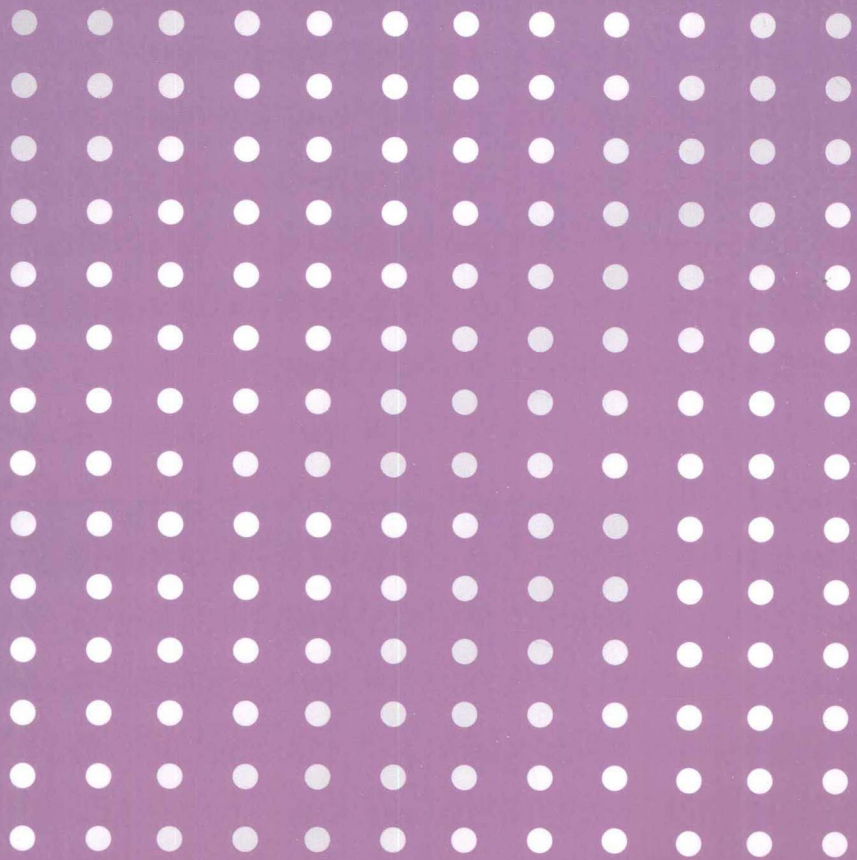


高等院校信息技术规划教材

数据结构 (C语言版)

