

高等学校教材·计算机科学与技术

可赠送课件素材

jsjc@tup.tsinghua.edu.cn

数据结构(C语言描述)

徐孝凯 贺桂英 编著



清华大学出版社

高等学校教材·计算机科学与技术

数据结构（C语言描述）

徐孝凯 贺桂英 编著

清华大学出版社
北 京

内 容 简 介

本书是利用 C 语言编写的一本数据结构教材, 适合在学习 C 语言之后使用。全书介绍了各种常用而具体的数据结构、对应的存储结构、以及各种典型运算的方法和算法。书中含有丰富而实用的算法实例, 这些算法都具有较好的可读性、结构化和时空有效性, 通过深入地学习和分析, 能够大大提高软件开发和设计能力。本书适合作为各级各类学校开设数据结构课程的教材或教学参考书, 也适合软件开发人员参考。

版权所有, 翻印必究。举报电话: 010-62782989 13901104297 13801310933

本书封面贴有清华大学出版社激光防伪标签, 无标签者不得销售。

图书在版编目(CIP)数据

数据结构 (C 语言描述) / 徐孝凯, 贺桂英编著. —北京: 清华大学出版社, 2004.10

(高等学校教材·计算机科学与技术)

ISBN 7-302-09358-X

I. 数… II. ①徐… ②贺… III. ①数据结构—高等学校—教材 ②C 语言—程序设计—高等学校—教材 IV. ①TP311.12 ②TP312

中国版本图书馆 CIP 数据核字 (2004) 第 088285 号

出 版 者: 清华大学出版社

地 址: 北京清华大学学研大厦

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

邮 编: 100084

社 总 机: 010-62770175

客户服务: 010-62776969

责任编辑: 郑寅堃

印 刷 者: 清华园胶印厂

装 订 者: 北京鑫海金澳胶印有限公司

发 行 者: 新华书店总店北京发行所

开 本: 185×260 印张: 18 字数: 442 千字

版 次: 2004 年 10 月第 1 版 2004 年 10 月第 1 次印刷

书 号: ISBN 7-302-09358-X/TP·6537

印 数: 1~5000

定 价: 25.00 元

本书如存在文字不清、漏印以及缺页、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) 62770175-3103 或 (010) 62795704

清华大学出版社计算机教材

编审委员会成员

(按地区排序)

清华大学	周立柱 教授	北京大学	杨冬青 教授
	覃 征 教授		陈 钟 教授
	王建民 教授		陈立军 副教授
	刘 强 副教授	中国人民大学	王 珊 教授
	冯建华 副教授		孟小峰 教授
北京航空航天大学	马殿富 教授		陈 红 教授
	吴超英 副教授	北京信息工程学院	孟庆昌 教授
	姚淑珍 教授	石油大学(北京)	陈 明 教授
北京交通大学	阮秋琦 教授	天津大学	艾德才 教授
北京科技大学	杨炳儒 教授		
南开大学	吴功宜 教授		
复旦大学	吴立德 教授	上海交通大学	傅育熙 教授
	吴百锋 教授		蒋建伟 副教授
	杨卫东 副教授	华东师范大学	杨宗源 教授
华东理工大学	邵志清 教授		应吉康 教授
东华大学	乐嘉锦 教授	上海第二工业大学	蒋川群 教授
浙江大学	吴朝晖 教授	南京大学	骆 斌 教授
	李善平 教授	南京理工大学	张功萱 教授
南京航空航天大学	秦小麟 教授	苏州大学	龚声蓉 教授
南京邮电学院	朱秀昌 教授		
江苏大学	宋余庆 教授		
武汉大学	何炎祥 教授	华中科技大学	刘乐善 教授
中南财经政法大学	刘腾红 教授		朱定华 教授
武汉理工大学	李中年 教授	华中师范大学	魏开平 教授
			王林平 副教授

