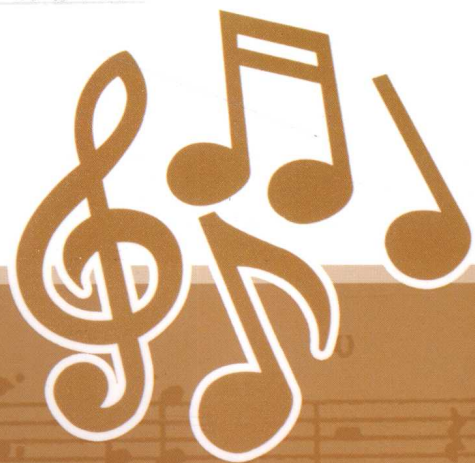
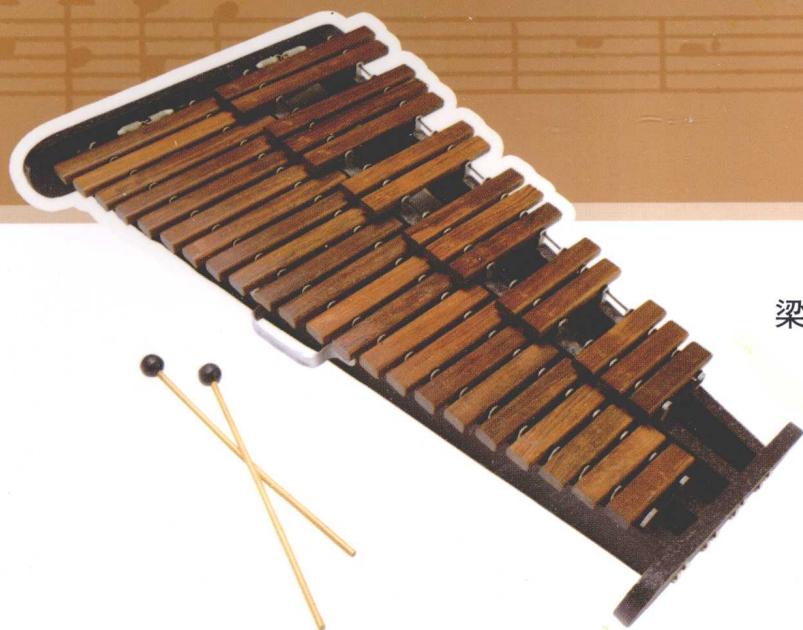


21世纪 应用型本科计算机科学与技术专业规划教材



数据结构与算法



龚丹 主 编
梁兴柱 蔡爱杰 王婧 副主编



清华大学出版社

21 世纪应用型本科计算机科学与技术专业规划教材

数据结构与算法

龚 丹 主编

梁兴柱 蔡爱杰 王 婧 副主编

清华大学出版社
北 京

内 容 简 介

本书为高等学校计算机科学与技术及相关专业“数据结构与算法”课程的教材。全书共分12章,较为系统地阐述了数据结构与算法的核心知识单元,包括基本数据结构、递归、数据类型和数据抽象、面向对象的程序设计、算法分析的基本方法和基本计算算法以及常用的算法设计策略等。

本书内容翔实、语言生动,注重理论叙述的完整性,更强调应用与实践,是培养应用型人才的院校,或希望较快地理解和掌握数据结构与算法相关实用知识并加以运用的学习者理想的教材形式。书中所有算法都有完整的C++程序,结构清晰、构思精巧,并在VC++ 6.0环境下编译及正确运行,既有助于读者更好地理解数据结构与算法的理论知识点,同时,又能充分地训练程序设计的能力。通过本书的学习,可以使读者建立科学合理的分析问题、解决问题及优化解决方案的思维过程,为今后学习或从事复杂程序设计奠定基础。

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

数据结构与算法/龚丹主编. —北京:清华大学出版社,2010.10

(21世纪应用型本科计算机科学与技术专业规划教材)