秦锋 主编

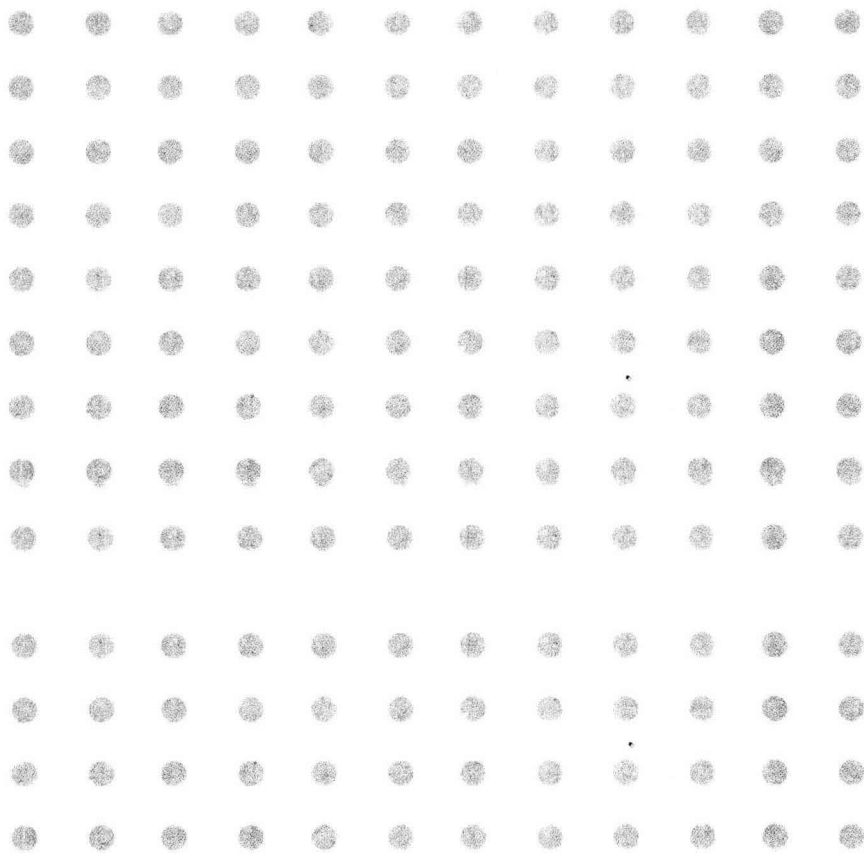


清华大学出版社

高等院校信息技术规划教材

数据结构 (C语言版)

秦 锋 主编
汤文兵 章曙光 汪 军 副主编
汤亚玲 陈学进 袁志祥



清华大学出版社
北京

内 容 简 介

本书全面系统地介绍了线性表、队列、堆栈、树、图等基本数据结构,以及这些数据结构在计算机中的存储及算法实现,系统地介绍了各种查找及排序算法的实现和效率分析,最后一章给出了数据结构综合应用实例。书中各种算法采用C语言描述,注重程序设计风格。

本书有配套教材《数据结构(C语言版)例题详解与课程设计指导》(ISBN: 9787302246282),书中包含各知识点的归纳与总结,也包含例题详解、习题解答以及课程设计指导。

有关教学参考资料的电子文档可通过 <http://jpkc.ahut.edu.cn/sjgg/> 下载。

本书语言流畅,内容通俗易懂,算法描述力求简练、易读。可作为计算机类及信息类专业教材,也可供广大计算机爱好者及软件开发人员自学提高时使用。

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(C语言版)/秦锋主编;汤文兵等副主编. —北京:清华大学出版社,2011.3
(高等院校信息技术规划教材)

ISBN 978-7-302-24603-9

I. ①数… II. ①秦… ②汤… 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(2011)第012439号

责任编辑:袁勤勇

责任校对:时翠兰

责任印制:孟凡玉

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:010-62795954, jsjic@tup.tsinghua.edu.cn

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

印 刷 者:三河市君旺印装厂

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:185×260

印 张:20

字 数:464千字

版 次:2011年3月第1版

印 次:2011年3月第1次印刷

印 数:1~3000

定 价:28.00元

前言

foreword

以计算机科学技术为核心的信息技术正在深刻地改变着人们的工作、生活和思维方式。软件是计算机的灵魂,程序设计是计算机科学技术最重要的基础,写出高质量的程序是每个软件开发者追求的目标。要达到这个目标仅靠学习几门高级语言是远远不够的,正如我们即使认识汉字并懂得中文语法也难以写出好文章一样。数据结构这门课程正是打开程序设计知识宝库的金钥匙,其主要目的是培养学生将现实世界抽象为数据和数据模型的能力以及利用计算机进行数据存储和数据加工的能力。学好数据结构,可以使读者掌握更多的程序设计技巧,为以后学习计算机专业课程及走上工作岗位从事计算机大型软件开发打下良好的基础。目前,数据结构是我国高校计算机类专业(包含计算机科学与技术、软件工程、网络工程、信息安全等)的核心课程之一,也是其他信息类专业(包含信息管理与信息系统、通信工程、信息与计算科学等)的必修课程之一。

本书内容共 10 章,第 1 章重点介绍数据结构与算法的基本概念,介绍评价算法优劣的主要指标以及衡量算法效率的时间复杂度和空间复杂度;第 2 章~第 4 章重点介绍线性表、栈、队列和字符串等线性结构的逻辑特性、存储结构,以及基本操作算法;第 5 章~第 7 章重点介绍多维数组、广义表、树、二叉树、图等非线性结构的逻辑特征、存储表示及基本操作算法的实现和具体应用;第 8 章和第 9 章介绍在软件开发中广泛运用的两种操作——排序和查找,对一些常用的查找、排序算法进行详细描述及效率分析;第 10 章将线性结构和树、图等非线性结构归纳在一起,运用树、图的遍历算法设计思想去解决实际问题,让读者理论联系实际,以加深对各种数据结构的理解。

