

DATA STRUCTURES AND ALGORITHMS

数据结构与算法

朱明方 吴及 编著

清华大学出版社

数据结构与算法

朱明方 吴及 编著

清华大学出版社
北京

内 容 简 介

本书是为非计算机专业开设数据结构与算法课程需要而编写的。作者在编写过程中力求做到内容精炼,同时反映该课程的新发展,知识点的介绍和实际应用紧密结合,使读者易学易用。

针对非计算机专业安排的学时少,学生相关基础知识相对薄弱且一般不独立开设算法设计课等特点,本教材从应用的角度,有重点地介绍数据处理中常用的数据结构——线性表、树与二叉树、图以及基本的数据处理技术——查找和排序方法。同时,把回溯法、分治法、贪心法、动态规划法等常用的算法设计方法融入其中,把数据结构介绍和算法设计讨论紧密结合。让读者更具体、更深刻地理解各种常用的数据结构及它们与算法之间的关系,从而学以致用。

本书可作为普通高等院校数据结构课程教材,也可供从事计算机应用开发的科技人员参考。

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/朱明方,吴及编著. —北京:清华大学出版社,2010.3

ISBN 978-7-302-21994-1

I. ①数… II. ①朱… ②吴… III. ①数据结构—算法分析—高等学校—教材

IV. TP311.12

中国版本图书馆 CIP 数据核字(2010)第 007758 号

责任编辑:王一玲 刘佩伟

责任校对:梁毅

责任印制:孟凡玉

出版发行:清华大学出版社

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

社 总 机:010-62770175

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

地 址:北京清华大学学研大厦 A 座

邮 编:100084

邮 购:010-62786544

印 装 者:北京嘉实印刷有限公司

经 销:全国新华书店

开 本:185×260 印 张:23.5

字 数:585 千字

版 次:2010 年 3 月第 1 版

印 次:2010 年 3 月第 1 次印刷

印 数:1~3000

定 价:35.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。
联系电话:010-62770177 转 3103 产品编号:029634-01

前言

随着信息技术的发展,计算机应用能力已经成为高等院校非计算机专业学生特别是理工科专业学生必须具备的基本能力之一。作为计算机应用的重要基础——数据结构,已经不只是计算机专业的重要专业基础课,也是许多非计算机专业特别是工科专业重要的技术基础课。但与计算机专业不同,对于非计算机专业,该课程的学时相对要少,先修课程的基础也不一样,而且一般都不再单独设置“算法设计”类课程。因此,如何针对非计算机专业的情况,能够在学时和基础有限的前提下,让学生较快地掌握数据结构和算法设计的基本知识,具备解决一般数据处理问题的初步能力,是非计算机专业的数据结构课程教学中非常重要而又十分具有挑战性的课题,教材建设则是解决这一问题的关键所在。

本教材是作者基于多年来为电子信息相关专业讲授数据结构与算法课程的教学实践,经过不断的改进编写而成的。本教材的特点是:针对非计算机专业学生的情况,从应用的角度介绍常用的数据结构,并且把数据结构应用和常用算法设计方法的讨论紧密结合,使常用的算法设计思路很自然地融入其中。教材中对各部分内容的介绍突出重点、力求精炼实用,对各种数据结构和算法设计方法的讨论力求深入浅出,并通过应用例子使其与实际问题相结合,从而使读者很容易把握书中的脉络,理解和掌握课程的重要知识点。同时本书以面向对象的观点来讨论数据结构,使其与该课程发展的新要求同步。

为方便非计算机专业仅有C语言基础的学生,教材中采用C/C++作为算法描述语言,只要具备C语言程序设计的基础就能阅读教材中的算法。同时为了兼顾有C++基础的学生理解数据结构与“类”的关系,教材中对C++的主要特点作了必要的归纳,并对各种数据结构给出了类的描述。

建议本教材的授课课时不少于48学时;标有“*”的章节,为可选内容,由教师视具体情况选定。若这些内容不作为课程要求,则建议授课课时为32~36学时,同时应该安排足够的包括上机大作业和实验在内的实习学时相配合。

本教材第1~4章由朱明方编写,第5~7章由吴及编写,朱峰做了算法的修改调试工作,全书由朱明方统稿。为配合学习,有《数据结构与算法——习题解答与实习指导》一书与本教材配套。

