

21世纪高等学校计算机科学与技术规划教材

数据结构

SHUJU JIEGOU

蹇强 罗宇 编著



北京邮电大学出版社
www.buptpress.com

21 世纪高等学校计算机科学与技术规划教材

数据结构

蹇强 罗宇 编著

北京邮电大学出版社

内容简介

本书共九章。第一章从逻辑结构、存储结构和数据运算三方面介绍了数据结构的基本概念;第二章至第四章介绍了线性结构,详细介绍了顺序表、链表、栈、队列、串、数组和广义表等各种常用数据结构;第五章介绍了树型结构;第六章简单介绍了图结构;第七章和第八章介绍了各种查找和排序的基本运算;最后在第九章简单介绍了文件结构。

本书注重应用、选材精练、图文并茂,对基本理论的叙述深入浅出、通俗易懂;精选的应用实例涉及领域相当广泛,给读者提供了思路和方法,有助于提高读者分析和解决实际问题的能力。本书结构合理,内容丰富,算法描述清晰,主要算法均给出C语言描述,便于自学,而且每章后都进行小结并配有适量习题,便于读者掌握各章的重点和难点并进行必要的训练。

本书可作为高等院校计算机专业或非计算机专业的教科书,亦可作为从事计算机应用的工程技术人员的自学参考书。

图书在版编目(CIP)数据

数据结构/蹇强,罗宇编著. —北京:北京邮电大学出版社,2004

ISBN 7-5635-0836-8

I. 数... II. ①蹇...②罗... III. 数据结构 IV. TP311.12

中国版本图书馆 CIP 数据核字(2004)第 038533 号

书 名: 数据结构

编 著: 蹇 强 罗 宇

策 划: 三文工作室

E-mail: sanwen99@mail.edu.cn

责任编辑: 陈露晓 付晓霞

出版发行: 北京邮电大学出版社

社 址: 北京市海淀区西土城路 10 号(100876)

电话传真: 010-62282185(发行部) 010-62283578(传真)

经 销: 各地新华书店

印 刷: 国防科技大学印刷厂印刷

开 本: 787mm×960mm 1/16

印 张: 19.75

字 数: 327 千字

版 次: 2004 年 5 月第 1 版 2005 年 1 月第 3 次印刷

ISBN 7-5635-0836-8/TP·107

定 价: 29.50 元

如有质量问题请与发行部联系

版权所有 侵权必究

21 世纪高等学校计算机科学与技术规划教材

编委会

- | | | |
|-----|-----|---------------------|
| 主 任 | 陈火旺 | 中国工程院院士,国防科技大学教授 |
| 委 员 | 周立柱 | 清华大学计算机系主任 |
| | 杨放春 | 北京邮电大学计算机科学与技术学院院长 |
| | 杨学军 | 国防科技大学计算机学院院长 |
| | 徐晓飞 | 哈尔滨工业大学计算机科学与技术学院院长 |
| | 李仁发 | 湖南大学计算机与通信学院院长 |
| | 卢正鼎 | 华中科技大学计算机学院院长 |
| | 李志蜀 | 四川大学计算机学院院长 |
| | 戴居丰 | 天津大学信息学院、软件学院院长 |
| | 蒋昌俊 | 同济大学计算机科学与工程系主任 |
| | 何炎祥 | 武汉大学计算机学院院长 |
| | 周兴社 | 西北工业大学计算机系主任 |
| | 陈志刚 | 中南大学信息学院副院长 |
| | 姜云飞 | 中山大学软件学院院长 |
| | 周昌乐 | 厦门大学软件学院院长 |
| | 齐 勇 | 西安交通大学计算机科学与技术系主任 |
| | 赵书城 | 兰州大学计算机学院院长 |
| | 孟祥旭 | 山东大学计算机学院院长 |

序

自 20 世纪 80 年代以来,高等学校计算机教育发展迅速,计算机教育的内容不断扩展、程度不断加深。特别是近十年来,计算机向高度集成化、网络化和多媒体化发展的速度一日千里;社会信息化不断向纵深发展,各行各业的信息化进程不断加速;计算机应用技术与其他专业的教学、科研工作结合更加紧密;各学科与以计算机技术为核心的信息技术的融合,促进了计算机学科的发展,各专业对学生的计算机应用能力也有更高和更加具体的要求。