书中算法采用 C 语言描述,每段代码都保持良好的设计风格,配有丰富的例题和习题。

数据结构是一门实践性很强的课程,读者在进行理论学习的同时,需要多动手编写程序上机调试,以加深对所学知识的理解,提高编程能力。

本书可作为高等院校计算机类或信息类相关专业数据结构课程教材,建议理论课时为 50~70 学时,上机及课程设计等实践课时为 20~30 学时。各院校可根据本校的专业特点和具体情况适当增删教学内容。

本书有配套教材《数据结构(C语言版)例题详解与课程设计指导》(ISBN: 9787302246282),书中有各知识点的归纳与总结,有例题详解及习题解答以及课程设计指导。

本书由秦锋教授任主编,由汤文兵、章曙光、汪军、汤亚玲、陈学进、袁志祥任副三编。

第 1 章和第 6 章由秦锋教授编写;第 3 章和第 5 章由汤亚玲编写;第 2 章由汪军编写;第 4 章由陈学进编写;第 8 章由章曙光编写;第 7 章由汤文兵编写;第 9 章由袁志祥编写;第 10 章由杨学兵编写,全书由秦锋教授负责修改并统稿,郑啸、程泽凯、储岳中、王森玉、秦飞、徐浩参加了部分编写和修改工作。因编者水平有限,书中难免有不足甚至错误之处,敬请广大读者批评指正。

为方便教学,读者可到 <http://jpkc.ahut.edu.cn/sjjg/> 下载或阅读相关教学参考资料。

建议或者意见请发电子邮件至 fqin@ahut.edu.cn。

编 者

2011 年 1 月

目录

Contents

第 1 章 绪论	1
1.1 什么是数据结构	1
1.1.1 数据结构的定义	1
1.1.2 学习数据结构的意义	4
1.2 基本概念和术语	5
1.2.1 数据与数据元素	5
1.2.2 数据的逻辑结构与存储结构	5
1.2.3 数据运算	6
1.2.4 数据类型与抽象数据类型	7
1.3 算法和算法描述语言	8
1.4 算法分析	9
1.4.1 算法评价	9
1.4.2 算法性能分析与度量	13
本章小结	18
习题	18
第 2 章 线性表	22
2.1 线性表的逻辑结构	22
2.1.1 线性表的定义	22
2.1.2 线性表的基本操作	23
2.2 线性表的顺序存储及运算实现	24
2.2.1 顺序表	24
2.2.2 顺序表上基本运算的实现	26
2.3 顺序表应用举例	30
2.4 线性表的链式存储和运算实现	31
2.4.1 单链表	32
2.4.2 单链表基本运算的实现	33

2.4.3	循环链表	38
2.4.4	双向链表	39
2.4.5	静态链表	40
2.4.6	单链表应用举例	41
2.5	顺序表和链表的比较	46
	本章小结	47
	习题	48
第3章	栈和队列	52
3.1	栈	52
3.1.1	栈的定义及基本操作	52
3.1.2	栈的顺序存储及操作实现	53
3.1.3	栈的链式存储及操作实现	56
3.2	栈的应用举例	59
3.3	递归	69
3.3.1	递归定义	69
3.3.2	递归和栈的关系	70
3.3.3	递归算法实例	71
3.4	队列	74
3.4.1	队列的定义及基本操作	74
3.4.2	队列的顺序存储实现及操作实现	75
3.4.3	队列的链式存储实现及操作实现	79
3.5	队列应用举例	82
	本章小结	84
	习题	84
第4章	串	88
4.1	串及其基本运算	88
4.1.1	串的基本概念	88
4.1.2	串的基本运算	89
4.2	串的顺序存储及基本运算	91
4.2.1	串的定长顺序存储	91
4.2.2	定长顺序串的基本运算	92
4.3	模式匹配	93
4.3.1	简单的模式匹配算法	94
4.3.2	KMP 算法	96
4.4	串的堆存储结构	100

