

可下载教学资料

<http://www.tup.tsinghua.edu.cn>

21世纪普通高校计算机公共课程规划教材

数据结构实用教程 (C语言版)

赵波 主编

霍利 郑巍 董靓瑜 李静 编著

清华大学出版社

21 世纪普通高校计算机公共课程规划教材

数据结构实用教程 (C语言版)

赵波 主编

霍利 郑巍 董靓瑜 李静 编著

清华大学出版社
北京

内 容 简 介

《数据结构实用教程(C语言版)》是为“数据结构”课程编写的教材。书中首先介绍了数据结构的概念及数据结构研究的逻辑结构、存储结构及运算三方面内容涉及的基本概念;然后针对经典的数据结构(即线性表、栈、队列、多维数组、广义表、树和图)的逻辑特征、常用的存储方式及各种基本运算的实现算法作了详细阐述;最后讨论了两类典型运算——排序和查找的各种实现方法。全书采用C语言作为数据结构和算法的描述工具。在一些重点部分,还给出了简单应用举例的完整C程序。

本书结构清晰,层次分明,深入浅出,通俗易懂,适用面广。可以作为普通高等院校计算机学科和信息类学科本科或专科教材,也可以作为其他理工类专业的选修教材。

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

数据结构实用教程: C语言版/赵波主编. —北京:清华大学出版社,2009.9

(21世纪普通高校计算机公共课程规划教材)

ISBN 978-7-302-20659-0

I. 数… II. 赵… 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(2009)第 144489 号

责任编辑:梁颖 林都嘉

责任校对:时翠兰

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:010-62776969, c-service@tup.tsinghua.edu.cn

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

印 装 者:北京鑫海金澳胶印有限公司

经 销:全国新华书店

开 本:185×260 印 张:17.75 字 数:439 千字

版 次:2009 年 9 月第 1 版 印 次:2009 年 9 月第 1 次印刷

印 数:1~4000

定 价:27.00 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:033744-01

相关课程教材推荐

ISBN	书 名	定价(元)
9787302130161	大学计算机网络公共基础教程	27.50
9787302197157	网络工程实践指导教程	33.00
9787302150503	网络存储导论	25.00
9787302155409	数据库技术——设计与应用实例	23.00
9787302183310	数据库原理与应用习题·实验·实训	18.00
9787302183983	数字图像处理实训教程	15.00
9787302193852	数字图像处理与图像通信	31.00
9787302154211	离散数学	25.00
9787302174585	汇编语言程序设计	21.00
9787302200765	汇编语言程序设计实验指导及习题解答	17.00
9787302201649	网页设计与开发——HTML、CSS、JavaScript 实例教程	29.00
9787302185635	网页设计与制作实例教程	28.00
9787302198796	Java 技术及其应用	27.00
9787302177104	C++ 语言程序设计教程	26.00
9787302180586	分布式对象技术及其应用	19.00
9787302200222	C++ 基础、上机指导及习题解答	25.00
9787302194422	Flash 8 动画基础案例教程	22.00
9787302183013	IT 行业英语	32.00
9787302176329	计算机专业英语：实用文案(档)写作	33.00
9787302191094	毕业设计(论文)指导手册(信息技术卷)	20.00
9787302174936	软件测试技术基础	19.80
9787302200628	信息检索与分析利用(第2版)	23.00

以上教材样书可以免费赠送给授课教师,如果需要,请发电子邮件与我们联系。

教学资源支持

敬爱的教师:

感谢您一直以来对清华版计算机教材的支持和爱护。为了配合本课程的教学需要,本教材配有配套的电子教案(素材),有需求的教师可以与我们联系,我们将向使用本教材进行教学的教师免费赠送电子教案(素材),希望有助于教学活动的开展。

相关信息请拨打电话 010-62776969 或发送电子邮件至 liangying@tup.tsinghua.edu.cn 咨询,也可以到清华大学出版社主页(<http://www.tup.com.cn> 或 <http://www.tup.tsinghua.edu.cn>)上查询和下载。

如果您在使用本教材的过程中遇到了什么问题,或者有相关教材出版计划,也请您发邮件或来信告诉我们,以便我们更好为您服务。

地址:北京市海淀区双清路学研大厦 A-708 计算机与信息分社 梁颖 收

邮编:100084

电子邮件: liangying@tup.tsinghua.edu.cn

电话:010-62770175-4505

邮购电话:010-62786544

出版说明

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程(简称‘质量工程’)”,通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