国防科技大学

赵克佳 教授
肖 依 副教授

湖南大学

林亚平 教授
邹北骥 教授

西安交通大学

沈钧毅 教授
齐 勇 教授
巨永峰 教授
陈莉君 副教授

长安大学

西安邮电学院

哈尔滨工业大学

长春工程学院

郭茂祖 教授
沙胜贤 教授

山东大学

孟祥旭 教授
郝兴伟 教授

中山大学

福州大学

潘小轰 教授
林世平 副教授

云南大学

西南交通大学

刘惟一 教授
杨 燕 副教授

中南大学

陈松乔 教授

西北大学

西安石油学院

周明全 教授
方 明 教授

吉林大学

何 桥 教授
徐一平 教授
毕 强 教授

山东科技大学

郑永果 教授

厦门大学

冯少荣 副教授

重庆邮电学院

王国胤 教授

出版说明

改革开放以来,特别是党的十五大以来,我国教育事业取得了举世瞩目的辉煌成就,高等教育实现了历史性的跨越,已由精英教育阶段进入国际公认的大众化教育阶段。在质量不断提高的基础上,高等教育规模取得如此快速的发展,创造了世界教育发展史上的奇迹。当前,教育工作既面临着千载难逢的良好机遇,同时也面临着前所未有的严峻挑战。社会不断增长的高等教育需求同教育供给特别是优质教育供给不足的矛盾,是现阶段教育发展面临的基本矛盾。

教育部一直十分重视高等教育质量工作。2001年8月,教育部下发了《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》,提出了十二条加强本科教学工作提高教学质量的措施和意见。2003年6月和2004年2月,教育部分别下发了《关于启动高等学校教学质量与教学改革工程精品课程建设工作的通知》和《教育部实施精品课程建设提高高校教学质量和人才培养质量》文件,指出“高等学校教学质量和教学改革工程”,是教育部正在制订的《2003—2007年教育振兴行动计划》的重要组成部分,精品课程建设是“质量工程”的重要内容之一,教育部计划用五年时间(2003—2007年)建设1500门国家级精品课程,利用现代化的教育信息技术手段将精品课程的相关内容上网并免费开放,以实现优质教学资源共享,提高高等学校教学质量和人才培养质量。

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,顺应并符合新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”。首批推出的特色精品教材包括以下三个系列:

(1) 高等学校教材·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) 高等学校教材·计算机科学与技术——高等学校计算机相关专业的教材。

(3) 高等学校教材·电子信息——高等学校电子信息相关专业的教材。

清华大学出版社经过近二十年的努力，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。清华版教材经过二十多年的精雕细刻，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。

总策划 李家强

策 划 卢先和 丁 岭

清华大学出版社教材编审委员会

E-mail: dingl@tup.tsinghua.edu.cn

luxh@tup.tsinghua.edu.cn

前 言

数据结构主要研究数据的逻辑结构、在计算机中的存储结构以及对数据进行各种非数值运算的方法和算法。数据的逻辑结构分为集合、线性、树（层次）、图（网状）等四种基本结构，由它们可以构成任何较复杂的逻辑结构。数据的存储结构分为顺序、链接、索引、散列等四种基本结构，同样由它们能够构成各种较复杂的存储结构。对数据进行的非数值运算主要包括查找、排序、插入、删除、修改、遍历等。对于同样的数据，若采用的逻辑结构和存储结构不同，对某一运算所采用的方法不同，则将得到不同的算法，进而在计算机上会有不同的运行时间和存储空间效率。通过该课程的学习，读者能够根据实际应用中对数据处理的要求，为数据选择和建立合适的逻辑结构和存储结构，接着选择和使用较好的数据处理方法，以及利用一种程序设计语言编写出相应的算法，最后在计算机系统上调试、运行和实现算法。

本书是根据一般计算机及相关专业对开设数据结构课程的知识结构要求编写的，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过时不用的内容；在写法上力求条理清楚、层次分明、内容连贯、循序渐进、便于阅读和自学；在各种运算方法和算法的分析上，力求细致、生动、深入、透彻、便于理解。

