

21世纪高等学校规划教材 | 计算机科学与技术



实用数据结构教程

颜辉 付宏 主编

潘冠宇 范丽梅 副主编

清华大学出版社

21世纪高等学校规划教材 | 计算机科学与技术



实用数据结构教程

颜辉 付宏 主编
潘冠宇 范丽梅 副主编

清华大学出版社
北京

内 容 简 介

本书共分 10 章,系统地介绍了各种典型的数据结构,主要内容包括绪论、线性表、栈和队列、串、数组和广义表、树和二叉树、图、查找、排序和文件。本书的内容编排由浅入深、语言通俗易懂,理论联系实际,强调应用能力的培养;在结构安排上,每章均以结构图形式系统地归纳和总结知识要点,配有应用案例、综合练习等内容,便于理解与掌握。

本书有配套教材《实用数据结构习题与实践》,在教学过程中与本书配合使用,既可作为高等院校计算机专业及相关专业的教材,也可作为计算机工程和应用技术人员的参考书。

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

实用数据结构教程/颜辉,付宏主编. —北京:清华大学出版社,2011.9

(21 世纪高等学校规划教材·计算机科学与技术)

ISBN 978-7-302-25415-7

I. ①实… II. ①颜… ②付… III. ①数据结构—高等学校—教材 IV. ①TP311.12

中国版本图书馆 CIP 数据核字(2011)第 074888 号

责任编辑:梁 颖

责任校对:李建庄

责任印制:王秀菊

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:北京四季青印刷厂

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:16.5 字 数:407 千字

版 次:2011 年 9 月第 1 版 印 次:2011 年 9 月第 1 次印刷

印 数:1~4000

定 价:26.00 元

产品编号:041056-01

编审委员会成员

(按地区排序)

清华大学	周立柱	教授
	覃征	教授
	王建民	教授
	冯建华	教授
	刘强	副教授
北京大学	杨冬青	教授
	陈钟	教授
	陈立军	副教授
北京航空航天大学	马殿富	教授
	吴超英	副教授
	姚淑珍	教授
中国人民大学	王珊	教授
	孟小峰	教授
	陈红	教授
北京师范大学	周明全	教授
北京交通大学	阮秋琦	教授
	赵宏	教授
北京信息工程学院	孟庆昌	教授
北京科技大学	杨炳儒	教授
石油大学	陈明	教授
天津大学	艾德才	教授
复旦大学	吴立德	教授
	吴百锋	教授
	杨卫东	副教授
同济大学	苗夺谦	教授
	徐安	教授
华东理工大学	邵志清	教授
华东师范大学	杨宗源	教授
	应吉康	教授
东华大学	乐嘉锦	教授
	孙莉	副教授

浙江大学	吴朝晖	教授
	李善平	教授
扬州大学	李 云	教授
南京大学	骆 斌	教授
	黄 强	副教授
南京航空航天大学	黄志球	教授
	秦小麟	教授
南京理工大学	张功萱	教授
南京邮电学院	朱秀昌	教授
苏州大学	王宜怀	教授
	陈建明	副教授
江苏大学	鲍可进	教授
中国矿业大学	张 艳	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
华中师范大学	叶俊民	教授
	郑世珏	教授
	陈 利	教授
江汉大学	颜 彬	教授
国防科技大学	赵克佳	教授
	邹北骥	教授
中南大学	刘卫国	教授
湖南大学	林亚平	教授
西安交通大学	沈钧毅	教授
	齐 勇	教授
长安大学	巨永锋	教授
哈尔滨工业大学	郭茂祖	教授
吉林大学	徐一平	教授
	毕 强	教授
山东大学	孟祥旭	教授
	郝兴伟	教授
中山大学	潘小轰	教授
厦门大学	冯少荣	教授
仰恩大学	张思民	教授
云南大学	刘惟一	教授
电子科技大学	刘乃琦	教授
	罗 蕾	教授
成都理工大学	蔡 淮	教授
	于 春	副教授
西南交通大学	曾华燊	教授

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。

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21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”。推出的特色精品教材包括:

(1) 21世纪高等学校规划教材·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) 21世纪高等学校规划教材·计算机科学与技术——高等学校计算机相关专业的教材。