ISBN 978-7-302-23576-7

I. ①数… II. ①龚… III. ①数据结构 ②算法分析 IV. ①TP311.12

中国版本图书馆CIP数据核字(2010)第158511号

责任编辑:索梅薛阳

责任校对:李建庄

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

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

印 装 者:北京国马印刷厂

经 销:全国新华书店

开 本:185×260 印 张:21 字 数:521千字

版 次:2010年10月第1版 印 次:2010年10月第1次印刷

印 数:1~3000

定 价:29.50元

产品编号:036144-01

编写委员会成员名单

名誉主任：李建中

主任：郝忠孝

副主任：周洪玉

委员：

王乃茂	王培东	王国权	苏晓东	周屹
牛皖闵	马慧彬	朱景富	黄凤岗	陈本土
高洪志	高巍巍	李人贤	常键斌	安齐国
韩凤来	杜凯	王克家	孙斐朗	贾宗福
马春华				

秘书：马宪敏



21世纪是信息产业大发展的时代,计算机技术成为信息社会的重要支柱。信息化社会对人才的培养提出了更高的要求 and 标准。掌握计算机技术并具有应用计算机的能力是适应信息化社会的基础。

这套计算机系列教材适用于培养应用型人才,突出实验教学,突出实用,培养学生动手能力,掌握最新技术,适应社会需求。

本套教材在编写模式和思路上有了较大变化,采取面向任务,面向目标,先提出问题,然后指出解决问题的方法和所需要的知识的项目驱动式教材编写指导思想。针对目标,明确任务,做什么项目,用什么知识;用什么,学什么,学什么,会什么;急用先学,学以致用;突出重点,突出有用;然后由此及彼,由表及里,由浅入深,先感性,后理性,先实践,后理论,先认识,后提高;先掌握基本应用,然后做理论讲解、扩展与延伸,最后落实到具体操作,指导学生动手设计,用实践检验对知识的掌握程度。

本套教材特点是:内容丰富,知识全面,项目驱动,图文并茂,案例教学,贯彻始终。结构严谨,层次分明,条理清晰,通俗易懂,由浅入深,深入浅出,循序渐进。减少交叉,避免重复,编排合理,精心设计,突出重点,化解难点。学习理论,上机实验,举一反三,学用结合,配备习题,提供试题,联系实际,提高能力。

我们从计算机技术的发展趋势和信息社会对人才培养的需求出发,实现知识传授与能力培养的有效结合,通过对教学内容的基础性、科学性和应用性的研究,体现以有效知识为主体,构建支持学生终身学习的计算机知识基础和能力基础,提高学生计算机的应用能力。本系列教材强调理论与实践相结合,既注重基本原理、基本概念的介绍,又注重基本操作、基本能力的培养,根据计算机技术的发展和应用,加重了项目实训的内容。提高学生的动手能力。本套教材由三个部分组成,一是教材本身,二是实践实验教程,三是配套电子课件和素材(可到清华大学出版社网站 www.tup.com.cn 上下载)。

教育是科学,其价值在于求真。教育是艺术,其生命在于创新。大学教育真正要教会学生的应该是学习精神、学习能力、应用和创新能力。学习应该是超越课本知识的一个过程。本系列教材内容广泛新颖、取材丰富实用、阐述深入浅出、结构合理清晰。本系列教材的出版,不仅是编者们的努力的结果,同时也凝结了编委会许多人的心血,清华大学出版社的编辑

们为系列教材的出版任劳任怨、一丝不苟。因此,本系列教材的出版是集体智慧的结晶,是各院校优势互补、突出学校特色、进行计算机应用型人才培养的一次有益尝试。在此,编委会向所有为本系列教材的出版付出辛勤劳动的教师们及清华大学出版社的同仁们表示崇高的敬意和衷心的感谢!本系列教材在编写过程中也得到黑龙江省教育厅的悉心指导以及许多高校的大力支持,特别是哈尔滨师范大学恒星学院院长邓中兴教授给予了热情帮助和大力支持,也得到了许多计算机公司的帮助,编委会在此向他们表示衷心感谢!