本书共包括 9 章。第 1 章概要介绍数据的逻辑结构和存储结构，以及算法的描述和评价。第 2 至 4 章主要介绍线性表、栈、队列、稀疏矩阵、广义表等线性数据结构的定义、存储结构以及相应运算的方法和算法。第 5、6 章主要介绍树、二叉树、二叉搜索树、堆、哈夫曼树等树结构的定义和存储结构，以及建立、遍历、查找等运算的方法和算法。第 7 章主要介绍图的定义和存储结构，以及图的建立及其遍历、求最小生成树、拓扑排序等运算的方法和算法。第 8 章主要介绍顺序、二分、索引、散列等查找的方法和算法。第 9 章主要介绍插入、选择、堆、快速、二路归并、外存文件等排序的方法和算法。

本书给出的所有算法和程序都在 C 语言运行环境下调试通过，同时，由于 C++ 兼容 C 语言，作者也在 VC++6.0 环境下调试通过所有算法和程序，因此能够保证算法是正确和有效的。

本书每章后面都配有丰富的练习题，读者通过做练习和上机调试程序，能够巩固和提高所学知识，提高分析问题和解决问题的能力。由于数据结构习题一般较难，有时不太容易做出，为了使读者得到帮助和启发，专门为本书配有学习指导与习题解答一书，并同本书一同出版发行。当然，读者要尽量独立思考完成作业，绝不能依赖现成的答案。

使用本书推荐讲授学时为 64，上机实验学时为 32，实验题目可采用教材中所给的调试程序，也可由实验指导教师指定。数据结构课程所涉及的内容较多，而教学时数有限，所以任课教师可根据实际情况对本书中的内容进行取舍。另外，对于加星号的内容和习题，由于较难，可以不作教学要求。

本书由徐孝凯教授和贺桂英副教授共同编写。在教学大纲构思和教材内容编写的过程中得到了巫家敏教授的真诚指导，在此表示衷心感谢！

由于作者水平有限，不足之处在所难免，敬请同行专家和广大读者指正。电子邮箱地址为 xuxk@crtvu.edu.cn，联系电话为 010-64910302。

徐孝凯

2004 年 2 月

目 录

第 1 章 绪论	1
1.1 基本概念	1
1.2 算法描述	8
1.3 算法评价	9
习题一	15
第 2 章 线性表	18
2.1 线性表的定义和操作	18
2.2 线性表的顺序存储结构和操作实现	20
2.2.1 线性表的顺序存储	20
2.2.2 顺序存储下线性表的操作实现	21
2.3 线性表的链接存储结构	31
2.3.1 链接存储的概念	31
2.3.2 线性表的链接存储	32
2.3.3 在单链表上的插入和删除操作	33
2.3.4 单链表中的结点类型	34
2.3.5 双向链表中的结点类型和插入与删除操作	37
2.3.6 带表头附加结点的线性链表	39
2.3.7 循环链表	39
2.4 线性表操作在单链表上的实现	40
习题二	51
第 3 章 稀疏矩阵和广义表	54
3.1 稀疏矩阵	54
3.1.1 稀疏矩阵的定义	54
3.1.2 稀疏矩阵的存储结构	56
3.1.3 稀疏矩阵的运算	58
3.2 广义表	66
3.2.1 广义表的定义	66
3.2.2 广义表的存储结构	68
3.2.3 广义表的运算	69
3.2.4 简单程序举例	73
习题三	74

