



世纪

信息管理与信息系统专业规划教材

可下载教学资料
<http://www.tup.tsinghua.edu.cn>

数据结构教程

王少波 孙夫雄 编著



清华大学出版社

21世纪信息管理与信息系统专业规划教材

数据结构教程

王少波 孙夫雄 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学科的必修课程。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。本书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。

本书是作者在二十多年数据结构教学经验总结的基础上编写而成的。全书共分为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、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。在附录A中还给出了VC++ 6.0编译环境介绍,在附录B中给出了本课程学习中应该完成的基本实验要求。每章的后面都附有相关的习题和部分习题答案。

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,孙夫雄编著. —北京:清华大学出版社,2011.7

(21世纪信息管理与信息系统专业规划教材)

ISBN 978-7-302-25407-2

I. ①数… II. ①王… ②孙… III. ①数据结构—高等学校—教材 IV. ①TP311.12

中国版本图书馆CIP数据核字(2011)第071749号

责任编辑:闫红梅 薛 阳

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

地 址:北京清华大学学研大厦A座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62795954, jsjic@tup.tsinghua.edu.cn

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

印 刷 者:三河市君旺印装厂

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:185×260 印 张:22.25 字 数:545千字

版 次:2011年7月第1版 印 次:2011年7月第1次印刷

印 数:1~3000

定 价:35.00元



编委会

顾 问:

马费成 武汉大学管理学院

张金隆 华中科技大学管理学院

主 任: 刘腾红 中南财经政法大学信息学院

副主任: 杨 超 华中科技大学管理学院

唐晓波 武汉大学管理学院

编 委: (按姓氏笔画为序)

马学才 华中农业大学信息管理系

王伟军 华中师范大学信息管理系

王 虎 武汉理工大学管理学院

邓仲华 武汉大学管理学院

史金平 湖北大学商学院

孙细明 武汉工程大学管理学院

宋克振 中南财经政法大学信息学院

张志清 武汉科技大学管理学院

张新兰 中国地质大学管理学院

沈先钊 湖北工业大学信息管理系

李禹生 武汉工业大学计算机与信息工程系

郑双怡 中南民族大学管理学院

柯 林 江汉大学商学院

鲁耀斌 华中科技大学管理学院



根据国家教育部 1998 年 7 月 6 日公布的《普通高等学校本科专业目录》的内容,将原经济信息管理、图书情报学、科技信息管理、林业信息管理和信息管理等专业合并为管理学科门类中的信息管理与信息系统专业。目前,我国已有二百多所高等院校设置了信息管理与信息系统专业。该专业的发展伴随着世界信息化的发展而发展,为我国培养了大量的信息化专门人才。

网络化、信息化、全球经济一体化是当今世界的主要特征。20 世纪 90 年代,信息技术不断创新,信息产业持续发展,信息网络广泛普及,信息化成为全球经济社会发展的显著特征,并逐步向一场全方位的社会变革演变。21 世纪,信息化对经济社会发展的影响更加深刻,信息资源日益成为重要的生产要素、无形资产和社会财富。我国信息化发展的进展十分迅速。

基于此,信息管理与信息系统专业人才培养的任务十分艰巨。首先要定位,再定向,还要定措施。不同的高校要根据自己的特色来定位,如:以经、法、管理为主的综合性人文社科大学,其信息管理与信息系统专业就要定位在和经济、法律、管理的结合上,培养的人才主要适合在经济管理部门、司法部门、企事业单位等从事信息系统建设和管理以及科学研究等工作。定向的具体内容由培养目标来确定,本专业直接以信息化建设的人才需求为培养目标与标准,培养熟练掌握现代信息技术手段和方法,具有坚实的现代管理科学理论知识,具备较强的计算机应用能力的综合型、实用型的高级专门人才。定措施则是要确定对培养目标的具体实施过程和方法,包括师资要求、全程教学计划和教材建设等。