本系列教材既可作为高等学校计算机专业的教材,也可作为信息技术的培训教材或参考书。

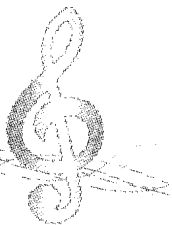
由于时间仓促,书中粗浅疏漏或叙述欠严密之处在所难免,恳请读者批评指正,热切期待着授课教师在教学实践中对系列教材提出宝贵意见和建议。我们将每年对系列教材进行一次认真的修订。

郝忠孝

2010年1月

前言

FOREWORD



“数据结构与算法”是计算机科学研究的主领域之一，是计算机科学与技术及相关专业学生必修的专业基础课，其核心知识单元包括：基本数据结构、递归、数据类型和数据抽象、面向对象的程序设计、算法分析的基本方法和基本计算算法以及常用的算法设计策略。

在我国，数据结构和算法相关技术的研究以及课程的教学已经有 30 余年的历史，但是随着计算机学科的高速发展和高等教育体制的深刻变革，对课堂教学用的教材，包括自学的参考教材都提出了更多的不同层次上的要求，特别是为满足我国当前社会生产、生活以及各类经济活动的实际需求，迫切地需要既有理论基础又有实践能力的应用型专业人才，这就要求在教材中既要包含核心的、系统的理论内容，又要包括丰富的、操作性强的示例，使读者在明确数据结构与算法“是什么”的基础上，能够深刻理解通过学习能够“做什么”、掌握“怎么做”，同时获得如何能“做得更好”的启发。

数据结构主要研究的内容是数据、数据的关系以及数据的基本运算，算法主要研究的是问题求解的方法和效率，二者研究内容不同却又密切相关，有如“米”与“炊”，虽互相独立却彼此依存。因此本书将数据结构和算法二者有机结合，以简洁明了、一气呵成的方式完成实用理论到典型应用的学习过程，使读者可以更好、更快地领会“数据结构+算法=程序”这一著名公式的要义。全书共分 12 章，总体划分为 3 大部分，第 1 部分为概述，即第 1 章；第 2 部分为第 2 章至第 11 章，共 10 章，分别介绍典型数据结构的抽象、表示和实现，包括线性表、串、栈、队列、数组、广义表、树、图，以及基本计算方法查找和排序；第 3 部分，即第 12 章，为经典算法的实例分析与应用，包括分治、动态规划和回溯，同时介绍了泛型编程的思想及 C++ 标准模板库的应用。

本书的主导思想是理论够用、知识连续，重在应用、培养能力，既注重数据结构与算法的基本原理，又突出数据结构相关理论在实际中如何运用，并将经典算法与实用技术相结合解决实际编程问题，强调程序设计的实战。学以致用，直观易懂，避免学习者，特别是以应用为目标的 learners 在学习抽象程度较高、理论深度较大的知识时的畏难心理，是应用型院校和希望较快速地理解和掌握数据结构与算法相关知识的学习者理想的教材形式。

书中的算法都有完整的 C++ 程序，结构清晰、构思精巧，并在 VC++ 6.0 环境下编译及

正确运行,能够帮助读者更好地理解相关的理论知识点,同时充分地训练程序设计的能力。本书旨在通过一系列核心知识点和应用实例的学习和训练后,帮助读者建立并完善分析问题、解决问题及优化解决方案的思维过程,以便更有信心去挑战更为复杂的程序设计问题。

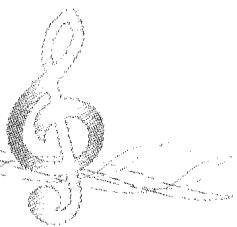
本书的编写团队主要来自应用型院校,编者希望将自身在应用型人才培养一线的教学经验通过本书加以体现,也欢迎各界的专家、同仁以及读者与我们进一步交流。主编龚丹,副主编梁兴柱、蔡爱杰、王婧,一起参与编写的还有姚璐、邢芳卉、黄磊、孙海龙、张振蕊、王振力、王家宁、王莉莉、张志维等老师,虽然经过反复的讨论和校对,但难免会有错漏之处,敬请原谅和指教。编者电子邮箱 pcgong1001@yahoo.com.cn。