第 4 章 栈和队列	76
4.1 栈	76
4.1.1 栈的定义	76
4.1.2 栈的运算概述	76
4.2 栈的顺序存储结构和操作实现	77
4.3 栈的链接存储结构和操作实现	81
4.4 栈的简单应用举例	84
4.5 算术表达式的计算	88
4.5.1 算术表达式的两种表示	88
4.5.2 后缀表达式求值的算法	89
4.5.3 把中缀表达式转换为后缀表达式的算法	92
4.6 栈与递归	95
4.7 队列	103
4.7.1 队列的定义	103
4.7.2 队列的运算概述	103
4.7.3 队列的顺序存储结构和操作实现	104
4.7.4 队列的链接存储结构和操作实现	109
4.7.5 队列的应用简介	111
习题四	112
第 5 章 树和二叉树	116
5.1 树的概念	116
5.1.1 树的定义	116
5.1.2 树的表示	117
5.1.3 树的基本术语	118
5.1.4 树的性质	119
5.2 二叉树	120
5.2.1 二叉树的定义	120
5.2.2 二叉树的性质	121
5.2.3 二叉树的运算概述	122
5.2.4 二叉树的存储结构	123
5.3 二叉树遍历	126
5.4 二叉树的其他运算	130
5.5 树的存储结构和运算	136
5.5.1 树的运算概述	136
5.5.2 树的存储结构	136
5.5.3 树的运算	139
习题五	145

第6章 二叉树的应用	151
6.1 二叉搜索树	151
6.1.1 二叉搜索树的定义	151
6.1.2 二叉搜索树的运算概述	151
6.1.3 二叉搜索树的运算	152
6.2 堆	160
6.2.1 堆的定义	160
6.2.2 堆的运算概述	160
6.2.3 堆的存储结构	161
6.2.4 堆的运算	162
6.3 哈夫曼树	167
6.3.1 基本术语	167
6.3.2 构造哈夫曼树	168
*6.3.3 哈夫曼编码	170
习题六	174
第7章 图	176
7.1 图的概念	176
7.1.1 图的定义	176
7.1.2 图的基本术语	177
7.2 图的存储结构	180
7.2.1 邻接矩阵	180
7.2.2 邻接表	182
7.2.3 边集数组	185
7.3 图的遍历	186
7.3.1 深度优先搜索遍历	186
7.3.2 广度优先搜索遍历	189
7.3.3 非连通图的遍历	191
7.4 图的生成树和最小生成树	194
7.4.1 普里姆算法	196
7.4.2 克鲁斯卡尔算法	199
7.5 拓扑排序	204
习题七	211
第8章 查找	213
8.1 查找的基本概念	213
8.2 顺序表查找	214
8.2.1 顺序查找	214

8.2.2 二分查找	215
8.3 索引查找	219
8.3.1 索引的概念	219
8.3.2 索引查找算法	222
8.3.3 分块查找	224
8.4 散列查找	225
8.4.1 散列的概念	225
8.4.2 散列函数	227
8.4.3 处理冲突的方法	229
8.4.4 散列表的运算	232
8.5 B 树查找	240
8.5.1 B_树的定义	240
8.5.2 B_树查找	241
8.5.3 B_树的插入	243
8.5.4 B_树的删除	245
习题八	247
第 9 章 排序	249
9.1 排序的基本概念	249
9.2 插入排序	250
9.3 选择排序	252
9.3.1 直接选择排序	252
9.3.2 堆排序	253
9.4 交换排序	257
9.4.1 气泡排序	257
9.4.2 快速排序	258
9.5 归并排序	262
9.6 各种内排序方法的比较	265
9.7 外排序	266
习题九	274
参考文献	276

第1章 绪 论

数据结构课程是计算机及相关专业中的一门专业基础课，它介绍和研究数据在计算机中的组织、存储和处理的方法。这里所说的数据的概念是广义的，它不仅表示单一的数据，如字符、数值等，而且表示带结构的数据，如记录、数组、矩阵、登记表、结构图等。数据在计算机中的组织和存储方法有顺序、链接、散列、索引等多种，根据数据处理的需要可从中选择一种或几种的组合来存储数据。对数据进行处理的方法又叫做算法，它是根据数据处理的实际需要而逐渐产生和发展起来的。现在人们已经总结出进行数据处理的各种具体、实用和有效的算法，根据这些算法和存储在计算机中的数据，再利用一种算法描述语言（如 C 语言）和面向过程或对象的程序设计方法就能够编写出进行数据处理的程序，通过计算机运行这个程序自动完成特定的数据处理任务。学习数据结构课程除了要学习和研究已有的数据存储结构和数据处理算法之外，更重要的是根据自己解决实际问题的需要，进行有效的数据存储和数据处理。