本系列教材立足于计算机公共课程领域,以公共基础课为主、专业基础课为辅,横向满足高校多层次教学的需要。在规划过程中体现了如下一些基本原则和特点。

(1) 面向多层次、多学科专业,强调计算机在各专业中的应用。教材内容坚持基本理论适度,反映各层次对基本理论和原理的需求,同时加强实践和应用环节。

(2) 反映教学需要,促进教学发展。教材要适应多样化的教学需要,正确把握教学内容和课程体系的改革方向,在选择教材内容和编写体系时注意体现素质教育、创新能力与实践能力的培养,为学生知识、能力、素质协调发展创造条件。

(3) 实施精品战略,突出重点,保证质量。规划教材把重点放在公共基础课和专业基础课的教材建设上;特别注意选择并安排一部分原来基础比较好的优秀教材或讲义修订再版,逐步形成精品教材;提倡并鼓励编写体现教学质量和教学改革成果的教材。

(4) 主张一纲多本,合理配套。基础课和专业基础课教材配套,同一门课程有针对不同层次、面向不同专业的多本具有各自内容特点的教材。处理好教材统一性与多样化,基本教材与辅助教材、教学参考书,文字教材与软件教材的关系,实现教材系列资源配套。

(5) 依靠专家,择优选用。在制定教材规划时要依靠各课程专家在调查研究本课程教

材建设现状的基础上提出规划选题。在落实主编人选时,要引入竞争机制,通过申报、评审确定主题。书稿完成后要认真实行审稿程序,确保出书质量。

繁荣教材出版事业,提高教材质量的关键是教师。建立一支高水平教材编写梯队才能保证教材的编写质量和建设力度,希望有志于教材建设的教师能够加入到我们的编写队伍中来。

21 世纪普通高校计算机公共课程规划教材编委会

联系人: 梁颖 liangying@tup.tsinghua.edu.cn

前 言

“数据结构”是计算机及相关专业的专业基础课和核心课。随着计算机应用范围逐渐深入到各个学科领域,在培养适应社会需求的多学科、复合型、应用型人才的过程中,本课程已经成为其他很多专业的热门选修课程。“数据结构”所研究的知识内容和技术方法,不论对学习计算机学科的其他相关课程,还是对从事软件设计和开发工作,都是重要的理论基础。

本书主要讨论数据处理问题中各种经典的逻辑结构及特点;数据在计算机中的存储结构及常用的存储方法;定义在逻辑结构上、实现在存储结构上的各种典型运算的算法。通过本书的学习,能够熟练掌握三大经典结构(线性表、树、图)的逻辑特征,能够采用常用的存储方法设计出合理的存储结构,并对典型运算设计多种实现算法。在深入理解和掌握本书内容的基础上,训练复杂程序设计的能力,并学会运用基本理论和基础知识解决实际问题。

教材中共包含 8 章内容:第 1 章绪论中主要介绍数据结构的概念及数据结构研究的三方面内容涉及的基本概念;第 2 章和第 3 章介绍了三种最基本的线性结构,即线性表、栈和队列;第 4 章至第 6 章叙述非线性结构,分别是多维数组、广义表、树和图;第 7 章和第 8 章讨论数据处理过程中使用频率最高的两种典型运算——排序和查找。鉴于目前“C 语言程序设计”已经普遍地成为数据结构的先修课,全书采用 C 语言作为数据结构和算法的描述工具。利用数组、结构体、指针等重要数据类型结合 C 函数,完成书中所有基本运算的实现算法。在一些重点部分,书中还给出了简单应用举例的完整 C 程序,旨在掌握如何利用数据结构中基本运算来解决实际问题。书中所有的算法都经过上机调试通过。

本书在内容选取上符合复合型、应用型人才培养目标的要求,遵循教学规律和认知规律。组织编排上体现先理论、后应用、理论与应用相结合的原则,注重课程内容的前后联系,理清来龙去脉,强调条理性 and 系统性,兼顾学科的广度和深度。本书结构清晰,层次分明,深入浅出,通俗易懂,适用面广。可以作为普通高等院校计算机学科和信息类学科本科和专科教材,也可以作为其他理工类专业的选修教材,讲授学时可以为 64~80 学时。教师可以根据本校的教学大纲及学时安排,选讲部分内容。