本教材的编写得到了清华大学教务处和清华大学出版社的大力支持,在此表示衷心的感谢。

限于作者的水平,书中疏漏之处难免,敬请读者批评指正。

作者

2009年10月

教师反馈表

感谢您购买本书！清华大学出版社计算机与信息分社专心致力于为广大院校电子信息类及相关专业师生提供优质的教学用书及辅助教学资源。

我们十分重视对广大教师的服务，如果您确认将本书作为指定教材，请您务必填好以下表格并经系主任签字盖章后寄回我们的联系地址，我们将免费向您提供有关本书的其他教学资源。

您需要教辅的教材：			
您的姓名：			
院系：			
院/校：			
您所教的课程名称：			
学生人数/所在年级：	_____人/ 1 2 3 4 硕士 博士		
学时/学期	_____学时/_____学期		
您目前采用的教材：	作者：_____ 书名：_____ 出版社：_____		
您准备何时用此书授课：			
通信地址：			
邮政编码：		联系电话	
E-mail：			
您对本书的意见/建议：		系主任签字	
		盖章	

我们的联系地址：

清华大学出版社 学研大厦 A602, A604 室

邮编：100084

Tel: 010-62770175-4409, 3208

Fax: 010-62770278

E-mail: liuli@tup.tsinghua.edu.cn; hanbh@tup.tsinghua.edu.cn



第1章 绪论	1
1.1 预备知识	1
1.1.1 数据抽象	1
1.1.2 数据抽象与二元关系	3
1.1.3 二元关系的基本性质和几种重要的关系	5
1.2 什么是数据结构	6
1.2.1 数据结构的引出	6
1.2.2 数据的逻辑结构和存储结构	8
1.2.3 数据结构的表示	11
1.3 抽象数据类型	11
1.3.1 什么是抽象数据类型	11
1.3.2 面向对象方法与抽象数据类型	13
1.3.3 抽象数据类型的实现	15
1.4 算法与算法分析	17
1.4.1 什么是算法	17
1.4.2 算法描述与算法描述语言	19
1.4.3 常用的算法设计方法	29
1.4.4 算法分析	35
习题	38
第2章 线性表及其顺序存储	42
2.1 线性表的概念	42
2.1.1 什么是线性表	42
2.1.2 线性表的抽象数据类型	44
2.2 线性表的顺序存储及其运算实现	45
2.2.1 线性表的顺序存储——顺序表	45
2.2.2 顺序表的基本运算	47
2.2.3 顺序表的类定义	52
2.3 栈	53
2.3.1 什么是栈	53
2.3.2 栈的抽象数据类型	56

2.3.3	栈的顺序存储及其运算	56
2.3.4	顺序栈的类定义	59
2.4	栈与算法设计	60
2.4.1	栈与优先级处理	60
2.4.2	栈与回溯法	68
2.4.3	栈与分治法	73
2.4.4	栈与递归	75
2.5	队列	85
2.5.1	队列及其抽象数据类型	85
2.5.2	顺序队列及其运算	87
2.5.3	队列应用例	92
*2.5.4	优先队列	96
2.6	数组与特殊矩阵的表示	98
2.6.1	数组的顺序存储	98
2.6.2	规则矩阵的压缩存储	100
*2.6.3	稀疏矩阵的三列二维数组表示——三元组顺序表	102
习题		105

第3章 链表 107

3.1	线性表的链式存储——线性链表	107
3.1.1	线性链表的概念	107
3.1.2	线性链表的运算	109
3.1.3	线性链表的类定义	116
3.2	链式栈与链式队列	116
3.2.1	链式栈	116
3.2.2	链式队列	120
3.3	循环链表	123
3.3.1	循环链表的结构特点	123
3.3.2	循环链表的基本运算	124
3.3.3	循环链表应用例	128
*3.4	双向链表与十字链表	136
3.4.1	双向链表	136
3.4.2	稀疏矩阵的十字链表表示	139
*3.5	广义表	140
3.5.1	广义表的概念	141
3.5.2	广义表的存储方式	143
3.5.3	广义表的基本运算	144
习题		148

