几个例子。

**案例 1:** 电话号码簿的使用及字典的使用。

当用户拿起一本厚厚的电话号码簿,查找自己需要的单位或个人的电话号码时,一定是从电话号码簿的分类目录开始,查找相应的大类别,然后根据所查找到的大类别后面指定的页码,翻到大类别的起始页,再从特定的大类别中查找小类别,从检索到的小类别下面顺序地找到用户所要的单位或个人的电话号码,如图 1.1 所示。

行业名称	页码
党政机关	7
大学	12
企业	25
⋮	⋮
旅游	32
⋮	⋮

1

行业名称	页码
省委	55
市委	127
区委	224
⋮	⋮

7

行业名称	页码
综合大学	325
理工类大学	327
人文类大学	334
⋮	⋮

12

单位名称	电话
省委办公厅一处	88060001
省委办公厅二处	88060002
⋮	⋮

55

单位名称	电话
市委办公一处	85800203
市委办公二处	85800105
⋮	⋮

127

单位名称	电话
华中科技大学	87870001
武汉大学	86880206
⋮	⋮

325

图 1.1 电话号码簿

大、小类别可以看作电话号码簿的目录或索引。如果电话号码簿缺少类别索引,而是按照电话安装先后的顺序进行排列或毫无规律地排列,用户会使用这样的电话号码簿吗?如果电话号码簿只记载着一个部门的十几部或几十部电话,还需要分类吗?

问题就是数据量的多少。少量数据的查找无须考虑数据的组织形式,而大量数据如果没有合理的组织形式,查找过程就只能顺序地进行,效率低下,从而导致用户无法接受。在合理的数据组织结构下,用户会按照从大类别到小类别,然后在确定的小范围内顺序检索到

所需要的信息。我们将检索过程称为检索的算法,每次检索都是按照同一规律进行的。电话号码簿的变更不会改变其数据的分类结构和数据的检索算法。

可见,这个问题的数据模型是对数据的两级分类索引结构,相应检索算法是“先大类,后小类”,然后在小类中顺序查找。无论是检索还是变更,都要保证数据模型本身的完整性。

类似地,当用户在一本厚厚的汉语字典中查找某一个汉字时,首先必须知道所使用的字典的编码方法,然后才能按照偏旁部首或者拼音等相应的编码方法较快地查到需要查找的汉字。用户能如此顺利地在几万个汉字中找到所需要的汉字,是因为字典中的每一个汉字都是按偏旁部首或者拼音的规律严格地安排在其应处的页行(编码)上。倘若不按某一规律,将几万个汉字任意安排,用户为了查找某一个汉字就不得不从字典的第一页开始逐页地查找了。

可以看出,查找效率与字典中汉字的排列规律相关,也就是与数据组织形式密切相关。自然,对于不同的数据组织形式,必须采用相应的查找办法才能达到提高效率的目的。

### 案例 2: 车厢调度问题。

有若干个发往同一方向不同城市的货车车厢以随机的次序到达货站的进车道,如图 1.2 和图 1.3,整个货站由 3 部分构成:进车道(入轨)、出车道(出轨)和调度道(缓冲轨,中间多条车道)。

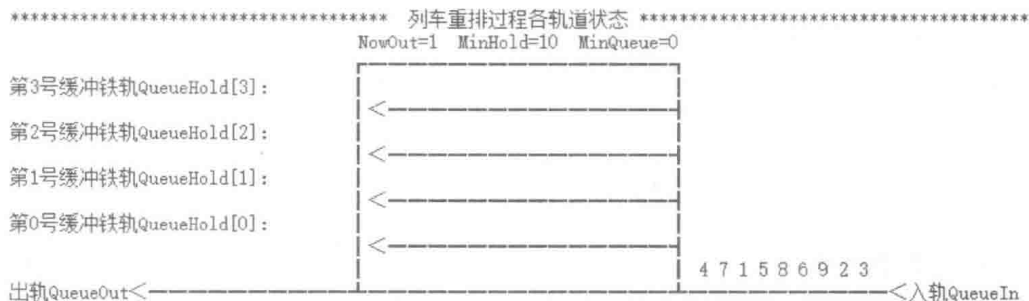


图 1.2 车厢调度转轨(初态)