本书的主编一直从事数据结构的教学和研究工作,参加编著过多本教材。本书是作者多年教学经验的结晶,在难点内容的叙述及讲解方法上有独到之处。主编完成全书的整体策划,并承担统稿工作,也参与了部分章节的编写。其他作者分工如下:第 1 章、第 2 章、第 3 章由郑巍编写;第 4 章、第 5 章由董靓瑜编写;第 6 章由李静编写;第 7 章、第 8 章由霍利编写。编写过程中参考了大量的著作、教材等资料,在此一并表示感谢。

虽然全体参编人员都尽心尽力、力求完美,但由于时间仓促、水平有限,书中难免出现遗漏或不妥之处,敬请广大读者不吝指正,不胜感激。

编 者

2009 年 5 月

目 录

第 1 章 绪论	1
1.1 基本术语	1
1.2 数据结构的定义及研究的内容	2
1.2.1 数据的逻辑结构	2
1.2.2 数据的存储结构	5
1.2.3 数据的运算	7
1.3 算法	8
1.3.1 算法的概念及特性	8
1.3.2 算法的描述	8
1.3.3 算法的评价	13
1.4 学习数据结构的意义和目的	17
习题	18
第 2 章 线性表	19
2.1 线性表的定义及运算	19
2.1.1 线性表的定义及逻辑特征	19
2.1.2 线性表上运算的定义	20
2.1.3 线性表的存储结构	21
2.2 顺序表	21
2.2.1 顺序表的定义及表示	21
2.2.2 线性表运算在顺序表上的实现	23
2.2.3 顺序表应用举例	27
2.3 链表	34
2.3.1 链表的定义及形式	34
2.3.2 单链表	35
2.3.3 循环链表	49
2.3.4 双链表	50
*2.3.5 静态链表	52
2.3.6 单链表的应用举例	54
2.4 顺序表和链表的比较	66

习题	67
第3章 栈和队列	68
3.1 栈.....	68
3.1.1 栈的定义及运算	68
3.1.2 顺序栈及运算的实现	69
3.1.3 链栈及运算的实现	71
3.1.4 栈的应用	72
3.1.5 栈与递归	75
3.2 队列.....	79
3.2.1 队列的定义及运算	79
3.2.2 顺序队列及运算的实现	80
3.2.3 链队列及运算的实现	83
3.3 栈与队列的比较.....	86
习题	87
第4章 多维数组及广义表	88
4.1 多维数组.....	88
4.2 矩阵的压缩存储.....	90
4.2.1 特殊矩阵	90
4.2.2 稀疏矩阵	94
4.3 广义表.....	99
4.3.1 广义表的定义	99
4.3.2 广义表的运算.....	100
习题.....	101
第5章 树.....	102
5.1 树的定义	102
5.2 二叉树	104
5.2.1 二叉树的定义及性质.....	105
5.2.2 二叉树的存储.....	108
5.2.3 二叉树的遍历及实现算法.....	112
5.3 线索二叉树	119
5.3.1 中序线索二叉树的定义.....	120
5.3.2 中序线索二叉树上遍历的实现.....	122
5.3.3 利用中序线索实现前序遍历和后序遍历.....	123
5.4 树和森林	127
5.4.1 树和森林的遍历.....	128
5.4.2 森林与二叉树的转换.....	128

* 5.4.3 树的存储	130
5.5 哈夫曼树	132
5.5.1 哈夫曼树的定义及建立	132
5.5.2 哈夫曼编码及译码	137
5.5.3 哈夫曼树应用举例	141
* 5.6 树与等价类问题	144
习题	147
第 6 章 图	149
6.1 图的概念	149
6.2 图的存储	153
6.2.1 邻接矩阵	154
6.2.2 邻接表	156
6.2.3 边集数组	159
6.3 图的遍历	160
6.3.1 深度优先搜索遍历	161
6.3.2 广度优先搜索遍历	164
6.3.3 非连通图的遍历	167
6.4 最小生成树	176
6.4.1 普里姆算法	177
6.4.2 克鲁斯卡尔算法	181
6.5 最短路径	183
6.5.1 单源最短路径	184
6.5.2 任意两点间最短路径	190
6.6 拓扑排序	192
* 6.7 关键路径	196
习题	201
第 7 章 排序	203
7.1 排序的基本概念	203
7.2 插入排序	205
7.2.1 直接插入排序	205
7.2.2 希尔排序	207
7.3 交换排序	209
7.3.1 起泡排序	209
7.3.2 快速排序	211
7.4 选择排序	214
7.4.1 直接选择排序	214
7.4.2 堆排序	215