基于近年来计算机学科的长足发展,以及国家教育部关于计算机基础教学改革的指导思路,我们确立了这套“21 世纪高等学校计算机科学与技术规划教材”的编写计划与编写思想。教材是教学过程中的“一剧之本”,是高校计算机教学的首要问题。该套系列教材编写计划的制定凝聚了编委会和作者的心血,是大家多年来计算机学科教学和研究成果的体现,并得到了陈火旺院士的亲自指导与充分肯定。

这套系列教材由北京邮电大学出版社三文工作室精心策划和组织。编写过程中,充分考虑了计算机学科的发展和《计算机学科教学计划》中内容和模块的调整,使得整套教材具有科学性和实用性。整套系列教材体系结构按课程设置进行划分。每册教材均涵盖了相应课程教学大纲所要求的内容,既具备学科设置的合理性,又符合计算机学科发展的需要。从结构上遵循教学认知规律,基本上能够满足不同层次院校、不同教学计划的要求。

各册教材的作者均为多年来从事教学、研究的专家和学者,他们有丰富的教学实践经验,所编写的教材体系结构严谨、内容充实、层次清晰、概念准确、论理充分、理论联系实际、深入浅出、通俗易懂。

教材建设是一项长期、艰巨的系统工程,尤其是计算机科学技术发展迅速、内容更新快,为使教材更新能跟上科学技术的发展,我们将密切关注计算机科学技术的发展新动向,以使我们的教材编写在内容上不断推陈出新、体系上不断完善,以适应高校计算机教学的需要。

21 世纪高等学校计算机科学与技术规划教材编委会

2003 年 12 月

前 言

“数据结构”这门课一直是高等院校计算机专业中的一门主要技术基础课程,当前,在各高等院校计算机专业的教学计划中,均把该课程列为核心课程。计算机在各个领域的应用过程中,都会涉及到数据的组织与程序的编排等问题,都会用到各种各样的数据结构,特别是针对各种特殊数据的表示,就更需要学会分析和研究计算机数据处理对象的特性,选择最合适的数据组织结构及其存储表示方法,以及编制相应实现算法的方法。这是从事计算机科学及其应用的科技工作者必须掌握的重要知识。

本书介绍了各种最常用的数据结构,阐述了各种数据结构所内涵的逻辑关系,讨论了它们在计算机中的存储表示,以及在这些数据结构上的运算(操作)和实际的执行算法,并对算法的效率进行了简要的分析和讨论。

本书共九章。第一章从逻辑结构、存储结构和数据运算三方面介绍了数据结构的基本概念和基本知识,并给出了算法评价指标和评价方法;第二章介绍了线性表的逻辑结构和运算,重点讨论了顺序表和链表的结构及其基本运算实现,并通过实例揭示了有关知识的应用方法;第三章介绍了栈、队列这两种特殊的线性结构的逻辑结构和运算,介绍了顺序栈、链栈、顺序(循环)队列和链队列及其基本运算的实现,并通过实例着重介绍了栈和队列的应用;第四章介绍了串、多维数组以及广义表的概念、运算和存储结构,介绍了特殊矩阵和稀疏矩阵的压缩存储等内容;第五章介绍了树的逻辑结构、存储结构和运算,介绍了二叉树的概念、性质、存储结构,重点讨论了二叉树的遍历运算及其应用,描述了线索二叉树的概念、作用和有关运算,并给出了相应的例题及分析;最后介绍了哈夫曼树的有关知识及其应用;第六章介绍了图的有关概念及两种常用的存储结构,重点讨论了深度和广度遍历算法及其应用,最后介绍了包括最小生成树、拓扑排序、关键路径、最短路径等图结构的几个典型应用;第七章讨论了排序这一软件设计中常用的运算,重点介绍了插入排序、希尔排序、冒泡排序、快速排序、直接选择排序、堆排序、归并排序、基数排序的方法、算法及其性能分析;第八章所介绍的查找也是软件设计中常用的基本运算,着重讨论了在顺序表、树表和散列表等几种表结构上的查找方法实现及其性能;第九章对有关文件方面的基本知识作了简要介绍。

通过本课程的学习,一方面,学生将得到必要的程序设计技能训练;另一方面,为学习《操作系统》、《编译原理》、《数据库系统》、《软件工程》、《人工智能》等后续课程提供必要的知识准备。因此通过本课程的学习,学生应达到两个目的:

1. 知识方面。通过学习常用的数据结构(如线性表、栈、队列、树、图等),了解数据结构内在的逻辑关系,了解数据结构在计算机中的表示方法。

2. 技能方面。通过典型应用,理解数据在不同的存储结构中,实现各种数据运算的不同方法。通过算法设计和算法分析,提高学生分析问题和解决问题的能力,包括提高程序设计的能力。

本书在内容上丰富充实,结构上紧凑合理,对各种常用数据结构的介绍尽量做到从实例出发,通过对实例的分析,使学生了解数据结构的基本概念。在算法介绍中,主要选取常规、简单的算法,舍弃结构复杂的算法。书中重要的算法都给出了C语言描述,为学生上机提供方便。本书每章后都进行小结并配有适量习题,便于读者掌握各章的重点和难点并进行必要的训练。学生可通过完成习题,加深对基本概念的理解,特别是通过完成算法类的习题,提高编程能力和程序调试能力。

本书由国防科技大学蹇强、罗宇编著,既可作为高等院校计算机科学与技术专业的数据结构课程的教材,也可作为非计算机专业数据结构类课程的教科书。对从事计算机应用的工程技术人员,本书也是一本十分有价值的参考书。书中不打“*”的章节可用60学时的时间讲授,全部内容讲授可在80~100学时内完成。

由于本书作者水平有限,加之时间仓促,难免存在错误或不当之处,敬请广大读者和专家批评指正。

编 者
2004年3月

目 录

第一章 绪 论	(1)
1.1 数据结构概述	(1)
1.2 数据结构的基本概念	(3)
1.2.1 基本概念	(3)
1.2.2 数据结构及其分类	(4)
1.2.3 数据结构课程的内容	(7)
1.3 数据类型和抽象数据类型	(8)
1.3.1 数据类型	(8)
1.3.2 抽象数据类型	(9)
1.4 算法和算法分析	(11)
1.4.1 算法概念	(11)
1.4.2 算法分析	(12)
1.5 小结	(14)
习题 1	(16)
第二章 线性表	(17)
2.1 线性表的类型定义	(17)
2.1.1 线性表的定义	(17)
2.1.2 线性表的基本操作	(18)
2.2 线性表的顺序表示和实现	(19)
2.2.1 顺序表的存储结构	(19)
2.2.2 顺序表的基本操作	(20)
2.2.3 顺序表其他算法举例	(22)
2.3 线性表的链接存储结构及其运算	(27)
2.3.1 单链表和指针	(28)
2.3.2 单链表的基本操作	(30)
2.3.3 单链表的其他操作举例	(37)
2.3.4 循环链表	(47)
2.3.5 双向链表	(49)
2.4 小结	(51)

习题 2	(53)
第三章 栈和队列	(55)
3.1 栈.....	(55)
3.1.1 栈的定义及基本操作.....	(55)
3.1.2 栈的表示和实现.....	(56)
3.2 栈的应用举例.....	(60)
3.3 队列.....	(77)
3.3.1 队列的定义.....	(77)
3.3.2 队列的表示和实现.....	(78)
3.4 队列应用举例.....	(82)
3.5 小结.....	(86)
习题 3	(86)
第四章 串、数组和广义表.....	(88)
4.1 串的定义和操作.....	(88)
4.2 串的存储结构.....	(90)
4.2.1 串的顺序存储.....	(90)
4.2.2 串的堆分配存储.....	(92)
4.2.3 串的链式存储.....	(95)
4.3 串的模式匹配算法.....	(96)
4.3.1 Brute-Force 算法	(96)
* 4.3.2 KMP 算法	(99)
4.4 串操作应用举例	(103)
4.5 数组	(105)
4.5.1 数组的概念	(105)
4.5.2 数组的顺序存储结构	(107)
4.5.3 数组的应用举例	(109)
4.6 矩阵的压缩存储	(112)
4.6.1 特殊矩阵的压缩存储	(112)
4.6.2 稀疏矩阵及存储	(114)
4.6.3 三元组表	(115)
* 4.6.4 十字链表	(118)
* 4.7 广义表	(119)
4.7.1 广义表的递归定义	(119)
4.7.2 广义表的存储结构	(121)

