

21世纪管理学系列教材

数据结构

DATA STRUCTURES

戚桂杰 姚云鸿 主 编

山东人民出版社



数据结构

DATA STRUCTURES

主 编：戚桂杰 姚云鸿

副主编：李明明 焦丽媛

山东人民出版社

图书在版编目(CIP)数据

数据结构/戚桂杰,姚云鸿主编. — 济南:山东人民出版社,2007.6

(21世纪管理学系列教材)

ISBN 978-7-209-04254-3

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

中国版本图书馆CIP数据核字(2007)第063100号

责任编辑:于宏明

数据结构

戚桂杰 姚云鸿 主编

山东出版集团

山东人民出版社出版发行

社址:济南市经九路胜利大街39号 邮编:250001

网 址:<http://www.sd-book.com.cn>

发行部:(0531)82098027 82098028

新华书店经销

青岛星球印刷有限公司印装

规格 16开(180mm×235mm)

印 张 19.5

字 数 360千字 插 页 2

版 次 2007年6月第1版

印 次 2007年6月第1次

印 数 1—3000

ISBN 978-7-209-04254-3

定 价 29.00元

如有质量问题,请与印刷厂调换。(0532)88194567

《21世纪管理学系列教材》编委会

主 编：徐向艺

副主编：赵炳新(常务) 杨蕙馨 刘 岗 吉小青

编 委：（按姓氏笔画排列）

丁荣贵 王兴元 王益明 吉小青

刘 岗 刘洪渭 陈志军 杨蕙馨

张体勤 赵炳新 赵景华 胡正明

徐向艺 戚桂杰 潘爱玲



总 序

管理的实践可以追溯到遥远的古代人类文明,但直到 20 世纪初泰勒的开创性贡献——《科学管理原理》一书的问世,才标志着人类告别经验管理时代,进入了科学管理时代。100 多年来,由于世界经济环境的发展变化,科学技术尤其是信息技术的突破性进展和大范围的应用,市场竞争的日趋激烈和国际化思潮风起云涌,都推动着现代管理思想、管理理论、管理方法和管理手段的日新月异。在 21 世纪,全球的管理者,无疑面对着更大范围的管理创新和变革。如何迎接这一世纪性的挑战,则成为世界各国政府、企业界和理论界共同关注的课题。

回顾上一个世纪,管理理论经历了古典管理理论阶段、行为科学阶段和现代管理理论阶段的演变,形成了庞大的知识体系。进入 20 世纪 90 年代后,随着知识经济的崛起和全球经济一体化进程的加快,市场环境的变化更加迅速,导致管理学家面临许多前所未有的新情况和新问题。于是当代管理新思潮也竞相涌现。对于历经改革开放风雨洗礼,正在进入世界大市场、融入世界经济大循环的中国来说,更是面临进入市场经济后复杂的现实管理问题。

在 21 世纪未来的岁月里,中国的企业、各级政府、事业团体等各类组织在管理方面所面对的主要挑战在于:

1. 变革与创新。今天的变化快过以往任何时候,原来的流程管理和职能管理方法难以适应当今的变化。世界在发展,新的问题层出不穷,需要运用已有的理论和方法去解决,更需要我们去大胆探索和创新。创新是管理理论发展的路径,是任何组织立于不败之地的法宝。

2. 国际化。互联网的兴起无疑大大改变了当今的经营环境。这是过去一个多世纪没有、也不可能预见到的。国际化的进程要求我们认真研究全球化战略,而不仅仅是国际化或跨国界运作战略。全球一盘棋的思想、跨文化的研究和应用必须扎根于当今领先企业领袖的头脑中。我们知道东方文化与西方文化是有很大的差异的,这与东西方人的思维方式的差异有关:感性与理性;严谨与灵性;

实验验证与感悟修证。这些差异反映在管理领域,产生了很多耐人寻味的现象。

3. 知识经济。知识的管理大大超过了过去一个世纪管理学者的想像。这里不是为知识而知识的知识管理,而应该是如何真正把知识应用于组织,实现知识转移、知识共享,推动组织的发展。