7.5 归并排序	220
*7.6 基数排序	223
7.7 内排序方法的比较	225
习题	226
第8章 查找	228
8.1 查找的基本概念	228
8.2 顺序表查找	229
8.2.1 顺序查找	229
8.2.2 二分查找	230
8.3 索引查找	234
8.3.1 索引表的组织	234
8.3.2 分块查找	236
8.4 树表查找	238
8.4.1 二叉排序树	238
*8.4.2 平衡二叉排序树	244
*8.4.3 B-树	251
8.5 散列表查找	256
8.5.1 散列表的概念	256
8.5.2 散列函数的设计	258
8.5.3 解决冲突的方法	259
8.5.4 散列表的应用举例	263
习题	269
参考文献	271

随着信息时代的到来,计算机软硬件技术得到了飞速的发展,计算机的应用领域也随之进一步拓展,从最初的主要是为了科学计算,发展到现在已经深入人类活动的各个领域。计算机科学是一门研究数据表示和数据处理的科学。目前,计算机处理的对象除了简单的数值数据,还包括声音、图形、图像、色彩等复杂的多媒体数据。此外,数据的处理量日益庞大。因此,对于结构复杂的海量数据,要开发出一套性能良好的、处理高效的软件系统,不仅要根据掌握一种计算机程序设计的高级语言和软件开发工具,更重要的是要研究数据的不同结构和组织方法以及在计算机中的存储表示,在此基础上选择数据处理的不同算法,通过分析和比较选择出较好的设计方案,这正是数据结构这门学科所要解决的问题。

在本章中主要介绍数据结构这门课程中用到的基本概念和基本方法,而所有给出的基本概念和基本方法都是围绕数据结构这个最重要的概念衍化出来的,其中很多概念和方法会贯穿整个课程始终。

1.1 基本术语

在本节中将对本书中使用的一些常用概念和术语赋以确定的含义。

数据(Data)是人们约定的符号,用它来表示客观事物及其活动,是信息的载体。数据是计算机程序加工处理的对象。在计算机领域中是指所有能被计算机识别、存储、处理、传输的符号的集合。数据是个广义的概念,可以指普通的数值、字符等,也可以指声音、图形、图像及色彩等,随着计算机软硬件的发展及应用领域的扩大,数据的范畴还将会拓广。例如,编译程序或文本程序的处理对象是字符串,音频软件处理的对象是声音。

数据元素(Data Element)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,在不同的情况下,又可以称为元素、结点、顶点或记录。数据是由数据元素构成的。例如,对于一个字符串数据来说,每个字符是它的数据元素;对于学生基本信息数据来说,每一个学生信息是它的数据元素;对于一个数组来说,每一个下标变量是它的数据元素;对于一个文件来说,每一个记录是它的数据元素。数据和数据元素是相对而言的,例如记录是文件数据中的一个数据元素,而相对它所含的数据项而言又可以看作数据。

数据项(Data Item)是构成数据元素不可分割的具有独立含义的最小标识单位。若数据元素可再分,则数据元素是由若干个数据项组成,如学生的信息由数据项姓名、学号、班级组成;如数据元素不可再分,数据元素和数据项是同一概念,如整型数据就是不可再分的。

数据类型(Data Type)是一个值的集合和定义在这个值集上一组操作的总称。如 C 语言中整型数据类型,其值集为某个区间上的整数(区间大小依赖于所使用的计算机及语言),定义在其上的操作为加、减、乘、除和取模等。现实世界中的数据是多种多样的,不同的用途