1.1 基本概念

数据（data）是人们利用文字符号、数字符号以及其他规定的符号对现实世界的事物及其活动所做的抽象描述。例如，一个人的名字可以用一个字符串来描述，一个图形可以用一个数组来描述，数组中的每一个值用来存储图形中对应点的坐标值。因此，一个文档、记录、数组、句子、单词、算式、符号等都统称为数据。在计算机领域，人们把能够被计算机加工的对象，或者说能够被计算机输入、存储、处理和输出的一切信息都叫做数据。

数据元素（data element）简称元素，它是一个数据整体中相对独立的单位。如对于一个文件来说，每个记录就是它的数据元素；对于一个字符串来说，每个字符就是它的数据元素；对于一个数组来说，每一个下标位置上的数据就是它的数据元素。数据和数据元素是相对而言的，如对于一个记录来说，它是所属文件的一个数据元素，而它相对于所含的数据项而言又可看做数据。因此，在本教材中，对数据和数据元素这两个术语的使用并不加以严格区分。

数据记录（data record）简称记录，它是数据处理领域组织数据的基本单位，数据中的每个数据元素在许多应用场合被组织成记录的结构。一个数据记录由一个或多个**数据项（item）**所组成，每个数据项可以是简单数据项（即不可再分，如一个数值、一个字符等），也可以是组合数据项（即数组或记录等）。如对于一个图书目录登记表，它由若干个记录所组成，每个记录都由登录号、书号、书名、作者、出版社、定价等数据项所组成。

在一个表或文件中，若所有记录的某个数据项的值都不同，也就是说，每个值能够惟一地标识一个记录时，则可把这个数据项作为该表或文件的关键数据项，简称**关键项（key item）**，关键项中的每一个值称为所在记录的**关键字（key word 或 key）**。如在图书目录登

记表中, 登录号数据项的值是互不相同的, 所以可把登录号作为图书目录登记表的关键项, 其中的每一个值 (即每本书的登录号) 就是该表的关键字。

在一个表或文件中, 能作为关键项的数据项可能没有, 可能只有一个, 也可能多于一个。当没有时, 可把多个有关的数据项联合起来, 构成一个组合关键项, 用组合关键项中的每一个组合值来惟一地标识一个记录, 该组合值就是所在记录的关键字。如在旅店管理中, 可把楼号和房间号联合起来作为关键字, 用它来惟一标识一个房间。

引入了记录的关键项和关键字后, 为简便起见, 在以后的讨论中, 经常利用关键项来代替所有记录, 利用关键字来代替所对应的记录。

数据处理 (data processing) 是指对数据进行查找、插入、删除、合并、拆分、排序、统计、计算、转换、输入、输出、传送等的操作过程。在早期, 计算机主要用于科学和工程计算, 而现在则主要用于数据处理。像计算机情报检索系统、网上订票系统、图书管理系统、物资调配系统、银行存取款系统、财务管理系统等都是计算机在数据处理领域的具体应用。

数据结构 (data structure) 是指数据以及相互之间的联系。上面提到, 数据的描述对象是现实世界的事物及其活动, 而任何事物及其活动都不是孤立存在的, 都是在一定意义上相互联系、相互影响的, 所以数据之间必然存在着联系。数据之间的相互联系, 被称为数据的**逻辑结构**。在计算机中存储数据时, 不仅要存储数据本身, 而且要存储它们之间的联系 (即逻辑结构)。一种数据结构在存储器中的存储方式称为数据的**物理结构**或**存储结构**。由于存储方式有顺序、链接、索引、散列等多种形式, 所以, 一种数据结构可以根据应用的需要表示成任一种或几种组合的存储结构。数据的逻辑结构和存储结构都反映数据的结构, 但通常所说的数据结构是指数据的逻辑结构, 不包含存储结构的含义。