编 者

2010年8月于北国冰城

目 录

CONTENTS



第 1 章 概述	1
1.1 数据结构的基本概念	1
1.1.1 数据结构概念的提出	1
1.1.2 数据结构中的常用术语	3
1.2 问题抽象与抽象数据类型	5
1.2.1 抽象——成功设计的基石	5
1.2.2 抽象数据类型的表示和实现	8
1.3 算法的基本概念	11
1.3.1 算法及算法设计的要求	11
1.3.2 算法效率的分析和量度	12
1.4 C++ 语言要点介绍	13
1.4.1 数组和指针	13
1.4.2 类	17
1.4.3 函数及函数重载	19
1.4.4 模板	22
1.4.5 其他自建类型的方法	25
1.5 小结	25
习题	26
第 2 章 线性结构	28
2.1 线性表的类型定义	28
2.1.1 线性表的逻辑特性	28
2.1.2 线性表的 ADT 定义	29
2.2 线性表的顺序存储表示	30
2.2.1 顺序表	30
2.2.2 线性表的其他顺序存储表示	39
2.3 线性表的链式存储	46
2.3.1 单链表	46
2.3.2 线性表的其他链式存储表示	53
2.4 线性表不同存储表示方法的对比	55
2.5 集合运算的模拟	55

2.5.1	问题描述与算法分析	56
2.5.2	算法实现	57
2.6	小结	58
	习题	59
第3章	串与文本编辑	60
3.1	串的类型定义	60
3.2	串的存储表示	62
3.2.1	串的顺序存储	63
3.2.2	串的链式存储	69
3.3	串的模式匹配算法	72
3.4	文本编辑	76
3.4.1	问题描述与算法分析	76
3.4.2	算法实现	77
3.5	小结	79
	习题	79
第4章	栈与递归	80
4.1	栈的类型定义	80
4.1.1	栈的逻辑特征	80
4.1.2	栈的 ADT 定义	81
4.2	栈的存储表示	82
4.2.1	栈的顺序存储	82
4.2.2	栈的链式存储	89
4.3	递归问题	94
4.3.1	$n!$	96
4.3.2	迷宫问题	98
4.4	小结	103
	习题	103
第5章	队列与离散事件模拟	105
5.1	队列的类型定义	105
5.1.1	队列的逻辑特性	105
5.1.2	队列的 ADT 定义	106
5.2	队列的存储表示	106
5.2.1	队列的顺序存储	106
5.2.2	队列的链式存储	110
5.3	离散事件模拟	114
5.3.1	问题描述与算法分析	114

5.3.2	算法实现	116
5.4	小结	118
	习题	119
第6章	数组与矩阵压缩	120
6.1	数组的类型定义	120
6.1.1	数组的逻辑特性	120
6.1.2	数组的 ADT 定义	121
6.2	数组的存储表示	121
6.3	矩阵压缩	122
6.3.1	特殊矩阵的压缩存储	123
6.3.2	稀疏矩阵的压缩存储	124
6.4	小结	129
	习题	130
第7章	广义线性表	131
7.1	广义表的类型定义	131
7.1.1	广义表的逻辑特性	131
7.1.2	广义表的 ADT 定义	132
7.2	广义表的存储表示	134
7.2.1	头尾表示法	134
7.2.2	孩子兄弟表示法	137
7.3	多级管理机构问题	139
7.3.1	问题描述与算法分析	139
7.3.2	算法实现	140
7.4	小结	142
	习题	143
第8章	树型结构及其应用	144
8.1	基本概念和术语	144
8.1.1	树	144
8.1.2	树的 ADT 定义	147
8.2	二叉树	148
8.2.1	二叉树的概念及特性	148
8.2.2	二叉树的存储表示	153
8.2.3	遍历二叉树	157
8.2.4	二叉树的线索化	160
8.3	树和森林	162
8.3.1	树的存储表示	162