需要不同种类的数据,而不同种类的数据对应不同种类的使用方法。数据类型明显或隐含地规定了在程序执行期间变量或表达式所有可能的取值范围,以及在这些值集上允许的操作。

按值的不同特性,高级程序设计语言中的数据类型可分为原子类型和结构类型两类。原子类型的值不可再分,一般都是高级语言提供的基本数据类型,如C语言的整型、实型、字符型、枚举型、指针类型和空类型。结构类型的值是由若干成分按某种结构组成,是可分解的,它的成分可以是原子的,也可以是结构的,如C语言中的数组、结构体等。

1.2 数据结构的定义及研究的内容

有了上面的基本概念,就可以给出数据结构的定义了。

世界是由客观事物(实体)构成的,而客观事物之间都不是孤立存在的,相互之间存在着一定的联系。在数据结构中,客观事物用数据元素来表示,那么数据元素之间就存在着一定的关系。把数据元素之间的关系称为**结构**,则数据结构可以看成是数据及数据之间的关系,但这只是从数学的角度认识数据结构。在本课程中,重点要讨论用计算机来处理数据,因此只研究数据及数据之间的关系是不够的,还要研究数据如何保存到计算机中并进行相应的处理,综上所述,可以将数据结构定义如下:

按照某种逻辑关系组织起来的一批数据,用一定的存储方式存储在计算机的存储器中,并在这些数据上定义一个运算的集合,就称为一个**数据结构(Data Structure)**。

上述定义中包含的三个方面的内容正是数据结构重点研究的内容。

- (1) 数据的逻辑结构:即数据之间的逻辑关系。
- (2) 数据的存储结构:即数据及数据之间的关系在计算机存储器中的表示。
- (3) 数据的运算:即对数据施加的各种操作。

数据的逻辑结构是从逻辑关系上描述数据,与计算机无关,因此,数据的逻辑结构可以看成是从具体问题抽象出来的数学模型;数据的存储结构是逻辑结构到计算机存储器的映射,是物理结构,它是依赖于计算机存储器的;数据的运算是定义在数据的逻辑结构上,每种逻辑结构都有一个运算的集合,常用的有查找、插入、删除、修改等,但这些运算只是在抽象的数据上所实施的一系列抽象的操作,若用计算机来实现这些运算,必须将数据及其关系按一定的原则存放到计算机的存储器中,也就是说,数据的运算是实现在存储结构上的;这就是数据结构研究的三方面内容之间的联系。

下面就从数据结构研究的三方面内容来给出常用的术语和概念。

1.2.1 数据的逻辑结构

数据的逻辑结构(Logical Structure)指的是数据元素之间的逻辑关系。它是人们根据实际问题的需要和问题本身所含数据之间的内在联系而抽象出来的数学模型,与如何利用计算机存储和处理无关,所以被称为数据的逻辑结构。

由于数据的逻辑结构是数学模型,可以借助数学方法来表示,具体的可以用离散数学中关系代数的二元组表示:

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

其中： D 是数据元素的有限集， D 为空集或表示为 $D = \{d_1, d_2, \dots, d_m\}$ ， S 是 D 上关系的有限集， S 为空集或表示为 $S = \{r_1, r_2, \dots, r_n\}$ ，若 S 为空集，表示 D 中数据元素之间不存在关系，或者说是一种离散的关系，就像数学中集合里的元素一样，无关系的数据不作为本课程讨论的内容。当讨论某一个数据结构时，通常取 S 中的一个关系 r_j 来进行讨论， r_j 可以表示为数据元素的序偶 $\langle d_i, d_j \rangle$ 的集合。如果集合中有序偶 $\langle d_i, d_j \rangle$ ，表示数据元素 d_i 和 d_j 之间有 r_j 这种关系。

针对用二元组表示的数据的逻辑结构，有如下的常用术语：

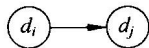
(1) 前趋结点、后继结点、相邻结点

设 $\text{Data_Structure} = (D, S)$ 是一个逻辑结构， $d_i, d_j \in D, r_j \in S$ ，若 $\langle d_i, d_j \rangle \in r_j$ ，则称 d_i 是 d_j 相对 r_j 的**前趋结点**， d_j 是 d_i 相对 r_j 的**后继结点**， d_i 和 d_j 称为**相邻结点**。