为了更确切地描述一种数据结构, 通常采用二元组表示:

$$B=(K,R)$$

B 代表一种数据结构, 它由数据元素的集合 K 和 K 上二元关系的集合 R 所组成。其中

$$K=\{k_i \mid 1 \leq i \leq n, n \geq 0\}$$

$$R=\{r_j \mid 1 \leq j \leq m, m \geq 0\}$$

其中 k_i 表示集合 K 中的第 i 个数据元素, n 为 K 中数据元素的个数, 特殊地, 若 $n=0$, 则 K 是一个空集, 此时 B 也就无结构而言, 有时也可以认为它具有任一结构。 r_j 表示集合 R 中的第 j 个二元关系 (以后均简称关系), m 为 R 中关系的个数, 特别地, 若 $m=0$, 则 R 是一个空集, 表明不考虑集合 K 中的元素之间存在任何关系, 彼此是独立的, 就像数学中集合里的元素一样。在本书所讨论的数据结构中, 一般只讨论 $m=1$ 的情况, 即 R 中只包含一个关系 ($R=\{r\}$) 的情况。对于包含有多个关系的数据结构, 可分别对每一个关系进行讨论。

K 上的一个关系 r (以后直接用大写 R 表示) 是序偶的集合。对于 R 中的任一序偶 $\langle x, y \rangle (x, y \in K)$, 我们把 x 叫做序偶的第一元素, 把 y 叫做序偶的第二元素, 又称序偶的第一元素为第二元素的直接前驱 (通常简称前驱), 称第二元素为第一元素的直接后继 (通常简称后继)。如在 $\langle x, y \rangle$ 的序偶中, x 为 y 的前驱, 而 y 为 x 的后继。

一种数据结构还能够利用图形形象地表示出来, 图形中的每个结点 (或叫顶点) 对应着一个数据元素, 两结点之间带箭头的连线 (称做有向边或弧) 对应着关系中的一个序偶,

其中序偶的第一元素为有向边的起始结点，第二元素为有向边的终止结点，即箭头所指向的结点。

作为例子，下面根据表 1-1 构造出一些典型的数据结构。

表 1-1 某个公司人事简表

职工号	姓名	性别	出生日期	职务	部门
01	万明华	男	1962-03-20	经理	
02	赵宁	男	1968-06-14	主管	销售部
03	张利	女	1964-12-07	主管	财务部
04	赵书芳	女	1972-08-05	主任	办公室
05	刘永年	男	1959-08-15	科员	销售部
06	王明理	女	1975-04-01	科员	销售部
07	王敏	女	1972-06-28	科员	财务部
08	张才	男	1967-03-17	科员	财务部
09	马立仁	男	1975-10-12	科员	财务部
10	邢怀常	男	1976-07-05	科员	办公室

表 1-1 中共有 10 条记录，每条记录都由六个数据项所组成。由于每条记录的职工号各不相同，所以可把每条记录的职工号作为该记录的关键字。在下面的例子中，我们将用记录的关键字来代表整个记录。

例 1-1 一种数据结构 $set=(K,R)$ ，其中

$K=\{01,02,03,04,05,06,07,08,09,10\}$

$R=\{\}$

在数据结构 set 中，只存在元素的集合，不存在关系的集合，或者说关系为空，这表明我们只考虑表 1-1 中的每条记录，并不考虑它们之间的任何联系。具有此种特点的数据结构被称为**集合结构**。对于集合结构，也可以看做元素按任一次序排列的线性结构，在存储空间中可以根据需要按任一种存储方式进行存储。