4.4.1 动态堆存储	100
4.4.2 静态堆存储	103
4.5 串的链式存储结构	106
4.6 文本编辑——串操作应用	108
本章小结	109
习题	110
第5章 数组和广义表	113
5.1 数组	113
5.1.1 数组的定义	113
5.1.2 数组的内存映像	114
5.2 特殊矩阵的压缩存储	114
5.2.1 对称矩阵	115
5.2.2 三角矩阵	115
5.2.3 稀疏矩阵	115
5.3 广义表	121
5.3.1 广义表的定义	121
5.3.2 广义表的存储	122
5.3.3 广义表基本操作的实现	123
本章小结	127
习题	127
第6章 树和二叉树	131
6.1 树的基本概念	131
6.1.1 树的定义及其表示	131
6.1.2 基本术语	133
6.2 二叉树	134
6.2.1 二叉树的定义	134
6.2.2 二叉树的性质	134
6.2.3 二叉树的存储结构	136
6.3 遍历二叉树	138
6.3.1 先序遍历	138
6.3.2 中序遍历	140
6.3.3 后序遍历	141
6.3.4 按层次遍历二叉树	144
6.3.5 遍历算法的应用举例	145
6.4 线索二叉树	147

6.4.1	线索的概念	147
6.4.2	线索的算法实现	149
6.4.3	线索二叉树上的运算	151
6.5	树与森林	154
6.5.1	树的存储结构	154
6.5.2	树、森林和二叉树的转换	156
6.5.3	树和森林的遍历	158
6.6	哈夫曼树	159
6.6.1	基本术语	160
6.6.2	哈夫曼树的建立	161
	本章小结	166
	习题	167
第7章	图	171
7.1	图的基本概念	171
7.1.1	图的定义和术语	171
7.1.2	图的基本操作	175
7.2	图的存储结构	176
7.2.1	邻接矩阵	176
7.2.2	邻接表	178
7.2.3	十字链表	180
7.2.4	邻接多重表	181
7.3	图的遍历	183
7.3.1	深度优先搜索	183
7.3.2	广度优先搜索	184
7.3.3	应用图的遍历判定图的连通性	186
7.3.4	图的遍历的其他应用	187
7.4	最小生成树	198
7.4.1	生成树及生成森林	198
7.4.2	最小生成树的概念	198
7.4.3	构造最小生成树的 Prim 算法	199
7.4.4	构造最小生成树的 Kruskal 算法	201
7.5	最短路径	203
7.5.1	从一个源点到其他各点的最短路径	204
7.5.2	每一对顶点之间的最短路径	207
7.6	有向无环图及其应用	209
7.6.1	有向无环图的概念	209
7.6.2	AOV 网与拓扑排序	210

7.6.3	AOE 图与关键路径	215
本章小结	219
习题	220
第 8 章	查找	224
8.1	基本概念	224
8.2	线性表的查找	226
8.2.1	顺序查找	226
8.2.2	有序表的查找	227
8.2.3	分块查找	229
8.3	树表查找	230
8.3.1	二叉排序树	230
8.3.2	平衡二叉树(AVL 树)	235
8.3.3	B-树和 B ⁺ 树	242
8.4	哈希表查找(杂凑法)	246
8.4.1	哈希表与哈希方法	246
8.4.2	常用的哈希方法	248
8.4.3	处理冲突的方法	249
8.4.4	哈希表的操作	252
8.4.5	哈希表查找及其分析	253
本章小结	254
习题	255
第 9 章	排序	259
9.1	基本概念	259
9.2	插入排序	261
9.2.1	直接插入排序	262
9.2.2	折半插入排序	263
9.2.3	希尔排序	265
9.3	交换排序	266
9.3.1	冒泡排序	267
9.3.2	快速排序	268
9.4	选择排序	272
9.4.1	简单选择排序	272
9.4.2	堆排序	273
9.5	归并排序	276
9.6	基数排序	278



9.6.1 多关键码排序	278
9.6.2 链式基数排序	279
本章小结	283
习题	283
第 10 章 数据结构综合应用	287
10.1 各种结构类型之间的关系概述	287
10.2 二叉树与分治策略	290
10.3 图的遍历及其应用	295
本章小结	305
习题	306
参考文献	307