现各个高校在信息管理与信息系统专业的教材使用上五花八门,教材主要由任课教师自己选定。计算机方面的教材主要选用计算机科学与技术专业的教材,管理方面的教材主要采用管理学科的教材。尽管近年来一些出版社陆续出版了几套信息管理与信息系统专业的教材,但仍然不能满足教学的需要。根据教育部 1998 年信息管理与信息系统专业课程要求,结合中国高等院校信息系统学科课程体系课题组撰写的《中国高等院校信息系统学科课程体系 2005》(征求意见稿)(清华大学出版社,2005 年 11 月)的内容,我们组织长期从事信息管理与信息系统专业教学和研究的教师,在清华大学出版社的大力支持下,经过多次讨论和研究,组织编委会,制定教材编写规划,审定编写大纲,并采取主编负责制,层层把关,力争使本套教材成为具有系统性、完备性的高水平、高质量的信息管理与信息系统专业教材。

本套教材的主要特点是:

1. 系统性。教材自成体系,系统地体现本专业的知识体系和结构。
2. 完整性。教材能完整、准确地反映本专业的教学内容,满足培养高层次人才的需要。
3. 新颖性。教材要反映本学科的最新发展动态和研究成果。

4. 理论性。教材注重理论基础的培养,使学生具备扎实的理论知识。

5. 实用性。教材注重理论与实践结合,把培养学生分析问题、解决问题的实际动手能力作为一项重要的内容予以体现。

本套教材的成功出版,凝聚了众多长期从事信息管理与信息系统专业建设的专家、学者及相关人员的心血。我们殷切希望从事信息管理与信息系统专业的教育工作者对本套规划教材提出宝贵建议,使教材质量不断得到提高。让我们共同为培养高素质的信息化人才而努力。

刘腾红 教授
本规划教材编委会主任
2007年8月



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。本书还可以作为计算机科技工作者及其有关专业人员的参考书。

“数据结构”发展至今,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。数据结构和算法的研究是计算机基础性研究之一,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、计算机辅助设计及其他信息管理的应用。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,受教育者通过《数据结构教程》的学习,不仅能从中学到专业知识,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,这门学科只有被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。计算机系统全面体现着数据结构所在,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。叙述时回避了复杂的数学定义与推导,力求通俗易懂。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。

本书系作者集多年来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。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。

数据结构不但是是一门理论性很强的课程,同时又是一门实践性很强的课程,因此在每一章的最后都安排了适当的习题,供读者练习。

本书共分 9 章,介绍了线性表和串、栈和队列、树、图、数组、排序、查找、文件等基本数据



结构、算法,提供了要应的应用例题。

本书由王少波、孙夫雄编著。王少波负责编写第1~4章及附录B;孙夫雄负责编写第5~9章及附录A,全书由王少波负责统稿和总纂。

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,书中错误与疏漏之处,恳请读者不吝指教。

编 者

2010年11月



目录

第 1 章 绪论	1
1.1 什么是数据结构	1
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	5
1.2.1 数据和信息	5
1.2.2 数据元素	6
1.2.3 结构类型	6
1.2.4 静态存储空间分配和动态存储空间分配	9
1.3 数据类型、抽象数据类型和数据结构	11
1.3.1 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
1.3.4 抽象数据类型的三元组表示	13
1.4 算法及算法分析、算法描述	14
1.4.1 算法和程序	14
1.4.2 程序性能和算法效率	15
1.4.3 算法分析	17
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	30
2.3 简单链表存储结构及操作	34
2.3.1 简单链表的存储	34
2.3.2 简单链表的操作	36