(2) 开始结点、终端结点、内部结点

设 $\text{Data_Structure} = (D, S)$ 是一个逻辑结构， $d_i \in D, r_j \in S$ ，若 d_i 相对 r_j 无前趋结点，则称 d_i 是相对 r_j 的**开始结点**；若 d_i 相对 r_j 无后继结点，则称 d_i 是相对 r_j 的**终端结点**；若 d_i 相对 r_j 既有前趋结点又有后继结点，则称 d_i 是相对 r_j 的**内部结点**。

数据的逻辑结构还能够利用更形象的图形表示。具体表示法为小圆圈代表数据元素（有时也叫结点或顶点），小圆圈间的带箭头连线表示数据元素间的关系（序偶），如序偶 $\langle d_i, d_j \rangle$ 可以表示为：



由于数据的逻辑结构是数据结构研究的三方面内容的前提部分，也就是说，没有数据的逻辑结构就不可能有数据的存储结构，也不可能有数据的运算。因此在不加严格区分的情况下，常常将数据的逻辑结构简称为数据结构。数据的逻辑结构有以下两大类。

(1) 线性结构：经典的线性结构是线性表。

线性结构的逻辑特征是：有且仅有一个开始结点和一个终端结点，其余的内部结点都有且仅有一个前趋结点和一个后继结点，也就是说结构中的数据元素间存在着一对一的相互关系。本书的第2章和第3章介绍的都是线性结构。

(2) 非线性结构：经典的非线性结构有树形结构和图形结构。

树形结构的逻辑特征是：有且仅有一个开始结点，可有若干个终端结点，其余的内部结点都有且仅有一个前趋结点，可以有若干个后继结点，也就是说结构中的数据元素间存在着一对多的层次关系。本书的第5章将介绍树形结构。

图形结构的逻辑特征是：可有若干个开始结点和终端结点，其余的内部结点可以有若干个前趋结点和若干个后继结点，也就是说结构中的数据元素间存在着多对多的网状关系。本书的第6章将介绍图形结构。

作为例子，下面根据表1.1某校围棋社团学生简表构造一些典型的数据结构。其中每个数据元素有5个数据项，学号是唯一的，为了简化，用学号代表某个学生对应的数据元素。

例1.1 在表1.1中，8名学生按学号从小到大排列，形成一个线性结构。假设表示这种逻辑结构的关系为 r_1 ，则 r_1 可以定义为学生按学号顺序递增排列的关系，该线性结构的逻辑结构可用二元组表示为：

$$L = (D, S), r_1 \in S$$

$$D = \{01,02,03,04,05,06,07,08\}$$
$$r_1 = \{ \langle 01,02 \rangle, \langle 02,03 \rangle, \langle 03,04 \rangle, \langle 04,05 \rangle, \\ \langle 05,06 \rangle, \langle 06,07 \rangle, \langle 07,08 \rangle \}$$

表 1.1 某校园棋社团学生简表

学号	姓 名	性别	出生日期	职务
01	黄家正	男	1982-08-05	团长
02	赵 芳	女	1981-08-15	组长
03	王 明	女	1983-04-01	组长
04	王 红	女	1982-06-28	组员
05	张小才	男	1984-03-17	组员
06	马立伟	男	1983-10-12	组员
07	孙 刚	男	1982-07-05	组员
08	刘 永	男	1982-12-09	组员

对应该线性结构的逻辑结构图示如图 1.1 所示。从图中可以明显地看出此数据结构为线性结构。其中 01 是开始结点,没有前趋结点,08 是终端结点,没有后继结点,其余的结点有唯一的前趋结点和后继结点,数据元素间是一对一的关系,符合线性结构的逻辑特征。

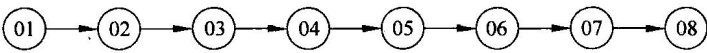


图 1.1 线性结构的图示

例 1.2 在表 1.1 中,学生之间还存在着领导与被领导关系,其中 01 号学生为团长,直接领导 02 和 03 号学生,他们分别是组长,02 号学生直接领导 04 和 05 号学生,03 号学生直接领导 06,07 和 08 号学生,假设表示这种逻辑结构的关系为 r_2 ,则 r_2 可以定义为学生之间的领导与被领导关系,该数据结构的逻辑结构可用二元组表示为:

$$T = (D,S), r_2 \in S$$
$$D = \{01,02,03,04,05,06,07,08\}$$
$$r_2 = \{ \langle 01,02 \rangle, \langle 01,03 \rangle, \langle 02,04 \rangle, \langle 02,05 \rangle, \\ \langle 03,06 \rangle, \langle 03,07 \rangle, \langle 03,08 \rangle \}$$