4. 网络技术。网络技术的应用是当今组织管理者尤其应该认真思考的问题。如何利用网络技术提高效率,节省成本,改善沟通,强化协调等等,有许多的新内容值得我们去研究。

面对这些挑战,谁能够最快地吸收各种管理新知识,谁就会获得竞争的主动权;谁拥有更多的知识,谁能够通过管理创新把更多的知识组合成独特的能力,谁就能赢得未来。

作为管理教育的从业者,我们理所应当承担起为我国管理科学的发展添砖加瓦的责任。一方面紧跟国际潮流,逐渐实现管理教育的规范化、国际化;另一方面则必须面对改革开放的丰富实践,推动管理教育创新。这就是我们编写这套系列教材的出发点。

《21 世纪管理学系列教材》主要是针对管理类各专业本科生教学编写的,亦可作为工商管理硕士(MBA)和经济、管理类各专业研究生参考书。教材的编写突出以下特点:

1. 扩大信息量,同时给教师授课留有一定的空间。作为管理类本科生、研究生专业课或基础课教材,要让学生掌握较大的信息量,管理学理论的重要成果不宜疏漏;同时,对于比较成熟理论的阐述,不要过细,让老师有充分的发挥余地。教材使用能达到教师好教、学生好学的目的。

2. 反映管理学的深邃与智慧。不过分追求“通俗易懂”,突出体现内容的专业性、学术性。突出管理学的哲学价值。管理过程充满着哲学道理,如决策中的“概率思考”,领导中的“权变观念”,管理控制中的“反馈原理”等,在阐述管理的基本问题时,体现管理学科的哲学性和科学性。

3. 反映管理学的最新理论进展。随着信息技术和知识经济的发展,传统的管理思想和管理方式受到了挑战。教材编写追踪管理理论前沿,反映管理学科的发展,体现理论的时代性和前沿性。

4. 处理好借鉴与创新的关系。教材编写既借鉴已有的理论成果,又注重理论的创新性,注重把作者自己的最新的理论成果写入教材。

本系列教材是山东大学管理学院教师们多年教学实践和科学研究成果的结晶。教材编写委员会统一指导各书编写,选择确定各书主编,精心组织参编教师,审查把关编写质量。力求使每种教材不但适合当前管理专业本科教学,同时也符合 21 世纪管理学科教育的发展大势。





编写一套贴近管理学科理论前沿、广泛介绍管理各学科成熟内容的教材,必然要参考国内外大量已有的研究成果,吸收近百年来管理理论的精华。我们谨向一切致力于管理学科理论繁荣的前辈与同仁致以崇高的敬意!

徐向艺

2005年1月24日



前言

信息时代,计算机应用已深入到社会、经济的各个领域,各行各业都需要对大量非数值型数据进行存储、加工和管理。这些数据并不是孤立的、杂乱无章的,而是有内在联系的,为了能有效地对数据进行处理,我们必须考虑数据之间的相互关系。各种数据本身及其它们之间的相互关系被称为“数据的逻辑结构”,或简称“数据结构”,它是人们对数据处理的依据,也是对数据进行存储和管理的基础,可以说,分析、确定数据的逻辑结构是进行数据组织的出发点;其次,由于要应用计算机处理数据,就需要将逻辑上组织好的数据存入计算机,即进行数据的物理组织,数据结构在计算机中的表示称为“数据的物理结构”或“数据的存储结构”,它包括对数据的存储和数据之间关系的存储;第三,要对存储在计算机中的数据结构进行运算或操作,就要考虑对数据进行操作的方法或对数据进行操作的步骤,称之为算法,这些算法的集合是定义在数据的逻辑结构上的,各个算法的实现是建立在存储结构上的。以上三方面就是数据结构所要研究的基本内容。