第4章 树与二叉树	151
4.1 树的基本概念	151
4.1.1 什么是树结构	151
4.1.2 树的定义与表示	154
4.1.3 树的性质	156
4.2 二叉树	157
4.2.1 二叉树的定义	157
4.2.2 二叉树的基本性质	158
4.2.3 二叉树的抽象数据类型	160
4.2.4 二叉树的存储结构	161
4.2.5 二叉树的遍历及其他运算	163
*4.2.6 线索二叉树	168
*4.2.7 二叉树的计数	172
4.3 二叉树应用	174
4.3.1 表达式线性化	174
4.3.2 最优二叉树	176
4.3.3 二叉搜索树	182
4.3.4 堆	188
4.3.5 二叉树与减治法	195
4.4 树的运算	196
4.4.1 树的抽象数据类型	197
4.4.2 树的存储结构	197
4.4.3 树的遍历	199
4.4.4 树的其他运算——树遍历的应用	200
4.5 树结构与算法设计	203
4.5.1 算法的一种描述方式——决策树	203
4.5.2 树与回溯法	204
4.5.3 树与划分等价类	210
*4.6 森林与二叉树	213
4.6.1 森林与二叉树的转换	213
4.6.2 森林的遍历	214
习题	215
第5章 图	217
5.1 图的基本概念	217
5.1.1 图的定义和概念	217
5.1.2 图的抽象数据类型	221
*5.1.3 欧拉路径和汉密尔顿路径	222

5.2 图的存储结构	224
5.2.1 图的邻接矩阵表示	224
5.2.2 图的邻接表表示	226
*5.2.3 图的其他表示方法	229
5.3 图的遍历	231
5.3.1 图的深度优先遍历	232
5.3.2 图的广度优先遍历	233
5.3.3 图遍历的应用	234
*5.3.4 广义图搜索	236
*5.3.5 图的连通性	237
*5.4 有向图与有向无环图	238
5.4.1 有向图的连通性和传递闭包	238
5.4.2 有向无环图与拓扑排序	241
5.4.3 关键路径	244
5.5 最小生成树与贪心算法	246
5.5.1 图的生成树与最小生成树	246
5.5.2 普里姆算法	248
5.5.3 克鲁斯卡尔算法	250
5.5.4 贪心算法	252
5.6 最短路径问题	256
5.6.1 单源最短路径	256
*5.6.2 带负权值边的单源最短路径	258
5.6.3 全源最短路径	261
5.6.4 动态规划算法	264
5.7 图应用例——城市间公路交通网问题	269
5.7.1 问题描述	269
5.7.2 问题求解思路	270
习题	270
第6章 查找	273
6.1 线性查找表	274
6.1.1 顺序查找	274
6.1.2 折半查找	274
*6.1.3 斐波那契查找	276
6.1.4 线性查找表的性能比较	276
6.2 二叉搜索树的查找性能	277
6.3 AVL 树	279
6.3.1 BST 的旋转操作	280
6.3.2 AVL 树的插入和平衡化旋转	281

* 6.3.3	AVL 树的删除	283
* 6.3.4	AVL 树的性能	285
6.4	B-树	286
6.4.1	多路动态搜索树	286
6.4.2	B-树的查找	287
6.4.3	B-树的插入	287
* 6.4.4	B-树的删除	289
* 6.5	2-3-4 树和红黑树	291
6.5.1	2-3-4 树	291
6.5.2	红黑树	292
6.6	散列方法	296
6.6.1	散列技术	296
6.6.2	散列函数	297
6.6.3	冲突处理	300
6.6.4	散列的删除	302
6.6.5	散列的性能	302
6.7	静态索引结构	303
6.7.1	索引查找	303
6.7.2	索引存储方式	303
* 6.7.3	索引文件结构	306
6.8	模式匹配算法	309
6.8.1	字符串的概念和 ADT	309
6.8.2	字符串的存储表示	310
6.8.3	字符串的模式匹配及简单匹配算法	311
6.8.4	KMP 算法	311
* 6.8.5	Boyer-Moore 算法	314
习题		317
第 7 章	排序	319
7.1	排序的概念及算法性能分析	319
7.2	基本排序方法	321
7.2.1	冒泡排序	321
7.2.2	插入排序	322
7.2.3	直接选择排序	326
7.2.4	基本排序方法的比较	328
7.3	快速排序	328
7.3.1	快速排序的过程	329
7.3.2	快速排序的性能分析	330
* 7.3.3	快速排序的改进算法	331