2.4	双向链表	42
2.4.1	双向链表的存储	42
2.4.2	双向链表的操作	43
2.5	单向循环链表和双向循环链表	45
2.5.1	单向循环链表的存储	45
2.5.2	双向循环链表的存储	46
2.6	模拟指针方式构造简单链表	47
2.6.1	模拟链表的存储	47
2.6.2	模拟链表的操作	48
2.7	多重链表	51
2.8	链表应用	53
2.8.1	节点移至表首运算	53
2.8.2	链表的逆向运算	54
2.8.3	多项式的相加运算	55
2.8.4	十字链表结构的应用	59
2.8.5	一个较复杂的机票售票系统的数据结构方案	61
2.9	串	63
2.9.1	串的定义	63
2.9.2	串的逻辑结构及运算	63
2.9.3	串的顺序存储结构	64
2.9.4	串的链式存储结构	64
习题 2		65
第 3 章	栈与队列	68
3.1	堆栈的定义	68
3.1.1	堆栈的逻辑结构	68
3.1.2	堆栈的抽象数据类型	68
3.2	堆栈的顺序存储及操作	69
3.2.1	堆栈顺序存储	69
3.2.2	堆栈顺序存储结构下的操作	70
3.3	堆栈的链式存储及操作	72
3.3.1	堆栈的链式存储	72
3.3.2	链式栈的操作	73
3.4	多个栈共享邻接空间	75
3.5	堆栈的应用	76
3.5.1	检验表达式中括号的匹配	76
3.5.2	表达式的求值	78
3.5.3	背包问题求解	79
3.5.4	地图四染色问题求解	81

3.6	队列的定义	86
3.6.1	队列的逻辑结构	86
3.6.2	队列的抽象数据类型	86
3.7	队列的顺序存储及操作	87
3.7.1	队列顺序存储	87
3.7.2	队列顺序存储结构下的操作	90
3.8	队列的链式存储及操作	92
3.8.1	队列的链式存储	92
3.8.2	链式队列的操作	93
3.9	队列的应用	95
3.9.1	列车重排	95
3.9.2	投资组合问题	98
习题 3		101
第 4 章	树和二叉树	103
4.1	树、森林的概念	103
4.1.1	树的定义	103
4.1.2	树的术语	103
4.2	二叉树的定义及性质	105
4.2.1	二叉树的定义	105
4.2.2	二叉树的性质	107
4.2.3	二叉树的抽象数据类型	109
4.3	二叉树的存储结构	109
4.3.1	二叉树的顺序存储	109
4.3.2	二叉树的链式存储结构	111
4.4	二叉树链式存储结构下的操作	111
4.4.1	二叉树的操作概念	111
4.4.2	二叉树的前序、中序、后序遍历操作	113
4.4.3	二叉树的层次遍历操作	120
4.4.4	二叉树的其他操作	123
4.5	线索树	125
4.5.1	线索树的概念	125
4.5.2	二叉线索树的操作	128
4.6	一般树的表示和遍历	137
4.6.1	一般树的二叉链表示以及它与二叉树的关系	137
4.6.2	二叉树、一般树及森林的关系	139
4.6.3	一般树的遍历概念	139
4.6.4	一般树的运算	140
4.7	树的应用	142

4.7.1 分类二叉树	142
4.7.2 堆树	147
4.7.3 树的路径长度和哈夫曼树	158
习题 4	168
第 5 章 图	172
5.1 图的概念	172
5.1.1 图的定义	172
5.1.2 图的术语	172
5.1.3 图的抽象数据类型	175
5.2 图的存储结构	176
5.2.1 邻接矩阵表示法	176
5.2.2 邻接表表示法	178
5.2.3 十字链表	181
5.2.4 邻接多重表	182
5.3 图的遍历	184
5.3.1 深度优先搜索遍历	184
5.3.2 宽度优先搜索遍历	187
5.3.3 图的连通性	189
5.4 最小生成树	192
5.4.1 生成树	192
5.4.2 最小代价生成树	192
5.5 最短路径	196
5.5.1 单源最短路径	196
5.5.2 任意两个顶点之间的路径	198
5.6 拓扑排序	201
5.6.1 有向无环图	201
5.6.2 AOV 网的概念	202
5.6.3 AOV 网的算法	204
5.7 关键路径	206
5.7.1 AOE 的概念	206
5.7.2 关键路径的概念	207
5.7.3 关键路径的算法	207
习题 5	210
第 6 章 数组、矩阵和广义表	214
6.1 数组的定义	214
6.1.1 数组的逻辑结构	215
6.1.2 数组的抽象数据类型	216