对应的逻辑结构图形表示如图 1.2 所示。可以看出,此数据结构为树形结构,与自然界的树相比,它是一棵倒立的树。其中 01 是开始结点,没有前趋结点,只有后继结点; 04,05, 06,07,08 是终端结点,没有后继结点,其余的结点都只有一个前趋结点但有若干个后继结点,数据元素间是一对多的关系,符合树形结构的逻辑特征。

例 1.3 在表 1.1 中,学生之间还有好友关系,如 01 和 02,03,05 号是好友,02 和 04 号是好友,03 和 05 号是好友,04 和 05,06 号是好友,06 和 07 之间是好友,08 无好友,假设表示这种逻辑结构的关系为 r_3 ,则 r_3 可以定义为学生之间的好友关系,该数据结构的逻辑结构可用二元组表示为:

$$G = (D,S), r_3 \in S$$
$$D = \{01,02,03,04,05,06,07,08\}$$
$$r_3 = \{ \langle 01,02 \rangle, \langle 02,01 \rangle, \langle 01,03 \rangle, \langle 03,01 \rangle, \langle 01,05 \rangle, \\ \langle 02,04 \rangle, \langle 04,02 \rangle, \langle 03,05 \rangle, \langle 05,03 \rangle, \langle 04,05 \rangle, \langle 05,04 \rangle, \langle 06,07 \rangle, \langle 07,06 \rangle \}$$

$\langle 05,01 \rangle, \langle 02,04 \rangle, \langle 04,02 \rangle, \langle 03,05 \rangle, \langle 05,03 \rangle, \langle 04,05 \rangle,$
 $\langle 05,04 \rangle, \langle 04,06 \rangle, \langle 06,04 \rangle, \langle 06,07 \rangle, \langle 07,06 \rangle\}$

对应的逻辑结构图形表示如图 1.3 所示,其中 08 号学生无好友,在图形中为孤立顶点;图形中每对顶点之间的两条相反的有向边,表示这两个顶点学生是一对好朋友。从图 1.3 可看出,数据元素间存在着多对多的关系,也就是说一个数据元素可以有任意多个前趋结点和任意多个后继结点,符合图形结构的逻辑特征。

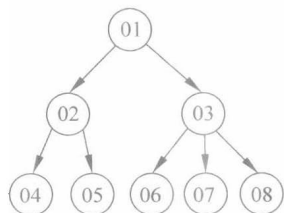


图 1.2 树形结构的图示

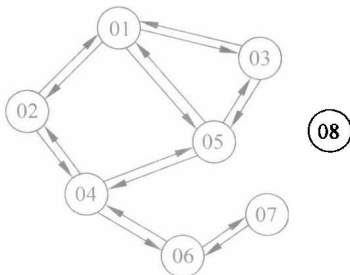


图 1.3 对应的逻辑结构图

从以上三个例子可以看出,虽然各例中数据元素是相同的,但是数据元素间的关系不同,所以它们是不同的数据结构。另外,上述三种结构也是数据结构中要重点研究的三种经典结构。

1.2.2 数据的存储结构

数据的存储结构(Storage Structure),是指数据的逻辑结构到计算机存储器的映射。对于数据的逻辑结构 $Data_Structure=(D,S)$,在映射中,一方面要将数据集 D 中的数据元素存放存储器中,另一方面还要体现关系集 S ,常见的体现关系 S 的方式有显示和隐含两种。

计算机存储空间是以字节为单位进行编址的,地址是从零开始的、连续的正整数。对计算机存储器存取操作的基本单位是字节(byte),每个字节都有唯一的地址标识。用户使用存储器通常是以存储单元为单位,每个存储单元由若干个连续的字节构成,一个单元可以存储一个数据元素,单元的大小取决于数据元素的类型。每个存储单元都有唯一的地址标识(即若干个连续字节的首地址),用户可以通过每个存储单元的地址实现对存储单元中数据元素的访问。数据可以存储在连续的存储单元中也可以存储在不连续的存储单元中。若干地址连续的存储单元称为一个存储区域,也可以说存储区域是存储单元的线性序列。