目前,《数据结构》不仅是计算机专业的核心课程之一,也是信息管理与信息系统等管理科学与工程类专业的主干课程之一。该课程的目的是使学生能通过对最常用的数据结构的逻辑特性、存储表示以及操作算法的学习与上机实习进行复杂程序设计的训练,培养学生组织数据、处理数据并设计“好的”程序的理念,为今后在相应专业领域的工作学习打下坚实的基础。

全书共包括八章内容。第一章介绍数据结构的基本概念,包括:数据结构、数据类型、抽象数据类型;数据的存储结构——顺序存储结构、链接存储结构;用类C语言描述算法及算法的评价标准等。第二章介绍线性表的类型定义,包括线性表的抽象数据类型定义;顺序存储及链式存储的线性表、不同存储表示线性表所需要的具体数据类型以及相应操作的实现算法等。第三章讨论栈和队列的抽象数据类型定义、表示及实现与应用。第四章主要以二叉树为例,讲述树结构及其应用,主要内容包括:二叉树的定义、性质,二叉树的存储结构及二叉树的遍历等各种操作的算法实现;线索二叉树;树、森林与二叉树的转化;赫夫曼树及其



应用。第五章介绍图的相关术语及图的类型等概念;图的存储表示方法;图的标准搜索方法——深度优先搜索、广度优先搜索;图的应用——拓扑排序、最小生成树、最短路径等。第六章介绍各种查找算法,包括:顺序查找、折半查找、分块查找等在顺序表上完成的查找算法;在二叉查找树、B树等树表结构上完成的查找方法;散列表查找等。第七章考虑数据处理中经常用到的插入排序、交换排序、堆排序、归并排序等排序方法。第八章讨论文件的相关概念、表示方法及各种运算的实现方法。

本书是根据高等院校信息管理 with 信息系统等专业对数据结构知识的要求编写的,它介绍的是数据结构学科成熟而实用的知识,摒弃那些深奥难懂而又较少应用的内容;在写法上力求条理清晰、层次分明、简明扼要、循序渐进,便于读者学习。本书主要采用类C语言作为数据结构算法的描述语言,它使本书对各种抽象数据类型的定义和实现简明清晰,既不拘泥C语言的细节,又容易转换成能上机执行的C或C++程序。同时,为便于读者进行从类C语言到C语言的转换,我们在附录中给出了部分章节的C语言程序,另外,第五、六章的算法基本上是按照C语言规范书写的。

本书的撰稿人(以章节先后为序):第一章至第四章由戚桂杰、李明明、焦丽媛撰写,第五章至第六章由姚云鸿、姜明撰写,第七章由戚桂杰、焦丽媛、李明明撰写,第八章由姚云鸿撰写,附录(部分章节的C语言程序)由李明明、王翠翠、聂广礼撰写并进行了上机调试。张伟收集了大量关于数据结构目前的前沿研究性资料,为本书写作团队提供了较大的帮助。全书由戚桂杰提出整体框架、组织讨论编写计划并负责统稿工作。本书的出版得到了山东大学管理学院领导及同事们的大力支持,山东人民出版社的于宏明老师为本书的顺利出版提供了莫大的帮助,在此表示衷心的感谢。此外,在本书的写作过程中,参阅和引用了国内外有关专家学者的论著,因篇幅所限,可能未全部一一列出,在此一并致谢。由于时间仓促及水平所限,书中难免有许多不当之处,敬请读者批评指正。

戚桂杰 姚云鸿

2007年3月于山东大学



目 录

第一章 绪论	(1)
第一节 数据结构实例	(2)
第二节 基本概念和术语	(4)
第三节 算法和算法的评价	(13)
第二章 线性表	(17)
第一节 线性表的类型定义	(18)
第二节 线性表的顺序表示和实现	(20)
第三节 线性表的链式表示和实现	(26)
第四节 线性表的其他链式表示	(33)
第五节 线性表的应用举例	(38)
第三章 栈和队列	(45)
第一节 栈	(45)
第二节 栈的应用举例	(53)
第三节 队列	(67)
第四节 队列的应用举例	(77)
第四章 树和二叉树	(80)
第一节 树	(80)
第二节 二叉树	(85)
第三节 线索二叉树	(103)
第四节 树和森林	(106)
第五节 赫夫曼树及其应用	(111)
第五章 图	(119)
第一节 图的数学结构	(119)
第二节 图的计算机表示	(125)
第三节 图的遍历	(138)
第四节 图的应用之一 —— 最小生成树	(144)