绪 论

数据结构是计算机、信息管理、信息与计算科学等信息类专业最重要的专业基础课程,掌握好数据结构的知识将直接关系到后续专业课程的学习。数据结构主要研究4方面的问题:(1)数据的逻辑结构,即数据之间的逻辑关系;(2)数据的物理结构,即数据在计算机内的存储方式;(3)对数据的加工,即基于某种存储方式的操作算法;(4)算法的分析,即评价算法的优劣。本章重点介绍数据结构研究问题所涉及的基本知识和概念。

【本章学习要求】

了解:研究数据结构的目的是意义。

了解:数据结构基本概念和相关术语。

掌握:算法基本概念和算法评价依据。

掌握:算法的时间复杂度和空间复杂度。

1.1 什么是数据结构

程序设计是计算机学科各个领域的基础。在计算机发展的早期,程序设计所处理的数据都是整型、实型等简单数据,绝大多数的应用软件都是用于数值计算。随着信息技术的发展,计算机逐渐进入到金融、商业、管理、通信以及制造业等各个行业,广泛地应用于数据处理和过程控制,计算机加工处理的对象也由纯粹的数值型数据发展到字符、表格和图像等各种具有一定结构的数据,这就给程序设计带来一些新的问题,数据结构的观念就是在这种背景下产生的。

1.1.1 数据结构的定义

计算机解决一个具体问题一般需要经过下列几个步骤:首先要从具体问题抽象出一个适当的数学模型,然后设计或选择一个解此数学模型的算法,最后编写出程序进行调试、测试,直至得到最终的解答。显然如何抽象出数学模型并设计解此数学模型的算法就成为问题的关键,由于早期计算机所涉及的运算对象是简单的整型、实型或布尔类型数据,数学模型或者比较简单或者就是某个数学公式,程序设计者的主要精力可以集中

于程序编码的设计上,无须重视数据的逻辑关系和存储结构。随着计算机应用领域的扩大和软、硬件技术的发展,非数值计算问题显得越来越重要。据统计,当今用计算机处理的非数值计算性问题约占 90% 以上的计算机时间。如图书资料的检索、职工档案管理、博弈游戏等,这类问题涉及的数据结构相当复杂,数据元素之间的相互关系无法用数学方程或数学公式来描述。这类问题的处理对象中的各分量不再是单纯的数值型数据,更多的是字符、字符串或其他编码表示的信息。对这类复杂问题的处理,设计者必须把重点放在如何将加工对象的各种信息按其逻辑特性组织起来,抽象出合适的数学模型并合理地存储到计算机中。只有做完了这些工作,设计者才能设计解决具体问题的算法,并编写出相应的程序进行调试。请看下面几个例子。

【例 1.1】 人口信息查询。

根据身份证号码来查询一个人的信息,解决此问题,首先要构造一张人口信息登记表,表中每个登记项至少有身份证号、姓名等内容,如图 1.1(a)所示。这张图表达出人口信息数据的逻辑关系,也可以说图 1.1(a)就是一个数学模型。要写出好的查找算法,取决于这个登记表的结构及存储方式。最简单的方式是把表中的信息按照某种次序(如登记的次序)依次存储在计算机内一组连续的存储单元中。用高级语言表述,就是把整个表作为一个数组,表中的每个记录(即一个人的身份证号、姓名等)是数组的一个元素。按身份证号查找时,从表的第一项开始,依次查对身份证号,直到找出指定的身份证号或确定表中没有要找的人。

身份证号	姓名	...
340105750209001	王成明	...
100202861021007	陈海龙	...
220104800923002	王利芳	...
...
340504560511039	丁凯非	...
100104730816011	李达明	...

(a) 无序的人口信息表

身份证号	姓名	...
100104730816011	李达明	...
100202861021007	陈海龙	...
220104800923002	王利芳	...
...
340105750209001	王成明	...
340504560511039	丁凯非	...