完成数据的逻辑结构到存储区域的映射可以有很多的方法,最常用的实现数据存储结构的方法有如下 4 种。

1. 顺序存储

顺序存储方法的基本思想是:把逻辑上相邻的数据元素存储在物理位置上相邻的存储单元里。数据元素间的逻辑关系由存储单元的邻接关系来体现,也就是说逻辑关系上相邻物理位置上也相邻,数据元素的逻辑次序与物理次序一致。这是一种隐含体现关系的存储方法,关系隐含在存储位置上。后继结点存储在前趋结点所在单元的下一个单元中。由此可以看出数据元素在存储区域中是连续存放的,这种存储方法称为顺序存储结构(Sequential Storage Structure),通常用计算机高级语言中的数组来描述。这种方法主要应

用于线性结构,但是并不局限于此,非线性的数据结构也可以通过某种线性化方法实现顺序存储,如多维数组、二叉树等结构的存储。

2. 链接存储

链接存储方法的基本思想是:通过附加指针域表示数据元素之间的关系。这种存储方法不要求逻辑上相邻的数据元素存储位置上也相邻,数据元素间的逻辑关系是通过附加指示其他数据元素位置的地址信息(指针)而得到的,这是一种显示体现关系的存储方法。数据元素在存储区域中可以是连续的,也可以是不连续的,通常用计算机高级语言中的指针来描述,称为**链接存储结构**(Linked Storage Structure)。这种方法具有很大的灵活性,适用于大多数的数据结构。但是与顺序存储方法相比,由于要存储其他数据元素的位置信息,所以它的空间开销要大一些。但由于不要求存储空间的连续性,很适合动态存储管理。

例 1.4 用上述两种方法存储有序序列 $A=(99,123,134)$,假设每个数据元素占 2 个字节,即一个存储单元为两个字节。

图 1.4(a)表示顺序存储结构,其中序列中的 3 个元素依次保存在从地址 0200 起的 3 个单元中,逻辑上相邻的数据元素,在存储位置上也相邻的。例如:针对结点 123,它的前趋结点 99 就存储在前一个单元中,地址为 $0202-2$;它的后继结点 134 就存储在后一个单元中,地址为 $0202+2$ 。由此可以看出,不需要附加其他信息就可以表示结点之间的关系,即由存储位置可以隐含关系。

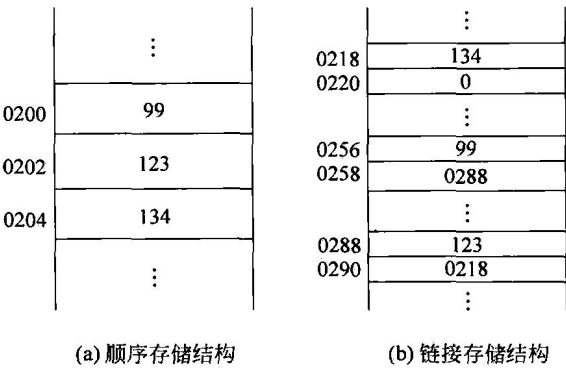


图 1.4 关系的映像方法

图 1.4(b)表示链接存储结构,数据元素 99 保存在地址为 0256 的单元中,它的后继结点 123 没有保存在相邻的下个单元中,而是保存在地址为 0288 的单元中,如何找到保存数据元素 123 的位置呢?这个地址是保存在地址 0258 的单元中,也就是在存储数据元素 99 的同时,也保存了它的后继结点 123 的存储位置(指针)0288,把这个域叫指针域。同理,在地址 0290 中可以找到第三个元素 134 的地址为 0218,元素 134 不再有后继结点,它的指针域为空。这样,从第一个元素开始,顺着指针域依次可以找到各个元素,就好像形成了一条链。这两种方法各有优缺点,后面还将继续讨论。

3. 索引存储

索引存储方法的基本思想是:除了存储数据元素,还要建立一个或若干个附加的索引表来标识数据元素的地址。索引表中的每一项称为索引项,是用来标识一个或一组数据元素的存储位置。索引项一般形式为(关键字,地址),其中的关键字是用来标识数据元素的数