4.7.3 广义表的递归算法	(124)
4.8 小结	(127)
习题 4	(129)
第五章 树和二叉树	(132)
5.1 树的概念	(132)
5.1.1 树的定义	(132)
5.1.2 树的基本术语	(134)
5.2 二叉树	(134)
5.2.1 二叉树的定义和基本术语	(134)
5.2.2 二叉树的几个基本性质	(136)
5.2.3 二叉树的存储结构	(137)
5.3 二叉树的遍历	(140)
5.3.1 二叉树的遍历方法	(140)
5.3.2 二叉树遍历算法描述	(141)
5.3.3 二叉树遍历应用举例	(149)
5.3.4 线索二叉树	(154)
5.4 树和森林	(158)
5.4.1 树和森林的存储结构	(158)
5.4.2 树和森林的遍历	(162)
5.5 树的应用	(166)
5.5.1 集合的表示	(166)
5.5.2 等价问题	(168)
* 5.6 哈夫曼树及其应用	(169)
5.6.1 哈夫曼树的基本概念	(169)
5.6.2 判定树	(171)
5.6.3 哈夫曼编码	(173)
5.7 小结	(176)
习题 5	(177)
第六章 图	(180)
6.1 图的概念	(180)
6.1.1 图的定义和术语	(180)
6.1.2 图有关术语的示例	(182)
6.2 图的存储结构	(184)
6.2.1 图的邻接矩阵存储表示	(184)

6.2.2 图的邻接表存储表示	(186)
6.2.3 图的建立算法	(188)
6.3 图的遍历	(191)
6.3.1 深度优先搜索遍历	(191)
6.3.2 广度优先搜索遍历	(195)
6.4 连通网的最小生成树	(197)
6.4.1 克鲁斯卡尔(Kruskal)算法	(199)
6.4.2 普里姆(Prim)算法	(200)
6.5 最短路径问题	(202)
6.5.1 单源最短路径	(202)
* 6.5.2 每一对顶点间的最短路径	(205)
* 6.6 拓扑排序	(207)
* 6.7 关键路径	(211)
6.8 小结	(213)
习题 6	(215)
第七章 查 找	(218)
7.1 基本概念	(218)
7.2 静态查找	(220)
7.2.1 静态查找的基本操作	(220)
7.2.2 静态查找表的顺序存储结构	(220)
7.2.3 顺序查找	(220)
7.2.4 折半查找	(222)
7.2.5 分块查找	(224)
7.3 动态查找	(227)
7.3.1 动态查找的基本操作	(227)
7.3.2 动态查找表的二叉链表存储结构	(227)
7.3.3 二叉排序树	(228)
7.3.4 平衡二叉树	(233)
* 7.3.5 B 树	(238)
7.4 散列表	(239)
7.4.1 散列表的概念	(239)
7.4.2 散列函数的构造方法	(241)
7.4.3 处理冲突的方法	(244)
* 7.4.4 散列表的查找和分析	(247)

7.5 查找操作应用举例	(249)
7.6 小结	(251)
习题7	(252)
第八章 排 序	(255)
8.1 基本概念	(255)
8.2 插入排序法	(257)
8.2.1 直接插入排序	(257)
8.2.2 希尔排序	(259)
8.3 交换排序法	(260)
8.3.1 冒泡排序	(260)
8.3.2 快速排序	(262)
8.4 选择排序法	(265)
8.4.1 直接选择排序	(265)
8.4.2 堆排序	(267)
8.5 归并排序法	(273)
8.5.1 两个有序序列的归并	(273)
8.5.2 一趟归并排序	(274)
8.6 基数排序法	(275)
8.6.1 多关键字排序	(275)
8.6.2 链式基数排序	(276)
8.7 各种内部排序法的比较	(281)
8.8 排序操作应用举例	(282)
8.9 小结	(284)
习题8	(285)
*第九章 文 件	(287)
9.1 外部存储设备简介	(287)
9.1.1 磁带存储设备	(287)
9.1.2 磁盘存储设备	(289)
9.1.3 光盘存储设备	(289)
9.2 有关文件的基本概念	(290)
9.3 顺序文件	(291)
9.4 索引文件	(293)
9.5 ISAM 文件和 VSAM 文件	(294)

9.5.1 ISAM 文件	(294)
9.5.2 VSAM 文件	(296)
9.6 倒排文件和多重表文件	(298)
9.7 散列文件	(299)
9.8 小结	(300)
习题 9	(300)
参考文献	(301)

第一章 绪 论

计算机的数据处理能力是计算机解决各种实际问题的基础,但是现实世界中的实际问题必须经过抽象,得出反映实际事物本质的数据表示后,才有可能被计算机处理。如何从实际问题抽象出它的数学模型,得出它的数据表示,这不属于本课程的内容,但如何用计算机所能接受的形式来描述这些数据(包括数据本身及数据与数据之间的关系),如何将这些数据以及它们之间的关系存储在计算机中,如何用有效的方法去处理这些数据,则是数据结构课程要研究的主要问题。