8.3.2	树与二叉树的转换	165
8.3.3	树和森林的遍历	167
8.4	表达式求值问题	168
8.4.1	问题描述与算法分析	168
8.4.2	算法实现	169
8.5	哈夫曼树	171
8.5.1	最优二叉树	173
8.5.2	哈夫曼树的存储表示	174
8.5.3	哈夫曼编码问题	176
8.6	小结	179
	习题	180
第9章	图型结构及其应用	182
9.1	图的类型定义	182
9.1.1	图的逻辑特性	182
9.1.2	图型结构中的基本概念和术语	183
9.1.3	图的 ADT 定义	185
9.2	图的存储表示	187
9.2.1	数组表示法	187
9.2.2	邻接表表示法	191
9.2.3	十字链表表示法	198
9.3	图的遍历	200
9.3.1	图的深度优先遍历	200
9.3.2	图的广度优先遍历	202
9.4	图的连通性与最小生成树问题	203
9.4.1	图的连通性	203
9.4.2	图的最小生成树	205
9.4.3	工程造价问题求解	211
9.5	图的拓扑排序与工程工期问题	213
9.5.1	图的拓扑	214
9.5.2	关键路径	217
9.5.3	工程工期问题求解	220
9.6	最短路径问题	223
9.6.1	单源点的最短路径	223
9.6.2	任意顶点对间的最短路径	225
9.7	小结	227
	习题	227

第 10 章 查找	229
10.1 基本概念和术语	229
10.2 静态查找表	230
10.2.1 简单顺序查找表	230
10.2.2 有序表的查找	232
10.2.3 索引顺序表的查找	235
10.2.4 静态树表查找	236
10.3 动态查找表	239
10.3.1 二叉排序树和平衡二叉树	239
10.3.2 B-树	252
10.3.3 B+树	255
10.4 哈希查找表	255
10.4.1 哈希表及散列存储	256
10.4.2 哈希表的构造	257
10.4.3 哈希表的查找	261
10.5 小结	263
习题	263
第 11 章 内部排序	265
11.1 基本概念和术语	265
11.2 插入排序	267
11.2.1 直接插入排序	267
11.2.2 希尔排序	269
11.3 交换排序	271
11.3.1 冒泡排序	271
11.3.2 快速排序	272
11.4 选择排序	274
11.4.1 简单选择排序	275
11.4.2 树型选择排序	276
11.4.3 堆排序	277
11.5 归并排序	279
11.6 基数排序	282
11.7 小结	282
习题	283
第 12 章 经典算法与实例分析	286
12.1 STL 基础知识	286
12.1.1 STL 与泛型编程	286

12.1.2	STL 的部件	288
12.1.3	各部件的协同工作	291
12.1.4	STL 部件的应用	292
12.2	分治	300
12.2.1	算法理论介绍	300
12.2.2	分治算法的应用	301
12.3	动态规划	306
12.3.1	算法理论介绍	306
12.3.2	动态规划算法的应用	308
12.4	回溯	312
12.4.1	算法理论介绍	312
12.4.2	回溯算法的应用	313
12.5	小结	320
习题	320
参考文献	321

概 述

计算机的应用从最初专门用于大规模数值计算至今,已经越来越多地应用在更广泛的领域内,比如事务处理、人工智能等领域。问题的多样化给计算机程序设计带来一些新的问题,当单纯的过程化处理无法使程序效率得到提高时,程序设计者开始将研究的重心转向对计算机所处理的数据的组织上。

1.1 数据结构的基本概念

1.1.1 数据结构概念的提出

计算机是应大量的数值计算任务而诞生的,对于数值计算可以用最常见的计算器来体会一下,如 $x=0.3 \times 8 - \log_2 128$ 。在计算器上我们会依次输入相关的操作数和操作符,其中求对数的操作 \log 虽然不是一种操作符,但其计算过程也已经预置在计算器中。和其他操作符的使用方法一样,整个 x 的求解过程处理的对象都是数值。当然,计算机比计算器要“智能”一些,它可以预先存入一定的指令序列,启动计算过程后就可以等待输出结果了。计算机这种高效的处理能力越来越多地被应用。现在,计算机应用的领域已经远不止于数值计算,更多地向与日常工作和生活密切相关的一些非数值处理发展,如人事管理、人机游戏、交通咨询等,这些非数值处理应用与数值计算的显著差别就是被处理的数据彼此之间不是孤立的,而是存在某种特定的关系,进而利用好这些关系完成实际问题的处理。