(b) 有序的人口信息表

地址	身份证号	姓名	...
北京 10 1	100104730816011	李达明	...
...	100202861021007	陈海龙	...
吉林 22 11567
...	11567 220104800923002	王利芳	...
安徽 34 35689
...	35689 340105750209001	王成明	...
...	...	丁凯非	...

(c) 人口信息索引表

图 1.1 人口信息表

这种查找速度太慢,对于一个人口数量不多的地区或许是可行的,但对于全国范围或者一个有几千万甚至几亿人口的地区就不行了。一种常用的做法是把登记表按照身份证号从小到大排序(见图 1.1(b)),并存储在计算机内一组连续的存储单元中,在查找时可以采用折半查找算法(在第 9 章有详细介绍),查找速度将大大提高。我们还可以根据身份证前两位数字代表省份这一特征将此表建成索引表的形式(见图 1.1(c)),查找时根据身份证前两个数字首先找到所在的地区或省份,然后再按照某种方法继续查找下去,这种查找速度也比较快。

从这个例子可以看出,人口信息登记表如何构造、如何存储在计算机内存中将直接影响查找算法的设计以及算法的执行效率。图 1.1 所示的表便是我们为解决问题而建立的数学模型。这类模型的主要操作是查询、修改、插入、删除等。诸如此类的还有库存管理、图书资料管理等。在这种数学模型中,计算机处理的对象之间通常存在一种简单的线性关系,故这类数学模型可称为线性的数据结构。

【例 1.2】 人机对弈问题。

1996 年 IBM“深蓝”计算机与国际象棋大师卡斯帕罗夫进行人机大战,结果“深蓝”取得胜利,震惊了整个世界。十年之后的 2006 年浪潮“天梭”计算机战胜了由柳大华等五位中国象棋大师组成的大师团队。计算机之所以能和人对局并取得胜是因为有人将对局的策略事先存入计算机,通过其超强的逻辑判断和计算能力寻找最佳走法。在对弈问题中,计算机操作的对象是对弈过程中可能出现的棋盘格局或棋盘状态。图 1.2(a)是井字棋的一个格局,而格局之间的关系是由比赛规则决定的。图 1.2(b)是由图 1.2(a)可能产生的各种棋局,如果将对弈过程从初始状态到最后状态的各种可能都描述出来就会形成一棵“博弈树”。显然这种关系不是线性的,因为从一个棋盘格局可以派生出许多格局。“博弈树”的树根是对弈开始之前的棋盘状态,而所谓的叶子就是可能出现的结局,对弈的过程就是从树根到某个叶子的过程。棋子走法越多,“博弈树”就越大,对计算机存储和计算机的要求就越高,如浪潮“天梭”计算机每秒最多可计算 42 亿步棋。“博弈树”就是解决这类问题的数学模型,它也是一种数据结构。

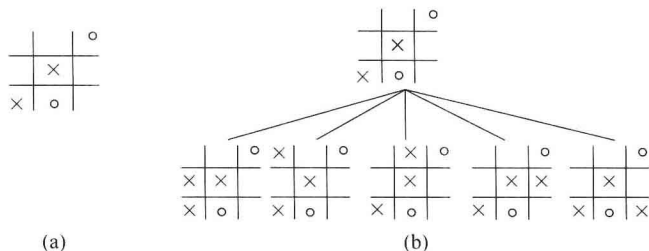


图 1.2 井字棋对弈树

【例 1.3】 最小代价问题。

几个村庄之间要架设输电线路,根据电能的可传递性,并不需要在每对村庄之间架设线路。如何用最小代价解决这个问题。

处理此类问题就需要用到图这种数学模型。用顶点代表村庄,每对顶点之间的边代

表村庄之间的线路,边的权值就是线路的建设费用。如图 1.3 所示。只要把这个图的有关信息存储在计算机中,利用图论中最小生成树算法,就可以解决最小代价问题。显然在图这种数学模型中,数据之间的关系更为复杂,但它能更准确的描述客观世界。图也是常见的数据结构之一。

从上述三个例子可见,描述非数值计算问题的数学模型不再是数学方程或数学公式,而是诸如表、树、图之类的数据结构。因此,简单地说数据结构是一门研究如何用计算机求解非数值计算问题的学科。