1.1 数据结构概述

早期的计算机价格昂贵,运算速度慢,存储容量也有限,主要应用于科学计算。其数据结构的特点是数据类型简单、算法复杂,所以侧重于建立程序,称之为数值计算。随着计算机技术的发展与普及,计算机从单纯的科学计算发展到控制管理和数据处理等非数值计算的工作中,像银行业务处理、工商企业管理和政府部门管理等。其应用特点是需要处理大量的数据,但数据处理过程相对比较简单,处理对象不仅可以是数值,还可以是字符、表格、声音、图像等各种信息。面对大批要处理的数据,为了有效地进行处理,就必须研究数据自身的内在结构。

例 1.1 一个大学的学生健康状况管理。

表 1.1 中的学生健康状况登记表就是一个数据结构。表中每个学生的情况为一个记录,它由姓名、学号、性别、年龄、班级和健康状况等 6 个数据项组成。计算机学生健康状况管理的主要功能包括:查询、浏览、插入、修改、删除和统计等。

表 1.1 学生健康情况登记表

姓名	学号	性别	年龄	班级	健康状况
王逸新	010631	男	18	五	良好
曲向丽	010632	女	17	六	良好
袁志坚	010633	男	19	七	良好
杨雁萍	010634	女	18	八	良好
.....					

在这种数据结构中,计算机处理的数据之间存在的是一种“一个对一个”的简单线性关系,称为线性数据结构。

例 1.2 一个大学的人事档案管理。

图 1.1 中的数据结构像一棵树,根在上树枝在下的倒挂的树,它清晰地描述了教师和学生所在的系和专业。在这种结构中可以从树根沿着某系某专业很快找到某个教师或学生的数据,查找的过程就是从树根沿树枝到某个叶子的过程。

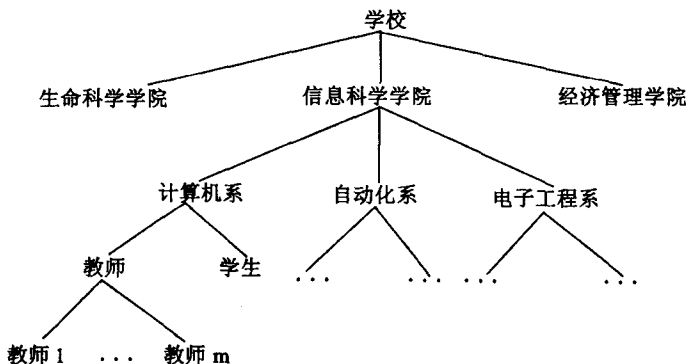


图 1.1 教师学生名册

在这种数据结构中,计算机处理的数据之间存在的是一种“一个对多个”的层次关系,称为树型数据结构。

例 1.3 在 n 个城市之间建立通信网络,要求在任意两个城市之间都有直接的或间接的通信线路,在已知某些城市之间直接通信线路预算造价的情况下,使网络的造价最低。

当 n 很大时,这样的问题只能用计算机来求解。用图 1.2(a)来描述 7 个城市之间的通信线路,其中:图中圆圈表示一个城市,两个圆圈之间的连线表示对应城市之间的通信线路,连线上的数值表示该通信线路的造价。这一描述的结构为图状结构,利用计算机可以求出满足要求的通信网络,如图 1.2(b)所示。

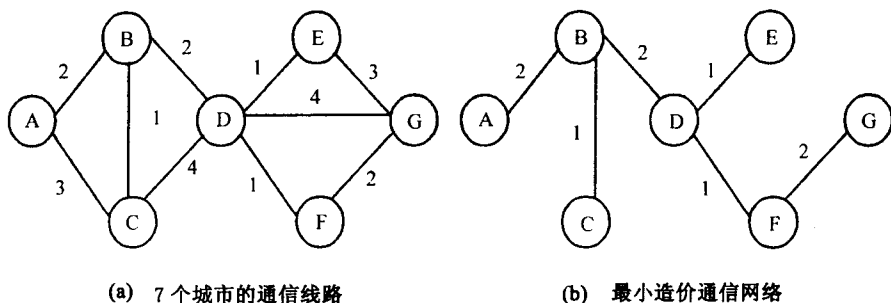


图 1.2 7 个城市的通信线路及最小造价通信网络

通过上面的 3 个例子可以看出:数据结构中数据元素之间存在着逻辑关系。一般说来,一个具体问题的软件设计通常包括三个步骤:(1)分析和确定功能以及该问题的逻辑数据结构;(2)设计该问题的具体存储结构;(3)设计该问题在具体存储结构下的操作实现算法。