第五节	图的应用之二 —— 拓扑排序与关键路径	(150)
第六节	图的应用之三 —— 最短路径	(155)
第六章	查找/检索	(165)
第一节	基本概念	(165)
第二节	静态查找表	(168)
第三节	二叉查找树	(176)
第四节	B-树	(191)
第五节	散列表查找	(200)
第七章	排 序	(210)
第一节	概述	(210)
第二节	插入排序	(212)
第三节	交换排序	(216)
第四节	选择排序	(219)
第五节	归并排序	(224)
第六节	各种内部排序方法的比较讨论	(226)
第七节	外部排序	(228)
第八章	文 件	(232)
第一节	文件概述	(232)
第二节	顺序文件	(236)
第三节	直接文件(散列文件)	(239)
第四节	索引文件	(243)
第五节	索引顺序文件	(245)
第六节	倒排文件	(250)
附录:	部分章节算法的 C 语言代码	(253)
主要参考文献		(304)



第一章 绪 论

在计算机发展初期,人们使用计算机的目的主要是处理数值计算问题。随着计算机软、硬件的发展和计算机应用领域的扩大,非数值计算问题显得越来越重要。解决这类问题的关键不再是数学分析和计算方法,如何设计出合适的数据结构常常是能否有效解决问题的关键所在。因为应用中需要处理的一批数据之间常常有一定的逻辑关系,我们应当根据数据的逻辑关系有效地将数据组织存储在计算机内,才能实现对数据的操作,否则毫无规律的将大批数据“塞”进计算机,不可能有效地去操作这些数据,如同将图书馆的图书杂乱无章地堆放在图书馆内,不可能迅速查阅或取出所需要的书籍,所以我们应研究数据之间的逻辑关系,合理组织它们,存储到计算机内,并依此实现操作数据的算法,这些正是数据结构研究的内容。

数据结构作为信息管理与信息系统专业的基础课,是十分重要的核心课程。所有的计算机系统软件和应用软件都要用到各种类型的数据结构,因此,要想更好地运用计算机来解决实际问题,充分有效地发挥计算机的性能,仅掌握几种程序设计语言是远远不够的,还必须学习和掌握好数据结构的有关知识。打好“数据结构”这门课程的扎实基础,对于学习本专业的其他课程如操作系统、数据库原理与应用、软件工程、信息系统分析与设计等都是十分有益的。

本章将介绍数据结构的基本概念,主要包括:

- (1)数据结构基本概念:数据、数据元素、数据对象、数据结构、数据类型、抽象数据类型;
- (2)数据的存储结构——顺序存储结构、链接存储结构;
- (3)算法和算法的评价:时间复杂性、空间复杂性。



第一节 数据结构实例

无论是进行科学计算、数据处理、过程控制、对文件的存储和检索以及数据库技术应用等,都是对数据进行加工处理的过程。因此,要设计出一个结构好、效率高的程序,必须研究数据的特性、数据间的相互关系及其对应的存储表示,并利用这些特性和关系设计出相应的算法和程序。下面我们从应用的角度来看数据结构实例。

[例 1-1] 飞机订票系统。应用本系统进行飞机订票的过程中,客户首先通过查询航班信息表,并根据自己的情况按照起始城市和起降时间查找具体航班,然后,根据票价和座位情况,确定航班号,进而进行订票操作。订票操作中,客户输入姓名、证件号等信息后,选择航班、订票数,即可完成订票。在此类系统中,计算机处理的对象之间通常存在着一种最简单的线性关系,如图 1-1 所示,即各个数据元素按照一定的顺序线性排列,我们把这类结构称为线性数据结构。