如一个企业的人事管理系统,企业员工的信息,如姓名、性别、职位等,都不是数值,而且它们合在一起才形成了对一个员工的描述,习惯上被称为“记录”,这样形成一个基本的人事信息表(如表 1-1 所示)。对人事信息的操作最基本的就是检索,为了在海量的信息中尽快地找到某一个员工的信息,要用不同的方式为人事信息建立分类索引表(如表 1-2 所示为“部门”索引,表 1-3 为“职位”索引)。

表 1-1 示例——人事管理系统中的人事信息表

员工编号	姓名	性别	出生日期	部门	职位	联系方式
1001	李强	男	19700606	行政部	主任	85853236
1002	李强	男	19731202	财务部	职员	85853369

续表

员工编号	姓名	性别	出生日期	部门	职位	联系方式
1003	赵明	男	19770604	行政部	职员	85853256
1004	王莉	女	19660627	财务部	主任	85853365
1005	蔡芳	女	19821107	人事部	职员	85854600

表 1-2 示例——人事信息表的按部门分类索引表

部门编号	部门名称	员工编号集
P01	行政部	1001,1003,...
P02	财务部	1002,1004,...
P03	人事部	1005
...		

表 1-3 示例——人事信息表的按职位分类索引表

职位编号	职位名称	员工编号集
T01	主任	1001,1004,...
T02	职员	1002,1003,1005,...
...		

不论是基本人事信息表还是各个分类索引表,在人工管理阶段它们都被存储在一个个的本子上或者卡片上。为了管理方便,本子中的信息不是孤立地、杂乱地记录,而是被加上了一定的次序形成了不同的表格,管理人员可以按照这些次序去查找员工的信息。

在棋牌对弈中,对弈双方总是根据对方的前一步,来考虑自己当前如何回手。回手的选择可能有多种,初级的选手会任意地在其中选择一个,将阵局向前推动,而越是好的选手越能够谋划出更多轮对弈后的情况,精心选择一个使自己的下一轮仍然有尽可能多的选择同时尽量封掉对方的后路,最后取胜。每一种阵局状态就存储在棋手的头脑中,当前的一个阵局状态可以衍生出零个或多个下一阶段的状态。用圆圈表示状态,圆圈之间的连线表示状态与状态之间的关系,则可将棋牌对弈过程中阵局的发展抽象地描述,如图 1-1 所示。

如果你的公司接下一个网络线缆安装工程,在最基本的互连要求下要怎么样施工才比较经济?首先,施工前都要实地测量,将各个站点和具备安装条件的站点间的安装距离绘制成一张图纸,如图 1-2 所示。站点与站点之间可能有通路也可能没有,一个站点可能只与另一个站点有通路,也可能与多个站点有通路,反之亦然,只有围绕这样一张图来进行一些计算后,才能在众多的能够满足互连要求的方案中找到最经济的一个,进而实际施工。