*7.3.4 三路划分的快速排序算法	332
7.4 归并排序	334
7.4.1 二路归并	334
7.4.2 自底向上的归并排序	335
7.4.3 自顶向下的归并排序	336
*7.5 锦标赛排序	337
7.6 堆排序	339
7.6.1 堆排序的思想	339
7.6.2 堆排序的实现	342
7.7 内排序方法分析	343
*7.7.1 排序方法的下界	343
7.7.2 内排序方法的比较	345
7.8 线性时间复杂度的排序算法	346
*7.8.1 计数排序	346
*7.8.2 箱排序	347
7.8.3 基数排序	348
*7.9 排序网络	351
7.9.1 排序网络的概念	351
7.9.2 巴彻尔奇偶归并排序	353
7.10 外部排序	356
7.10.1 外部排序方法	357
*7.10.2 基于败者树的 k 路归并算法	357
*7.10.3 排序-归并的改进	359
习题	363
参考文献	365



第1章

绪 论

当今是信息爆炸的时代,具有时代性标志的技术——计算机应用已经成为人们工作、学习和生活的重要组成部分,但其中绝大部分应用是处理非数值计算问题,或者是解决数据处理问题。与数值计算问题相比,数据处理问题的数据量大得多,结构也更复杂,因而数据的组织对处理效率的影响也就更大,因此人们对计算机所处理的数据的组织形式和相互关系的研究也越来越多、越来越深入,这也正是“数据结构”这门学科得以迅速发展的原因。

数据结构问题的提出,来源于程序设计技术的发展。对数据结构的研究是为了提高程序运行的效率(时间效率、空间效率、复用程度等)。数据结构所讨论的问题主要包括:数据集合中数据元素之间的逻辑关系,它们在计算机存储器中的表示以及基本运算的实现。

本章作为后续章节的预备知识,首先从实际问题的数据抽象入手,引出二元关系的定义及其在问题描述中的实际意义,然后介绍有关数据结构的基本概念,最后讨论算法及其性能评价问题。

1.1 预备知识

1.1.1 数据抽象

对数据处理问题而言,在计算机上解决问题有以下步骤:

建立数学模型——组织数据;确定求解方法——编码、处理。

显然,解决数据处理问题首先要把实际问题转化为数学模型,即对实际问题进行数据抽象,包括对问题中所涉及的实体和它们之间联系的抽象。实体是指问题中所涉及到的事、物、人,也就是客观事物,对它们抽象的结果就是问题中所要涉及的数据;实体间的联系抽象得到的是数据元素之间的关系。在经过数据抽象建立起数学模型的基础上,再确定数据的表示以及相应的求解方法,即确定数据的存储表示和采用的算法。最后进行编码写出解决问题的程序。

可以看出,数据抽象和数据组织是数据处理中要解决的基本问题,也是核心的问题。为了对问题作数据抽象,要分析问题中实体的特性,归纳出实体之间存在的或者是需要建立的

联系,从而为合理、有效地表示它们,高效率地进行处理确立前提。请看以下问题。

问题 1.1 统计全年级学生的成绩,要求:

- ① 按成绩排序输出全年级学生的姓名、学号、班级、成绩;
- ② 分班级,每班按成绩排序输出学生的姓名、学号、成绩。

问题分析与数据抽象:

学生集合——若干学生记录(姓名、学号、班级、成绩);

学生之间的成绩高低顺序——数据元素间的关系。

抽象结果:按成绩排序的学生成绩表。

问题 1.2 假设运动会比赛项目分类编号如下:

比赛项目:	100 米	200 米	400 米	800 米	1500 米	3000 米	4×100 米	4×200 米或 4×400 米
男子项目编号:	A ₁₁	A ₁₂	A ₁₃	A ₁₄	A ₁₅	A ₁₆	A ₁₇	A ₁₈
女子项目编号:	A ₂₁	A ₂₂	A ₂₃	A ₂₄	A ₂₅	A ₂₆	A ₂₇	A ₂₈
比赛项目:	跳高	撑竿跳高	跳远	三级跳远	铅球	铁饼	标枪	
男子项目编号:	B ₁₁	B ₁₂	B ₁₃	B ₁₄	C ₁₁	C ₁₂	C ₁₃	
女子项目编号:	B ₂₁	B ₂₂	B ₂₃	B ₂₄	C ₂₁	C ₂₂	C ₂₃	