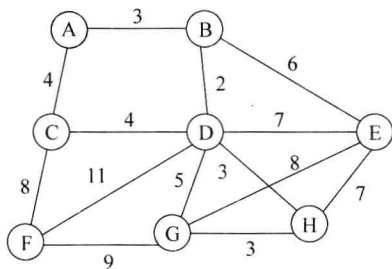


图 1.3 输电线路示意图

1.1.2 学习数据结构的意义

数据结构是计算机科学与技术专业的核心基础课程。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,仅掌握几种计算机程序设计语言是难以应付众多复杂问题的。要想有效地使用计算机、充分发挥计算机的性能,还必须学习和掌握好数据结构的相关知识。学好“数据结构”这门课程,将为学习计算机专业的其他课程(如操作系统、编译原理、数据库管理系统、软件工程、人工智能等)打下良好的基础。

数据结构作为一门独立的课程最早是在美国开设的,1968年美国唐·欧·克努特(D. E Knuth 1974年获图灵奖)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和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,结构程序设计成为程序设计方法学的主要内容,数据结构的地位显得更为重要。人们认为程序设计的实质是对确定的问题选一个好的结构,加上一种好的算法。当时,数据结构几乎和图论(特别是表、树的理论)为同义语。随后,数据结构这个概念被扩充到包括网络、集合代数、格、关系等方面,从而变成现在称之为“离散结构”的内容。然而,由于数据必须在计算机中进行处理,因此,不仅要考虑数据本身的数学特性,而且还要考虑数据的存储结构,这就进一步扩大了数据结构的内容。

瑞士计算机科学家沃斯(N. Wirth 1984年获图灵奖)曾以“算法+数据结构=程序”作为他的一本著作的名称。可见,程序设计的实质是对实际问题选择一种好的数据结构,并设计一个好的算法。因此,仅仅掌握几种计算机语言和程序设计方法,而缺乏数据结构知识,则难以解决众多复杂的问题。

数据结构的研究不仅涉及计算机硬件(特别是编码理论、存储装置、存取方法等),而且与计算机软件的研究有着密切的关系,无论是编译程序还是操作系统,都涉及如何组织数据,使检索和存取数据更为方便。因此,可以认为,数据结构是介于数学、计算机硬件和软件三者之间的一门核心课程。

目前,数据结构是我国高校计算机类相关专业的核心课程之一,也是其他信息类专业如信息管理、通信工程、信息与计算科学等必修课程之一。

1.2 基本概念和术语

本节中将对一些基本概念和术语加以定义和解释,这些概念和术语将在以后的章节中多次出现。

1.2.1 数据与数据元素

数据(data)是对客观事物的符号表示,它能被计算机识别、存储和加工处理。它是计算机程序加工的“原料”。例如,一个代数方程求解程序所用到的数据是整数和实数;一个编译程序处理的对象是字符串(源程序)。在计算机科学中,数据的含义相当广泛,如客观世界中的声音、图像等,经过某些处理也可被计算机识别、存储和处理,因而它们也属于数据的范畴。

数据元素(data element)是数据的基本单位,有时也称为元素、结点、顶点或记录。一个数据元素可能由若干数据项(data item)组成。例如,在例 1.1 中的人口信息查询系统中身份证号、姓名等数据项组成了一个数据元素,即一个数据元素包含了两个以上的数据项。数据项是最小标识单位,有时也称为字段、域或属性。数据元素也可以仅有一个数据项。

数据结构(data structure)是指数据元素之间的相互关系,即数据的组织形式。它一般包括以下三个方面的内容。

(1) 数据元素之间的逻辑关系,也称为数据的逻辑结构(Logical Structure)。它独立于计算机,是数据本身所固有的。

(2) 数据元素及逻辑关系在计算机存储器内的表示方式,称为数据的存储结构(storage structure)。它是逻辑结构在计算机存储器中的映射,必须依赖于计算机。