该类线性数据结构还可应用于其他各种订票系统、学籍管理、选课系统、网上购物、仓库账目管理等。

航班号	起降时间	起始城市	航班票价	票价折扣	是否满仓	
1001	8:30 ~ 10:00	济南—上海	760	五折	否	
1002	9:30 ~ 11:40	济南—北京	630	六折	是	
1003	9:00 ~ 11:30	广州—济南	1590	五折	否	
1004	8:00 ~ 10:00	济南—成都	1360	八折	否	
⋮	⋮	⋮	⋮	⋮	⋮	⋮

姓名	证件号	航班号	订票数量	
丁琳	130903 * * * * * 0123	1003	1	
张伟	370102 * * * * * 6586	1004	3	
李明	420202 * * * * * 2581	1001	1	
王强	230105 * * * * * 9316	1002	4	
⋮	⋮	⋮	⋮	⋮

图 1-1 订票系统存储文件示例

[例 1-2] 物料清单(BOM)。物料清单 BOM(Bill of Material)是一种描述配套件结构的零件表,其中包括所有子件、零件、原材料的清单,以及制造一个配件



所需要的所有物料的数量。BOM 是制造业信息系统的一个核心部件。如果我们将产品的配件按照线性数据结构罗列表示,如图 1-2(a)所示,将不利于产品的设计和改进;当把产品的配件按照一定的层次进行表示时,如图 1-2(b)所示,形成的结构图像一棵倒长的树,“树根”是该产品,而所有的“叶子”就是产品对应的零配件。这时的产品配件结构鲜明,易于分析和应用。我们把此类存储产品配件资料的数据结构称为“树”结构。

该产品结构图不仅可以应用于产品的设计和改进,还可应用到生产管理中。当接收到客户订单时,查询每个结点即配件是否有货,即进行“遍历”操作。如果无货,对于本厂生产的配件,制订出生产作业计划;对于外购的配件,向供货商发出订单。如果把图 1-2(b)的“树”结构转换成图 1-2(c)的结构,即每个结点只有不多于两个分叉,转换后的结构可称为“二叉树”,该“树”的结点的增加、删除、修改、遍历等操作的相应算法将比非“二叉树”的算法更加简单和成熟。

组件 a 1 件, 部件 b 3 件, 套件 c 2 件

零件 d 4 件(两件用于 a, 两件用于 c)

零件 e 3 件(用于 c)

零件 f 1 件(用于 a)

零件 g 2 件(用于 b)

零件 h 3 件(两件用于 a, 一件用于 b)

零件 i 1 件(用于 c)

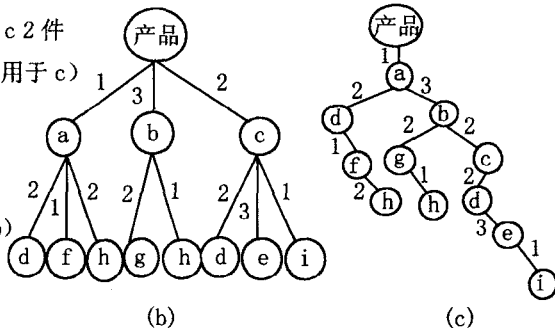


图 1-2 物料清单结构图

[例 1-3] 邮递员送信问题。如图 1-3(a)所示,现在邮递员从邮局出发,向以下几个地点 2、3、4 送信。要求选择一条路线,从邮局出发,经过每个地点,最终回到邮局,并且路线长度最短。

该问题描述的是经典的“中国邮路”问题,通常这类邮件发送问题的数学模型是一种称为“图”的数据结构。在此例中,如图 1-3(b)所示,可以通过一个顶点表示一个地点,通过两点之间的连线表示两地之间的通路。