比赛顺序的安排原则是:

- ① 100 米、200 米项目在其他比赛项目之前进行,接力赛在其他项目比赛结束之后进行;
- ② 中长跑项目、跳高跳远、铅球可以同时进行;
- ③ 铁饼、标枪项目安排在 400 米、800 米、1500 米、3000 米之后进行;
- ④ 径赛项目顺序:100 米、200 米、400 米、800 米、1500 米、3000 米;
- ⑤ 跳高跳远项目同时进行;

跳高项目顺序:跳高、撑竿跳高;

跳远项目顺序:跳远、三级跳远;

⑥ 投掷类项目顺序:铅球、铁饼、标枪;

⑦ 同名项目安排顺序:男子项目、女子项目。

要求根据预定的项目安排原则排出比赛顺序。

问题分析与数据抽象:

比赛项目(编号)——数据元素;

比赛顺序原则——数据元素之间的前后关系。

抽象结果:数据元素之间的关系图(图结构)。

数据元素之间的前后关系如图 1-1 所示。

存储表示与选择算法:

存储得到的图,采用拓扑分类方法处理——以顺序存储表示实现。

处理结果:比赛顺序安排表。

从上述两个简单的例子中可以看出,无论问题简单还是复杂,为了建立起解决问题的数学模型,必须归纳出数据元素之间的联系,必要时还可能需要人为地把数据元素按照某种要

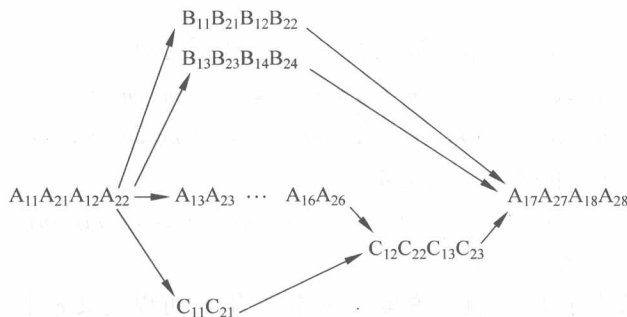


图 1-1 比赛顺序安排问题的数据抽象

求联系起来,这首先就会涉及到数据元素之间关系的描述问题。

1.1.2 数据抽象与二元关系

经过数据抽象建立起来的数据元素之间的联系,归纳起来有三种,即 1 对 1 的联系,1 对多的联系和多对多的联系。无论是哪种联系,都可以借助于“二元关系”简单明了地描述它。也就是说,“二元关系”是描述数据元素之间的联系的基础。

二元关系是一个数学概念,它定义于集合的基本运算——笛卡儿积 (Cartesian Product) 的基础上。为了便于理解,我们从集合的笛卡儿积的定义出发,进而说明二元关系的概念及其一般性质。

1. 集合的笛卡儿积

设有集合 A 和 B , 则集合 A 对集合 B 的笛卡儿积记做 $A \times B$, 定义为

$$A \times B = \{ \langle a, b \rangle \mid a \in A \text{ 且 } b \in B \}$$

根据定义,若有 $A = \{a_1, a_2\}$, $B = \{b_1, b_2\}$, 则有

$$A \times B = \{ \langle a_1, b_1 \rangle, \langle a_1, b_2 \rangle, \langle a_2, b_1 \rangle, \langle a_2, b_2 \rangle \}$$

其中, $\langle a_1, b_1 \rangle$ 等都是有序对 (序偶), a_1 是有序对的第一分量, b_1 是有序对的第二分量, 必须注意它的有序性。根据这个特点, 显然有

$$A \times B \neq B \times A$$

同样地, 有

$$(A \times B) \times C \neq A \times (B \times C)$$

当然, 笛卡儿积与集合的并、交运算结合起来, 还可以得到其他的性质。

实际上, 只要理解了两个集合的笛卡儿积的定义, 也就很容易理解 n 个集合的笛卡儿积的概念了。

2. 二元关系

(1) 二元关系的数学定义

二元关系是一个数学概念, 它的定义为

设有集合 M, N , 则 $M \times N$ 的任意一个子集 R 称为 M 到 N 的一个二元关系。

如果 $M = N$, 则称 R 为 M 上的一个二元关系, 简称关系。此时, 它表示的是 M 中元素

之间的某种关联情况。

例如, 设集合 $M = \{m_1, m_2, m_3, m_4\}$, $N = \{n_1, n_2, n_3\}$, 则

$$R_1 = \{\langle m_1, n_1 \rangle, \langle m_1, n_2 \rangle, \langle m_2, n_2 \rangle, \langle m_3, n_2 \rangle\}$$