例 1-2 一种数据结构 $linearity=(K,R)$ ，其中

$K=\{01,02,03,04,05,06,07,08,09,10\}$

$R=\{<05,01>,<01,03>,<03,08>,<08,02>,<02,07>,<07,04>,<04,06>,<06,09>,<09,10>\}$

对应的图形如图 1-1 所示。

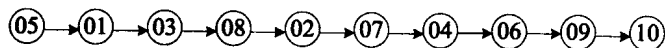


图 1-1 数据的线性结构示意图

结合表 1-1，细心的读者不难看出： R 是按职工年龄从大到小排列的关系。

在 $linearity$ 中，每个数据元素有且仅有一个直接前驱元素（结构中第一个元素 05 除外），有且仅有一个直接后继元素（结构中最后一个元素 10 除外）。这种数据结构的特点是数据元素之间的 1 对 1（1:1）联系，即**线性关系**，我们把具有这种特点的数据结构叫做**线性结构**。

例 1-3 一种数据结构 $tree=(K,R)$, 其中

$$K=\{01,02,03,04,05,06,07,08,09,10\}$$

$$R=\{<01,02>,<01,03>,<01,04>,<02,05>,<02,06>,<03,07>,<03,08>,<03,09>,<04,10>\}$$

对应的图形如图 1-2 所示。

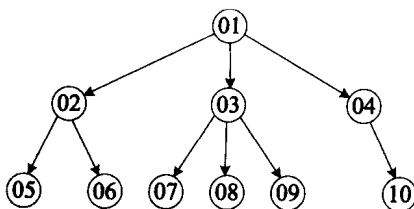


图 1-2 数据的树形结构示意图

结合表 1-1, 细心的读者不难看出: R 是人员之间领导与被领导的关系。

图 1-2 像倒着画的一棵树, 在这棵树中, 最上面的一个没有前驱只有后继的结点叫做树根结点, 最下面一层的只有前驱没有后继的结点叫做树叶结点, 除树根和树叶之外的结点叫做树枝结点 (实际上, 树根结点是一种特殊的树枝结点)。在一棵树中, 每个结点有且只有一个前驱结点 (树根结点除外), 但可以有任意多个后继结点 (树叶结点可看做具有 0 个后继结点)。这种数据结构的特点是数据元素之间的 1 对 N ($1:N$, $N \geq 0$ 联系), 即层次关系, 我们把具有这种特点的数据结构叫做**树形结构**, 简称树。

例 1-4 一种数据结构 $graph=(K,R)$, 其中

$$K=\{01,02,03,04,05,06,07\}$$

$$R=\{<01,02>,<02,01>,<01,04>,<04,01>,<02,03>,<03,02>,<02,06>,<06,02>,<02,07>,<07,02>,<03,07>,<07,03>,<04,06>,<06,04>,<05,07>,<07,05>\}$$

对应的图形如图 1-3 所示。

从图 1-3 可以看出, R 是 K 上的对称关系。为了简化起见, 我们把 $<x,y>$ 和 $<y,x>$ 这两个对称序偶用一个无序对 (x,y) 或 (y,x) 来代替; 在图形表示中, 我们把 x 结点和 y 结点之间两条相反的有向边用一条无向边来代替。这样 R 关系可改写为:

$$r=\{(01,02),(01,04),(02,03),(02,06),(02,07),(03,07),(04,06),(05,07)\}$$

对应的图形如图 1-4 所示。

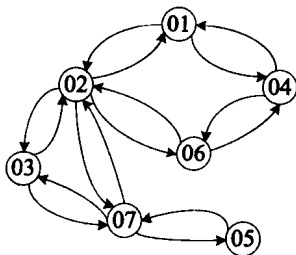


图 1-3 数据的图形结构示意图

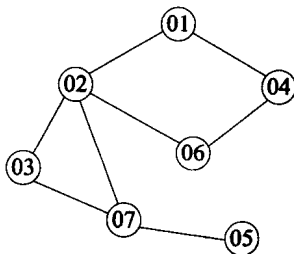


图 1-4 图 1-3 的等价表示

如果说 R 中每个序偶里的两个元素所代表的人员是好友的话, 那么 r 关系就是人员之

间的好友关系。

从图 1-3 或 1-4 可以看出, 结点之间的联系是 M 对 N ($M:N$, $M \geq 0$, $N \geq 0$ 联系), 即**网状关系**。也就是说, 每个结点可以有任意多个前驱结点和任意多个后继结点。我们把具有这种特点的数据结构叫做**图形结构**; 简称**图**。