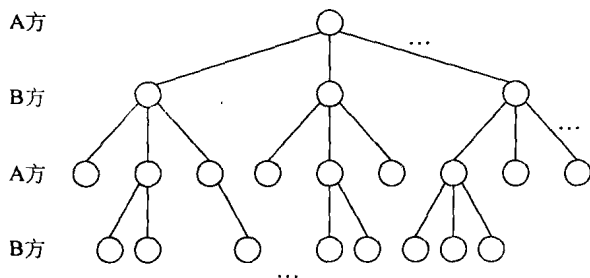


图 1-1 棋牌对弈时的阵局状态图。

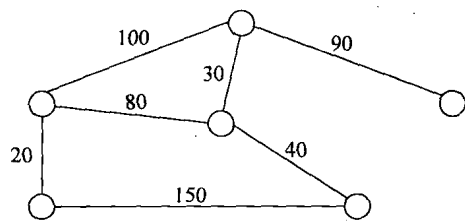


图 1-2 网络线缆安装工程现场测量图

以上所介绍的例子都是现实中常见的应用问题,并以人工操作的视角来分析,可以看出这些问题的求解都基于已有的或可以间接获得的数据,而这些数据不再是一些孤立的数值,合理地组织和表示出它们应有的关系才可以有效地解决问题。因此,当用计算机编制程序来代替人工操作时,就需要考虑应用问题中的数据在计算机中如何组织,以便完成问题求解

的目标。

“结构(structure)”在百科词典中即有“组成整体的各部分的搭配和安排”的释义,“数据结构(data structure)”就是问题域中数据及数据之间的特定关系在计算机中的表示和实现。

学界对于数据结构的定义尚未统一,对它的理解和定义有多种不同的表述方法,如:

Sartaj Sahni 在他的《数据结构、算法与应用》一书中称:数据结构是数据对象,以及存在于该对象的实例和组成实例的数据元素之间的各种联系。这些联系可以通过定义相关的函数来给出。他将数据对象(data object)定义为“一个数据对象是实例或值的集合”。

Clifford A. Shaffer 在《数据结构与算法分析》一书中的定义是:数据结构是 ADT (Abstract Data Type,抽象数据类型)的物理实现。

Lobert L. Kruse 在《数据结构与程序设计》一书中,将一个数据结构的设计过程分成抽象层、数据结构层和实现层。其中,抽象层是指抽象数据类型层,它讨论数据的逻辑结构及其运算,数据结构层和实现层讨论一个数据结构的表示和在计算机内的存储细节以及运算的实现。

国内的著作中,严蔚敏的《数据结构(C语言版)》中有两种解释,一是数据结构是相互之间存在一种或多种特定关系的数据元素的集合,另一个则是从研究内容上给出了数据结构这门学科的定义: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运算等的学科。

陈慧南的《数据结构与算法》一书中给出了一个较细致的表述:一个数据结构是由数据元素依据某种逻辑联系组织起来的。对数据元素间逻辑关系的描述称为数据的逻辑结构;数据必须在计算机内存储,数据的存储结构是数据结构的实现形式,是其在计算机内的表示;此外讨论一个数据结构必须同时讨论在该类数据上执行的运算才有意义。

徐绪松在《数据结构与算法》一书中对数据结构的定义非常简洁:数据结构指的是数据之间的结构关系,具体来说包括数据的逻辑结构和数据的物理结构。

以上各种表述有简有繁,虽然在定义上有所不同,但从中可以归纳出各方对数据结构研究内容上一致的观点:数据结构研究数据之间的关系,这些关系通过在该结构上定义的运算来表示,这一结构的实现离不开具体的计算机存储方式。

在许多类型的程序设计中,数据结构的选择是一个基本的设计考虑因素。精心选择数据结构可以大大提升问题解决方案的效率。许多大型系统的构造经验表明,系统实现的困难程度和系统构造的质量都严重地依赖于是否选择了最优的数据结构。

数据及其关系是系统构造的关键因素的理念也带来了许多优秀软件设计方法和程序设计语言的出现,面向对象的程序设计(OOP)及相关语言(如 C++)就是其中之一。

1.1.2 数据结构中的常用术语

数据是对客观事物的符号表示,在计算机科学中是指所有能输入到计算机中并由计算机程序处理的符号的总称。

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考虑。一个数据元素由若干个数据项组成。数据项是数据的不可分割的最小单位。数据元素可分为两类:一类是不可分割的原子型数据元素,如:整数“5”,字符“N”等;另一类是由多个数据项构成的数据元素,例如描述一个员工的信息的数据元素可由姓名、性别、出生日期、部门、职位、联系