6.2	数组的顺序表示及运算	217
6.2.1	数组的顺序存储结构	217
6.2.2	数组顺序存储结构描述	219
6.2.3	数组顺序存储结构下的操作	221
6.3	矩阵的存储及操作	222
6.3.1	矩阵的定义及操作	222
6.3.2	矩阵的顺序存储	222
6.3.3	特殊矩阵的压缩存储及操作	223
6.3.4	稀疏矩阵的压缩存储及操作	225
习题 6		238
第 7 章	排序	241
7.1	排序的基本概念	241
7.2	待排序数据对象的存储结构	243
7.3	插入排序	244
7.3.1	直接插入排序	244
7.3.2	折半插入算法	245
7.3.3	希尔排序	246
7.4	交换排序	248
7.4.1	冒泡排序	248
7.4.2	快速排序	250
7.5	选择排序	253
7.5.1	直接选择排序	253
7.5.2	堆排序	255
7.5.3	树形选择排序	255
7.6	归并排序	256
7.7	基数排序	259
7.7.1	用二维数组表示桶	260
7.7.2	用链式存储结构实现桶	262
7.8	内部排序方法比较	264
7.9	外排序	265
7.9.1	外部排序	265
7.9.2	多路平衡归并	267
习题 7		271
第 8 章	查找	274
8.1	查找的概念	274
8.2	静态查找技术	275
8.2.1	顺序查找	276

8.2.2 二分查找	277
8.2.3 分块查找	280
8.3 动态查找技术	283
8.3.1 平衡二叉树	283
8.3.2 B-树	294
8.3.3 B+树	301
8.4 哈希表的查找	301
8.4.1 基本概念	301
8.4.2 构造哈希函数的方法	302
8.4.3 哈希冲突的解决方法	304
8.4.4 哈希表的查找	306
8.4.5 哈希算法	307
8.4.6 哈希表的查找分析	309
习题 8	310
第 9 章 文件	313
9.1 外部存储设备	313
9.1.1 磁带	313
9.1.2 磁盘	314
9.1.3 光盘	315
9.1.4 闪存	315
9.2 基本概念	316
9.3 顺序文件	317
9.4 索引文件	318
9.5 索引顺序文件	319
9.6 直接存取文件	320
9.7 倒排文件	321
习题 9	322
附录 A VC++ 6.0 编译环境介绍	324
附录 B 实践内容及要求	338
参考文献	342



第1章

绪论

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”有别于数学概念中的一般计算,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),而且还包括数据存取、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,应用方面不再局限于数值计算的应用,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,数据的组织排列从某种意义上决定着非数值计算应用的有效性。

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,是典型的以程序为中心的设计过程。

程序设计中的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,所处理的数据对象的类型一般较为复杂,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,数据是大量的,由于数据的大量性,所以移动全部或部分数据将消耗大量的时间或空间。还有,对数据进行较复杂的四则运算、函数运算等相对较少,较多地是对数据进行“管理”运算,如存取、查找、排序、插入、删除、更新等。

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提高效率的目的,必须对这些被加工数据的组织形式加以研究,找出最佳的数据组织形式,并与好的程序设计技巧相配合,才能达到提高效率的目的。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,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学问,而非数值计算是先于计算机科学存在的学问。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,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,下面先讨论现实生活中的几个例子。