$$R_2 = \{\langle m_1, n_1 \rangle, \langle m_2, n_2 \rangle, \langle m_3, n_1 \rangle, \langle m_2, n_1 \rangle, \langle m_4, n_2 \rangle\}$$

是集合 M 到集合 N 的两个二元关系。

$$R_3 = \{\langle m_1, m_2 \rangle, \langle m_1, m_3 \rangle, \langle m_2, m_3 \rangle, \langle m_3, m_4 \rangle\}$$

则是集合 M 上的一个二元关系。

$M \times N$ 可以有若干子集, 也就是说集合 M 到集合 N 可以定义若干不同的关系, 其数目由集合 M 和 N 的元素个数决定。在实际应用中我们所取的关系是其中的很少一部分, 它们对所讨论的问题有用, 也是我们所关心的关系, 而对那些与我们所讨论的问题无关的关系, 此时我们不会去关心它。

为了描述关系中元素之间的关联情况, 对于关系 R 中的任意一个有序对 $\langle a, b \rangle$ (可记为 $\langle a, b \rangle \in R$ 或记为 aRb), 定义如下。

a 是 b 的关于 R 的前件(直接前驱);

b 是 a 的关于 R 的后件(直接后继)。

例如, 在上面的关系 R_1 中, 因为包含了有序对 $\langle m_1, n_2 \rangle$, 因此, m_1 是 n_2 的关于 R_1 的前件, 同样, n_2 是 m_1 的关于 R_1 的后件。

在上述关系 R_3 中, 有序对 $\langle m_1, m_2 \rangle$ 表明 m_1 是 m_2 的关于 R_3 的前件, 同样, m_2 是 m_1 的关于 R_3 的后件。

(2) 数据抽象的二元关系表示

二元关系的数学定义是在笛卡儿积的基础上给出的, 从实际意义来说, 二元关系表示了集合中两个元素之间的某种相关性。这两个元素可以来自两个不同集合, 也可以同属一个集合, 实际上这也就是实体之间关系的抽象。以下通过两个简单例子说明二元关系所表示的实际意义。

例 1.1 有编号为 A, B, C, D, E 的 5 个旅游团, 旅游公司派出甲、乙、丙、丁、戊 5 个带团导游。为了激励导游之间的竞争, 旅游公司向旅游团公布每个导游的特点和业绩, 旅游团可以根据自己的需要选择本团导游, 如果出现有导游未被任何旅游团选中, 旅游公司可以另派导游, 最后每个旅游团落实一个带团导游。选配结果为: 旅游团 A 的随团导游为丙, 旅游团 B 的随团导游为甲, 旅游团 C 的随团导游为丁, 旅游团 D 的随团导游为戊, 旅游团 E 的随团导游为乙, 现要求简单、明了地表示出各旅游团与其随团导游的对应关系, 我们可以用一个二元关系来达到此要求。

$$R_4 = \{\langle A, 丙 \rangle, \langle B, 甲 \rangle, \langle C, 丁 \rangle, \langle D, 戊 \rangle, \langle E, 乙 \rangle\}$$

这里 R_4 是一个二元关系, 其中每一个有序对表示每个旅游团和他所选的导游, 有序对的第一分量是旅游团, 第二分量是该旅游团所选的导游。这个二元关系清晰地表示了“旅游团”集合 $\{A, B, C, D, E\}$ 中的元素“旅游团”和“导游”集合 $\{\text{甲、乙、丙、丁、戊}\}$ 中的元素“导游”之间的关联。因此, R_4 也就是对这个问题中实体之间联系的数学抽象。

例 1.2 有编号为 1、2、3、4 的 4 个选手进行乒乓球循环赛, 比赛的结果为 1 号胜 2 号, 3 号胜 2 号, 3 号胜 4 号, 3 号胜 1 号, 1 号胜 4 号, 4 号胜 2 号。我们可以用一个二元关系很清楚地表示上述比赛结果为

$$R_5 = \{ \langle 1, 2 \rangle, \langle 3, 2 \rangle, \langle 3, 4 \rangle, \langle 3, 1 \rangle, \langle 1, 4 \rangle, \langle 4, 2 \rangle \}$$