从图形结构、树形结构和线性结构的定义可知, 树形结构是图形结构的特殊情况 (即 $M=1$ 的情况), 线性结构是树形结构的特殊情况 (即 $N=1$ 的情况)。为了区别于线性结构, 我们把树形结构和图形结构统称为**非线性结构**。

例 1-5 一种数据结构 $B=(K,R)$, 其中

$$K=\{k_1, k_2, k_3, k_4, k_5, k_6\}$$

$$R=\{r_1, r_2\}$$

$$r_1=\{<k_3, k_2>, <k_3, k_5>, <k_2, k_1>, <k_5, k_4>, <k_5, k_6>\}$$

$$r_2=\{<k_1, k_2>, <k_2, k_3>, <k_3, k_4>, <k_4, k_5>, <k_5, k_6>\}$$

若用实线表示关系 r_1 , 虚线表示关系 r_2 , 则对应的图形如图 1-5 所示。

从图 1-5 可以看出: 数据结构 B 是一种非线性的图形结构。但是, 若只考虑关系 r_1 则为树形结构, 若只考虑关系 r_2 则为线性结构。

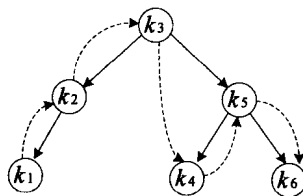


图 1-5 带两个关系的数据结构示意图

数据类型 (data type) 是对数据的取值范围、数据元素之间的结构以及允许施加操作的一种综合描述。每一种计算机高级语言都定义有自己的数据类型。在通用的计算机高级语言中, 一般都具有整数、实数 (浮点数)、枚举、字符、字符串、指针、数组、记录、文件等数据类型。如整数类型在计算机系统中通常用两个字节或四个字节表示, 若采用两个字节, 则整数表示范围在 $-2^{15} \sim 2^{15}-1$, 即 $-32768 \sim 32767$ 之间; 若采用四个字节, 则整数表示范围在 $-2^{31} \sim 2^{31}-1$, 即 $-2147483648 \sim 2147483647$ 之间。对整数类型的数据允许施加的操作通常有: 单目取正或取负运算, 双目加、减、乘、除、取模等运算, 双目等于、不等于、大于、大于等于、小于、小于等于等关系 (比较) 运算以及赋值运算。字符类型在机器中通常用一个字节表示, 无符号表示范围在 $0 \sim 255$ 之间, 能够对至多 256 种字符进行编码。对字符类型的数据允许进行的操作主要为赋值和各种关系运算。字符串类型为字符类型的顺序排列结构, 每一个字符序列 (其最大长度由具体语言规定) 都是字符串类型中的一个值。对字符串的操作主要有求串长度、串拷贝、两串连接、两串比较等。

数据类型可分为简单类型和结构类型两种。简单类型中的每个数据 (即简单数据) 都是无法再分割的整体, 如一个整数、实数、字符、指针、枚举量等都是无法再分割的整体, 所以它们所属的类型均为简单类型。结构类型由简单类型按照一定的规则构造而成, 并且结构类型仍可以包含结构类型, 所以一种结构类型中的数据 (即结构数据) 可以分解为若干个简单数据或结构数据, 每个结构数据仍可再分。如数组就是一种结构类型, 它由固定个数的同一类型顺序排列而成, 数组类型中的每一个数组值包含有固定个数的同一类型数据, 每个数据 (元素) 都可以通过下标运算符直接访问。记录也是一种结构类型, 它由固定个数的不同 (也可以相同) 类型顺序排列而成, 记录类型中的每一个记录值包含有固定个数的不同类型数据, 每个数据都可以通过成员运算符直接访问。另外, 字符串和文件也