(3) 数据运算,即对数据施加的操作。运算的定义直接依赖于逻辑结构,但运算的实现必依赖于存储结构。

1.2.2 数据的逻辑结构与存储结构

数据的逻辑结构是从逻辑关系上描述数据,不涉及数据在计算机中的存储,是独立于计算机的。可以说,数据的逻辑结构是程序员根据具体问题抽象出来的数学模型。数据中元素通常有下列几种形式的逻辑关系。

(1) 集合:任何两个元素之间都没有逻辑关系,每个元素都是孤立的。

(2) 线性结构:结构中的元素之间存在一对一的关系,即所谓的线性关系。例如例 1.1 中的人口信息表就是一个线性结构。在这个结构中,各元素(由一个人的身份证号、姓名和其他信息组成)排列成一个表,第一个元素之后紧跟着第二个元素,第二个元素之后紧跟着第三个元素,以此类推,整个结构就像一条“链”,故有“线性结构”之称。

(3) 树形结构:结构中的数据元素之间存在一对多的关系。这种结构像自然界中的倒长的“树”一样,呈分支、层次状态。在这种结构中,元素之间的逻辑关系通常称做双亲

与子女关系。例如,家谱、行政组织结构等都可用树形结构来表示。

(4) 图状结构: 结构中的元素之间存在多对多的关系,也就是说元素间的逻辑关系可以是任意的。在这种结构中,元素间的逻辑关系也称做邻接关系。

通常将树形结构、图状结构归纳为非线性结构。因此,数据的逻辑结构可分为两大类,即线性结构和非线性结构。

数据的存储结构(也称物理结构)是指数据在计算机内的表示方法,是逻辑结构的具体实现。因此,存储结构应包含两个方面的内容,即数据元素本身的表示与数据元素间逻辑关系的表示。数据的存储结构有下列4种基本方式。

(1) 顺序存储: 其方法是把数据元素依次存储在一组地址连续的存储单元中,元素间的逻辑关系由存储单元的位置直接体现,由此得到的存储表示称为顺序存储结构(sequential storage structure)。高级语言中,常用一维数组来实现顺序存储结构。该方法主要用于线性结构,非线性结构也可通过某种线性化的处理,实现顺序存储。

(2) 链式存储: 将数据元素存储在一组任意的存储单元当中,而用附加的指针域表示元素之间的逻辑关系,由此得到的存储表示称为链式存储(linked storage structure)。使用这种存储结构时,往往把一个数据元素及附加的指针放在一个结构体中作为一个结点。在高级语言中,常用指针变量来实现链式存储。

(3) 索引存储: 该方法的特点是在存储数据元素的同时,还建立附加的索引表。图1.1(c)就是一个索引表,索引表中每一项称为索引项。索引项的一般形式是:(关键字,地址)。关键字是指能唯一标识数据元素的数据项。若每个数据元素在索引表中均有一个索引项,则该索引表称为稠密索引(dense index)。若一个索引项对应一组数据元素,则该索引表称为稀疏索引(sparse index)。

(4) 散列存储: 该方法是依据数据元素的关键字,用一个事先设计好的函数计算出该数据元素的存储地址,然后把它存入该地址中。这种函数称为散列函数,由散列函数计算出的地址称为散列地址。

上述4种存储方式,既可单独使用,也可组合使用。逻辑结构确定后,采取何种存储结构,要根据具体问题而定,主要的考虑因素是运算方便、算法效率的要求。

存储结构的描述与程序设计语言有关。通常用高级语言中的类型说明来描述存储结构,不必涉及计算机的内存地址。

1.2.3 数据运算

数据运算是针对数据施加的操作。每种逻辑结构都有一个基本运算的集合,最常用的基本运算有:检索(查找)、插入、删除、更新、排序等。因为这些运算是在逻辑结构上施加的操作,因此它们同逻辑结构一样也是抽象的,只规定“做什么”,无须考虑“如何做”。只有确定了存储结构后,才能考虑“如何做”。简而言之,运算是逻辑结构上定义,在存储结构上实现。

必须注意,数据结构包含逻辑结构、存储结构和运算三方面的内容。同一逻辑结构采用不同存储结构,得到的是不同的数据结构,我们常用不同的数据结构名来标识它们。例如,线性结构采用顺序存储,我们称之为顺序表;采用链式存储时则称为链表。同样,