这里 R_5 也是一个二元关系, 其中每个有序对表示一场比赛的结果, 每个有序对的第一分量是该场比赛的胜者, 第二分量是该场比赛的负者。这个二元关系表示了同一个集合 $\{1, 2, 3, 4\}$ 中的元素“选手”之间的比赛胜负关系。这里 R_5 是本问题中实体之间联系的数学抽象。

从上述例子中可以看出, 利用二元关系可以简单、明了、确切地表示出集合中两个元素之间的某种相关性。因而可以利用它来描述问题中实体之间的联系, 也就是说二元关系是数学抽象的重要结果。

1.1.3 二元关系的基本性质和几种重要关系

了解二元关系的基本性质, 对分析数据元素之间的关系是有意义的; 以下给出二元关系的几个主要的性质。

1. 二元关系的基本性质

设 R 是集合 M 上的一个关系, 则

- (1) 若对于每一个 $a \in M$, 都有 $\langle a, a \rangle \in R$, 那么称 R 是自反关系(自反性)。
- (2) 若对于任何 $a \in M$, 都有 $\langle a, a \rangle \notin R$, 那么称 R 是反自反关系(反自反性)。
- (3) 如果有 $\langle a, b \rangle \in R$, 则必有 $\langle b, a \rangle \in R$, 那么称 R 是对称的(对称性)。
- (4) 如果 $\langle a, b \rangle \in R$ 且 $a \neq b$ 时, 定有 $\langle b, a \rangle \notin R$, 那么称 R 是反对称关系(反对称性)。
- (5) 如果 $\langle a, b \rangle \in R$ 且 $\langle b, c \rangle \in R$ 时, 必有 $\langle a, c \rangle \in R$, 那么称 R 是传递关系(传递性)。

假设有集合 $M = \{1, 2, 3, 4, 5\}$, M 上的一个关系为

$$R = \{ \langle 1, 2 \rangle, \langle 2, 3 \rangle, \langle 3, 4 \rangle, \langle 3, 5 \rangle, \langle 1, 3 \rangle, \langle 1, 4 \rangle, \langle 1, 5 \rangle, \langle 2, 4 \rangle, \langle 2, 5 \rangle \}$$

根据上述性质定义, 很容易看出该关系 R 是反自反的、反对称的、传递的。

在实际问题中, 利用二元关系的基本性质, 可以更方便、更确切地分析和描述数据集合中数据元素之间的关联规律和特点, 从而有利于对它们的处理。

2. 等价关系与等价类

等价关系是一类重要的二元关系, 在处理问题时经常会用到它。等价关系的定义是:

如果非空集合 S 上的关系 R 是自反的、对称的和传递的, 则称 R 为 S 上的一个等价关系。如果 R 是 S 上的等价关系, a, b 是 S 的任意元素, 若有 $\langle a, b \rangle \in R$, 则称 a 等价于 b , 记作 $a \sim b$ 。

等价关系在现实世界中广泛存在。例如, 假设有学生集合 $S = \{a, b, c, d, e, f\}$, S 上的“同班”关系 R 就是等价关系。因为:

- ① 任何一个人都和自已同一个班, 即具有自反性。
 - ② 若 a 和 b 是同一个班的, 那么 b 当然也就和 a 同在一个班。即具有对称性。
 - ③ 如果 a 和 b 在一个班, b 与 c 是同班的, 那么 a 和 c 一定是同班的。亦即具有传递性。
- 由此可以看出, 这个“同班关系”是一个等价关系。

假设 a, b, c 是同班的, d, e, f 在一个班, 则同班关系为

$$R = \{ \langle a, a \rangle, \langle b, b \rangle, \langle c, c \rangle, \langle a, b \rangle, \langle b, a \rangle, \langle b, c \rangle, \langle c, b \rangle, \langle a, c \rangle, \langle c, a \rangle, \langle d, d \rangle, \langle e, e \rangle, \langle f, f \rangle, \langle d, e \rangle, \langle e, d \rangle, \langle e, f \rangle, \langle f, e \rangle, \langle d, f \rangle, \langle f, d \rangle \}$$

$$\langle e, f \rangle, \langle f, e \rangle, \langle d, f \rangle, \langle f, d \rangle\}$$