1.2 数据结构的基本概念

1.2.1 基本概念

1. 数据和信息

数据是客观事物的符号表示,是对现实世界的事物采用计算机能够识别、存储和处理的形式进行描述的符号的集合;信息指的是其中的含义,不同的形式可以传达同样的信息。计算机能处理多种形式的数。例如,科学计算软件处理的是数值数据;文字处理软件处理的是字符数据;多媒体软件处理的是图像、声音等多媒体数据。

2. 集合和关系

通俗地说,集合就是一堆东西,其中每个东西成为该集合的一个元素。就集合本身而言,元素之间无所谓先后之分。一个班的学生是一个集合,其中选出的若干个人也构成一个集合,并称为上述“班”集合的子集。通常所说的集合不含有相同的元素,例如 3 个整数 2、2 和 5 不能构成一个集合。集合通常借助花括弧表示,一种方法是直接列出集合中的元素,如集合 $\{11, 12, \dots, 99\}$;另一种方法是用条件规定,如上述集合可以描述为 $\{x \mid x \text{ 是整数且 } 10 < x < 100\}$ 。用一个字母代表集合可以简化叙述,例如用 Z 表示整数集合,记为 $Z = \{0, \pm 1, \pm 2, \dots\}$ 。

在实际问题中,集合的元素之间常常具有某种关系。例如学生的一个班组成的集合中,班长与组长以及组长与组员之间都有“领导”关系,班内成员之间可能具有“友好”关系、“敬慕”关系等。一般地说,对于一个集合,可以在其上定义一种或多种关系。例如一个3人小组的集合表示为 $A = \{e_1, e_2, e_3\}$, 其中 e_1 是组长, e_2 和 e_3 是同桌同学,则在集合 A 上存在两种关系:领导关系 $r_1 = \{e_1 \rightarrow e_2, e_1 \rightarrow e_3\}$, 同桌关系 $r_2 = \{e_2 \rightarrow e_3\}$ 。

3. 数据元素(data element)

数据元素是数据的基本单位,在计算机程序中通常作为一个整体进行考虑和处理。数据元素可以是不可分割的“原子”。例如一个整数“5”或一个字符“N”,也可以由若干款项组成,例如在学籍管理程序中处理的一个学生的基本情况就是一个数据元素,它包括学号、姓名、性别和出生日期等4项,其中每一个款项称为一个“数据项(data item)”。如果一个数据项由若干款项组成(如出生日期),则称为组合项;否则称为原子项(如性别)。非原子的数据元素也是一种组合项。图 1.3 所示即为上述数据元素的内部结构。数据元素有时也称为记录、结点或顶点。

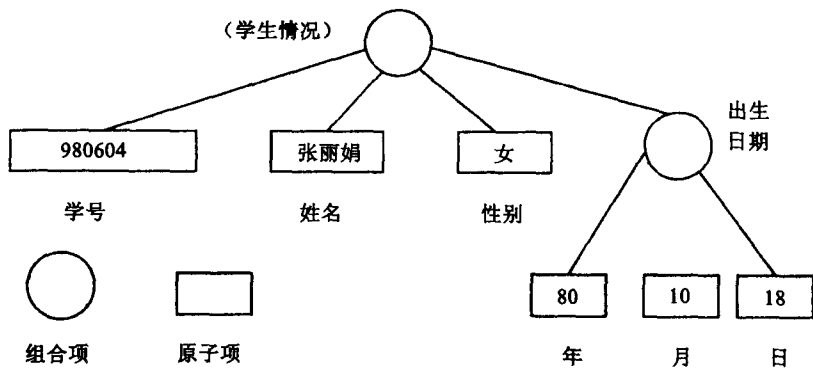


图 1.3 数据元素的内部结构

4. 关键码(key)

关键码指的是数据元素中能起标识作用的数据项,例如,学生情况中的学号和姓名。其中能起唯一标识作用的关键码称为“主关键码”(简称主码),如学号;反之称为“次关键码”(简称次码),如姓名。

1.2.2 数据结构及其分类

1. 数据结构的定义

一个特性相同的数据元素的集合,如果在数据元素之间存在一种或多种特