问题 1: 电话号码簿和字典的使用问题。

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。

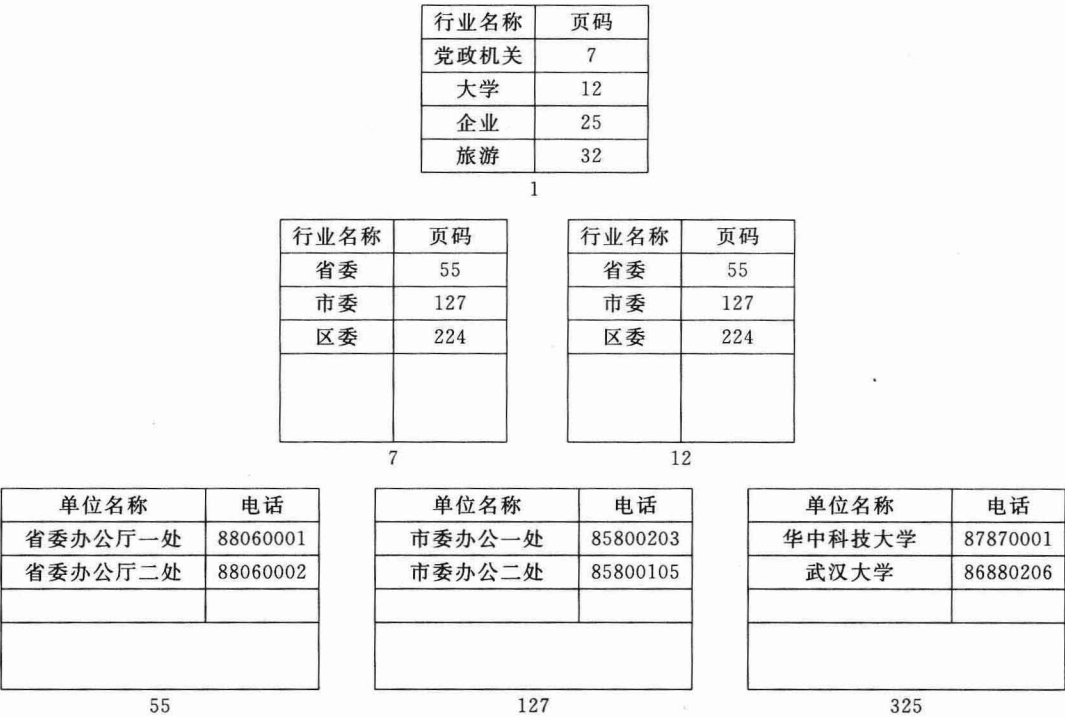


图 1.1 电话号码簿

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,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所需要查找的汉字。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它应处的页行(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。

可以看出查找效率是与字典中汉字的安排规律,也就是数据组织形式密切相关的。自然,对于不同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。

问题 2: 车皮调度问题。

有若干个发往同一方向不同城市的货车车皮以随机的次序到达货站的进车道,如图 1.2 所示,整个货站由三部分构成:进车道、出车道和调度车道(中间多条车道)。



图 1.2 车厢调度转轨

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

可问题是,不同的货运委托用户的车皮是以随机的次序到达货站的进车道的,为了便于说明,我们假设发往较远站点的车皮的编号比发往较近站点的车皮的编号要小,从图 1.2 中可以看到进车道上是不同的货运委托用户的车皮到达的次序。这些车皮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,即出车道上的次序。

为实现调整过程,货运站通过调度车道完成调度,调度车道由若干条缓冲铁轨组成,调度过程可描述为:

(1) 将入轨上第一个车皮任意选一个缓冲铁轨进入,一般选择第一个缓冲铁轨。

如:编号为 3 的车皮进入①号缓冲铁轨。

(2) 后面到达的车皮,如果其编号大于或等于已停放了车皮的每个缓冲铁轨上的最后一个车皮的编号,则选择与当前进入车皮编号最接近的缓冲铁轨进入。

如:编号为 5 的车皮进入①号缓冲铁轨,放在 3 号车皮的后面,因为 $5 > 3$ 。

(3) 后面到达的车皮,如果其编号小于已停放了车皮的每个缓冲铁轨上的最后一个车皮的编号,则另选择一个没有停放车皮的缓冲铁轨进入。

如:编号为 4 的车皮进入②号缓冲铁轨,因为 $4 < 5$ 。编号为 7 的车皮进入时,因为 $7 > 5$ 且 $7 > 4$,选择接近的编号 5 所在的缓冲铁轨,编号为 7 的车皮进入①号缓冲铁轨,放在 5 号