由等价关系可以进一步给出等价类的概念。假设 R 是非空集合 S 上的等价关系, 则 S 上互相等价的元素构成了 S 的若干不相交的子集, 这些子集称为 S 的 R 等价类。

下面给出等价类的一般定义:

设 R 是非空集合 S 上的等价关系, 对任意的 $x \in S$, 由 $[x]_R = \{y \mid y \in S \wedge xRy\}$ 给出的集合 $[x]_R$, 称为 x 关于 R 的等价类, 简称为 x 的等价类, 简记为 $[x]$ 。

例如, 对于上述学生集合的同班关系, 根据“同班”关系得到的不相交的学生子集 $\{a, b, c\}$ 和 $\{d, e, f\}$ 分别是学生 a 和 d 的等价类。当然, 它们也是学生 b, c 和学生 e, f 的等价类。即有

$$[a] = [b] = [c] = \{a, b, c\}, \quad [d] = [e] = [f] = \{d, e, f\}$$

在实际问题中, 当需要根据某种关系来划分集合中的元素时, 划分等价类是常用的方法。后续章节将结合具体应用讨论等价类的划分方法。

3. 偏序关系和全序关系

偏序关系和全序关系也是重要的关系, 它们提供了比较集合中元素的工具。

偏序关系的定义是: 如果集合 S 上的关系 R 是自反的、反对称的和传递的, 则称 R 是集合 S 上的偏序关系。称序偶 $\langle S, R \rangle$ 为偏序集合, 也记为 $\langle S, \leq \rangle$ 。

例如, 集合 $S = \{2, 4, 6, 8\}$ 和其上的整倍数关系 $R = \{\langle 2, 2 \rangle, \langle 4, 4 \rangle, \langle 6, 6 \rangle, \langle 8, 8 \rangle, \langle 4, 2 \rangle, \langle 6, 2 \rangle, \langle 8, 2 \rangle, \langle 8, 4 \rangle\}$, 则构成偏序集合 $\langle S, R \rangle$ 。

对于偏序集 $\langle S, \leq \rangle$ 中的 $x, y \in S$, 如果有 $x \leq y$ 或 $y \leq x$ 成立, 我们说 x 和 y 是可比较的。由偏序集合的定义可知, 偏序集合中的元素不一定是可比较的, 也就是说, 它们在偏序中不一定都有确定的位置上的先后关系。若要使集合中的元素之间都可比较, 则需要构造下面定义的全序关系。

全序关系的定义是: 在偏序集合 $\langle S, \leq \rangle$ 中, 如果对于任意的 $x, y \in S$, x 和 y 都是可比较的, 则称此偏序关系 $\leq$ 为 S 上的全序关系, 称序偶 $\langle S, \leq \rangle$ 为全序集合。

例如, 集合 $S = \{1, 2, 3, 4\}$ 上的小于等于关系: $R_1 = \{\langle 1, 1 \rangle, \langle 2, 2 \rangle, \langle 3, 3 \rangle, \langle 4, 4 \rangle, \langle 1, 2 \rangle, \langle 1, 3 \rangle, \langle 1, 4 \rangle, \langle 2, 3 \rangle, \langle 2, 4 \rangle, \langle 3, 4 \rangle\}$, 即是全序关系, $\langle S, R_1 \rangle$ 为全序集合。

由全序集合的定义可以看出, 全序集合中所有元素之间都是可以比较的, 这为我们处理实际问题时从给定的一些“前提条件”(即偏序关系)出发, 确定集合中各成员之间的先后位置关系提供了办法。因为根据偏序集合与全序集合的关系, 我们能够以非空有限偏序集合 $\langle S, \leq \rangle$ 为基础, 在集合 S 的元素上构造一个包含了原来的偏序关系的全序, 从而使得集合中的全体成员在满足给定的偏序关系的前提下都可比较, 这就是在第 5 章中将专门讨论的拓扑排序(也叫拓扑分类)问题。

1.2 什么是数据结构

1.2.1 数据结构的引出

前面已经提到, 数据结构学科的形成与发展是和程序设计技术的发展、计算机应用的日