此外,在诸如城市物流配送、建筑网络布线、城市道路规划等问题中,均会用到“图”数据结构。

由以上三个例子可见,我们不是用数学方程,而是通过诸如线性表、树、图之类的数据结构来描述这类非数值计算问题的数学模型。可以说数据结构是一门研究非数值计算的程序设计问题中计算机的操作对象以及它们之间的关系和操作等的学科。

数据结构作为一门独立的课程在国外开始于 1968 年。但在此之前,其有关

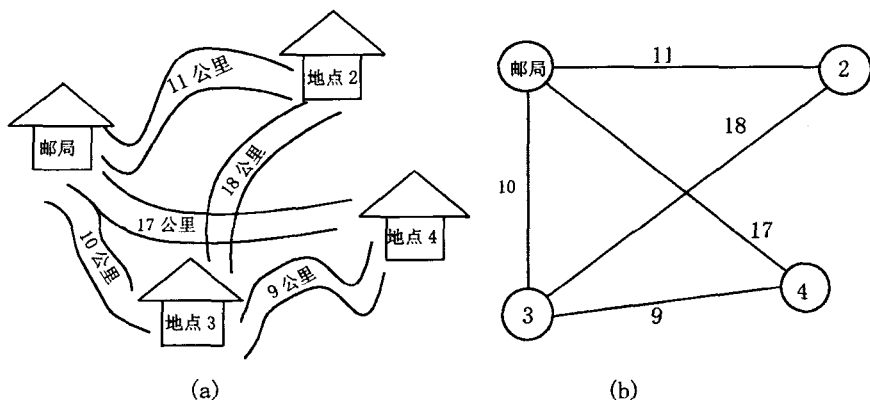


图 1-3 邮递员送信问题

内容已散见于编译原理及操作系统之中。20 世纪 60 年代中期,美国的一些大学开始设立有关课程,但当时的课程名称并不叫数据结构。1968 年美国唐·欧·克努特教授开创了数据结构的最初体系,他所著的《计算机程序设计技巧》第一卷《基本算法》,是第一本较系统地阐述数据的逻辑结构、存储结构及其操作的著作。从 20 世纪 60 年代末到 70 年代初,出现了大型程序,软件也相对独立,结构化程序设计成为程序设计方法学的主要内容,人们开始越来越重视数据结构。从 70 年代中期到 80 年代,各种版本的数据结构著作相继出现。经过了几十年的发展演变,数据结构成为了介于数学、计算机硬件和计算机软件之间的一门课程。

目前,数据结构的发展仍在继续。一方面,面向各专门领域中特殊问题的数据结构正在被研究和发展,如多维图形数据结构等;另一方面,从抽象数据类型和面向对象的观点来讨论数据结构已成为一种新的趋势,越来越被人们所重视。

第二节 基本概念和术语

一、数据

数据(Data)是人们利用各种符号对客观事物及其活动进行的抽象描述。从计算机程序设计的角度来讲,数据则是计算机程序加工和处理的“原料”和对象。例如,一个简单的数值计算程序所使用的数据是一些整数和实数;一个编译程序



所加工、处理的数据则是用某种高级语言编写的源程序。就我们本教材所讨论的内容而言,我们可做如下定义:数据是描述客观事物用到的数、字符以及所有能输入到计算机中并能被计算机处理的符号的集合,它是计算机程序使用、加工的原料和输出的结果。一般把数据分为数值性数据和非数值数据两类。数值性数据常用于科学计算和商业事务处理等,包括整数、实数、复数、双精度数;非数值数据一般是指字符、声音、图像、文字等数据。

数据元素(Data Element)是数据整体中相对独立的基本单位,通常作为一个整体在计算机程序中进行处理。针对处理具体问题的不同,数据元素可以是简单的数、字符到复杂的对象(例如 C 语言中的结构体类型),只要它们能被计算机接受并处理,包括图形、声音等。

一个数据元素可以由若干个数据项组成,数据项是数据的最小单位,例如银行储蓄业务管理系统中处理的数据元素可包括如下数据项:

账号,姓名,存入,支出,余额,利息;

学生档案管理系统中的数据项常常是:

学号,姓名,性别,年龄,籍贯,学习成绩。

数据对象(Data Object)是性质相同的数据元素的集合,是数据的一个子集。例如,整数数据对象是集合 $N = \{0, \pm 1, \pm 2, \dots\}$, 字母字符数据对象是集合 $C = \{'A', 'B', \dots, 'Z'\}$ 。

二、数据结构

现实世界中,客观事物都是相互联系的,所以数据之间也必然存在着联系。数据结构(Data Structure)就是指数据元素及其相互之间的关系。

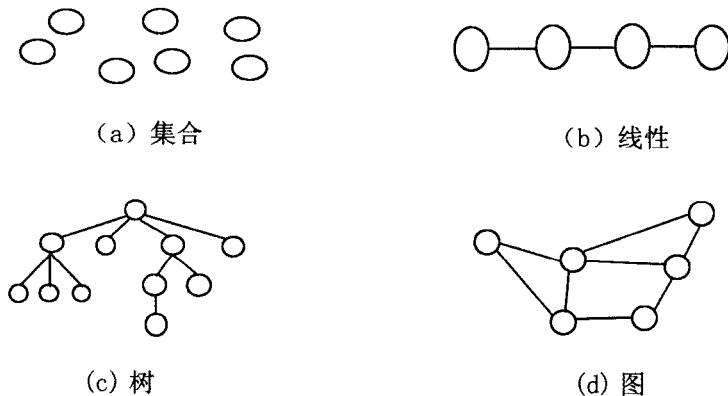


图 1-4 数据结构的分类



数据结构

数据元素之间的相互联系称为数据的逻辑结构。根据数据元素之间关系的不同特性,数据结构可以分为集合、线性结构、树型结构、图结构 4 类基本结构,见图 1-4。(1)集合:数据元素间的关系同属一个集合,除此之外,别无其它关系。(2)线性结构:结构中的元素间存在一对一的关系。(3)树型结构:结构中的元素间存在一对多的关系。(4)图(网状)结构:结构中的元素间存在多对多的关系。在上节中,我们已经看到了分别为线性结构、树结构和图结构的三个例子。

为了更确切地描述数据的逻辑结构,通常采用二元组表示数据结构:

$$\text{Data_Structure} = (D, S)$$

其中:D 是数据元素的有限集合,S 是 D 上关系的有限集合。

下面结合上节中的具体例子进行说明。

[例 1-4] 在例 1-1 的飞机订票系统中,定义一种数据结构为:

$\text{Linearity} = (\text{Name}, \text{Relation})$, 其中

$\text{Name} = \{\text{丁琳}, \text{张伟}, \text{李明}, \text{王强}\}$

$\text{Relation} = \{ \langle \text{李明}, \text{王强} \rangle, \langle \text{王强}, \text{丁琳} \rangle, \langle \text{丁琳}, \text{张伟} \rangle \}$

这是按照航班号的大小顺序排列的线性关系。对应的图形如图 1-5 所示。

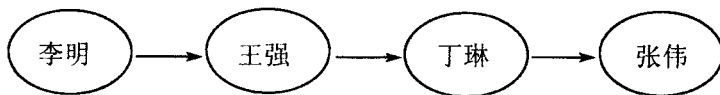


图 1-5 数据的线性结构

[例 1-5] 在例 1-2 的物料清单结构中,定义一种数据结构为:

$\text{Tree} = (\text{Accessory}, \text{Relation})$, 其中

$\text{Accessory} = \{a, d, f, h\}$

$\text{Relation} = \{ \langle a, d \rangle, \langle a, f \rangle, \langle a, h \rangle \}$

对应的图形如图 1-6 所示。

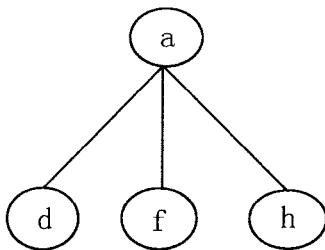


图 1-6 数据的树结构