由于货车的车厢从当前货站发往不同目的地,所以,货车发出本站时,列车的尾部所挂载的车厢应该是发往距本货站最近的车厢,这样可以保证到达某车站时从尾部“甩下”若干到站的车厢。

可问题是,不同的货运委托用户的车厢是以随机的次序到达货站的进车道(入轨),为了便于说明,假设发往较远站点的车厢的编号比发往较近站点的车厢的编号要小,从图 1.2 中可以看到进车道(入轨)上是不同的货运委托用户的车厢到达的次序。这些车厢发出本站前应调整它们的次序,调整为小编号在前,大编号在后的排列次序,即出车道(出轨)上的次序。

为实现调整过程,货运站就可以通过调度车道(缓冲轨)完成,调度车道由若干条缓冲铁轨组成,调度过程可描述如下:

(1) 将入轨上第一个车厢任意选择一个缓冲铁轨进入,一般选择第一个缓冲铁轨。例如,编号为 4 的车厢进入 0 号缓冲铁轨。

(2) 后面到达的车厢如果其编号大于或等于已停放了车厢的每个缓冲铁轨上的最后一

个车厢的编号,则选择与当前进入车厢编号最接近的缓冲铁轨进入。例如,编号为7的车厢进入0号缓冲铁轨,放在4号车厢的后面,因为 $7 > 4$ 。

(3) 如果下一个从入轨上进入的车厢编号正好比出轨上最后一个进入的车厢编号大1,则可将入轨上的这节车厢直接放到出轨上。例如,编号为1的车厢进入时,可以直接放到出轨上,

(4) 如果下一个从入轨上进入的车厢编号小于已停放了车厢的每个缓冲铁轨上的最后一个车厢的编号,则另选择一个没有停放车厢的缓冲铁轨进入。例如,编号为5的车厢进入1号缓冲铁轨,因为 $5 < 8$ 。

图 1.3 是编号为8的车厢进入后的状态。8号车厢进入时,因为 $8 > 7$ 且 $8 > 5$ ,选择接近的编号7所在的缓冲铁轨,编号为8的车厢进入0号缓冲铁轨,放在7号车厢的后面。如果8号车厢放在5号车厢的后面,下一个6号车厢就不能放在5号车厢的后面了,只能选择一个新的缓冲轨道进入了。多占用一个缓冲轨道,调度失败的可能性就会增加。

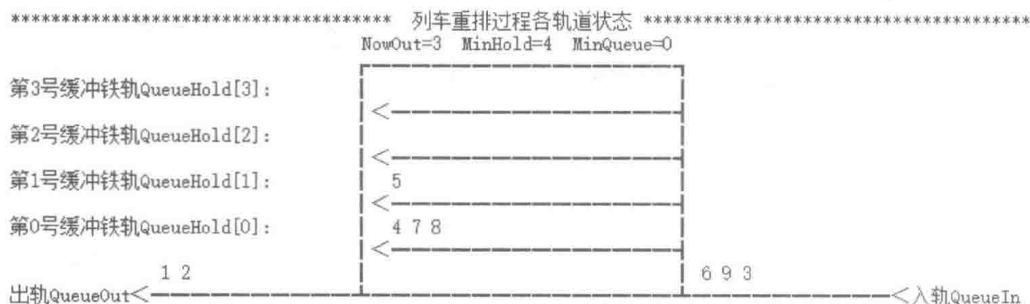


图 1.3 车厢调度转轨

这一问题是数值计算无法完成的,它属于非数值计算问题。数据在调度过程中构成了有序“队列”,对于每个车厢的进入,采用同一运算方法进入缓冲铁轨。这就是数据结构中讨论的“队列”结构问题,即从一头进入,从另一头出去。

**案例 3:** 某省各城市之间要架设电话通信线路,要保证各城市间互通,又要使架设成本最少。就是数据结构中讨论的图结构的应用——最小生成树。

图 1.4 中给出了各城市之间的距离,要实现各城市间互通,就要城市之间都有线路连接,要使成本最少,就要选择较近的城市之间架设线路,因此应该采用的架设结构如图 1.5 所示。

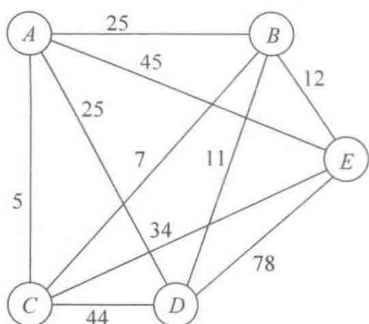


图 1.4 城市连接及距离

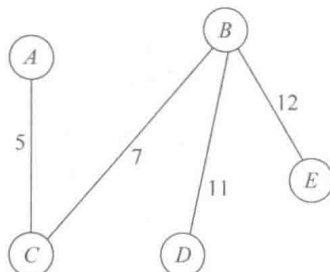


图 1.5 最小生成树



本案例中,数据的组织结构称为“图”结构,算法过程称为求取“最小生成树”。

可以看到,算法过程就是保留连接线路中值最小的连接线路,且保证任何一城市都至少与另一个城市相连接。这样既保证了城市之间“连通”,又使“连通”成本最小。

这个问题中数据的组成结构是由顶点(城市)集合和边(线路)集合构成的,求得的结果中,顶点(城市)集合没有发生改变,边(线路)集合减少了,但整个数据结构仍然是“图”结构。

对上述3个案例归纳可以发现:

第1个案例中,数据可以看作由3个表存放,表与表之间的关系是由每个表项中的页码关联的,每个表中的数据之间的关系是由前后存放的次序进行关联的。检索时,前两个表之间利用页码指向跳跃查找,在第3张表中则顺序查找。这个问题涉及大量数据,所以在处理中要合理排列数据。

第2个案例中,数据在进车道是按照先来排在前,后来排在后,不允许中间插入的原则形成一个队列。在调度过程中,将进车道中的队列分散到不同的缓冲铁轨,每个缓冲铁轨又是一个有序的子队列。最后,再形成一个有序的完整队列进入出车道。每个队列中元素以进入的先后进行关联。这个问题的处理中总是从队列的一端进入数据,从另一端移出数据。

第3个案例中,数据是由两个集合组成,“顶点”集合中存放着顶点名称,“边”集合中存放着边的长度和边所关联的是哪两个顶点,两个集合中的数据可以是无序的;运算过程是找出较小的边,且保证保留的边使所有顶点间“连通”,但不形成环形路。

从上面的3个案例可以看出,这些案例的模型都无法用数学公式、方程或函数来解决,它们都是非数值计算的问题。

### 1.1.2 数据结构的定义

综上所述,可以这样给数据结构这门学科下个定义:数据结构就是研究大量数据在计算机中存储的组织形式,并定义且实现对数据相应的高效运算,以提高计算机的数据处理能力的一门科学。

对该定义需要说明的是,数据的组织形式(结构)具有两个层面:

一个是与计算机本身无关的逻辑组织结构,简称逻辑结构,它由数据的值、数据与数据之间的关联方式两个部分组成。如电话号码簿中“大类名、大类所指的页码”“小类名、小类页码”“单位名、单位的电话号码”是数据的值,数据间的关联是由大类所指的页码和小类页码共同实现的。

另一个是将具有逻辑组织结构的数据在计算机的存储介质上如何存放的物理组织结构,简称物理结构。如电话号码簿中,每一张表内的数据可以在存储介质上连续地按顺序存放,但表与表之间是在非连续的存储介质上分开存放的。

“运算”或“操作”是数据结构讨论内容的一个核心问题。不同的实际问题具有不同的处理要求,所有的处理都要求事先定义,并用计算机的某种语言给予描述,这就是算法设计。

算法不仅要实现问题的要求,而且应该是高效地完成。低效的算法无法满足用户的需求或根本不能运用于实际。用一个高效的处理算法设计的程序结合高速运算的计算机可以最大程度地满足用户处理要求,而一个低效的处理算法设计的程序即使运用高速运算的计算机也不能满足用户的处理要求。

对实际问题的处理能力不仅取决于计算机硬件本身的处理能力,更多地取决于对数据