(3) 21世纪高等学校规划教材·电子信息——高等学校电子信息相关专业的教材。

(4) 21世纪高等学校规划教材·软件工程——高等学校软件工程相关专业的教材。

(5) 21世纪高等学校规划教材·信息管理与信息系统。

(6) 21世纪高等学校规划教材·财经管理与应用。

(7) 21世纪高等学校规划教材·电子商务。

(8) 21世纪高等学校规划教材·物联网。

清华大学出版社经过二十多年的努力,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。

清华大学出版社教材编审委员会

联系人:魏江江

E-mail:weijj@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,也是计算机课程体系中的核心课程之一,是设计和实现系统软件及大型应用软件的技术基础,主要研究各种基本数据的逻辑结构、存储结构和基本运算的实现以及查找排序等常用算法的实现。通过学习数据结构课程,学会如何把现实世界的问题转化为计算机内部的表示和处理,进而学会组织数据,选择合适的逻辑结构和存储结构,设计算法,形成良好的程序设计风格,提高逻辑思维和抽象思维的能力。

本书主要面向应用型本科及大专院校的计算机类专业的学生,本着“培养技术应用型人才”的原则,力求以应用为主体,对基本理论进行深入浅出的阐述。本书采用类 C 语言来描述算法,具有如下特点。

(1) 简洁通俗:理论知识的阐述由浅入深、通俗易懂,对多年来形成的数据结构课程的教学内容进行了合理地剪裁,略去了一些抽象理论推导和数学证明的过程,淡化算法的设计分析和复杂的时空分析,简明扼要地介绍学生需要掌握的基础知识和技术。

(2) 实践性强:各章均配有具体案例,以案例驱动知识要点,以实践指导教学过程。同时贯彻分层次教学的思想,在案例中增加了难度较大的“拓展训练”内容,供能力较强的学生练习。

(3) 系统归纳:各章均以结构图形式系统地归纳和总结知识要点,以多角度深入讲解重点难点。

本书共分 10 章。第 1 章介绍数据结构的一般概念和算法分析的初步知识;第 2 章~第 5 章分别讨论线性表、栈与队列、串、数组和广义表等线性逻辑结构、存储结构以及对于不同存储结构的各种基本操作的算法实现;第 6 章和第 7 章分别介绍树和图这两种重要的非线性逻辑结构、存储方法及重要的应用;第 8 章和第 9 章分别讨论各种查找算法和排序算法;第 10 章介绍常用的文件结构。

本书由颜辉、付宏任主编,潘冠宇、范丽梅任副主编,汤赫男、吴大亲、胡海燕、贾玉红参编完成。第 1 章由付宏编写;第 2 章由颜辉、汤赫男编写;第 3 章由汤赫男编写;第 4 章由吴大亲编写;第 5 章由胡海燕编写;第 6 章由潘冠宇编写;第 7 章由范丽梅编写;第 8 章由吴大亲、胡海燕编写;第 9 章由颜辉、范丽梅编写;第 10 章由贾玉红编写。参加编写的还有杨明莉、高婷、刘磊、成桂玲等。全书由颜辉统稿,付宏主审。

本书编者都是多年从事本课程教学的一线教师,但由于水平有限,书中难免存在不足与疏漏之处,敬请读者及同行们予以批评指正。

编者

2011 年 6 月

目 录

第 1 章 绪论	1
1.1 数据结构概论	1
1.1.1 数据结构的研究对象	1
1.1.2 数据结构的基本概念	3
1.1.3 数据结构的研究内容	5
1.2 算法	7
1.2.1 算法的概念	7
1.2.2 算法的特征	7
1.2.3 算法的描述	8
1.2.4 算法分析及评价	9
1.3 C/C++	11
1.3.1 C/C++ 中常用的数据类型	11
1.3.2 C/C++ 语句	15
本章小结	17
综合练习	17
第 2 章 线性表	19
2.1 线性表及其逻辑结构	19
2.1.1 线性表的定义	19
2.1.2 线性表的逻辑特征	20
2.2 线性表的顺序存储结构	22
2.2.1 线性表的顺序存储结构——顺序表	22
2.2.2 顺序表基本运算的实现	23
2.3 线性表的链式存储结构	27
2.3.1 线性表的链式存储结构——链表	27
2.3.2 单链表基本运算的实现	27
2.3.3 顺序存储和链式存储的比较	31
2.3.4 双链表	31
2.3.5 循环链表	33
2.3.6 静态链表	33
2.4 线性表的应用案例	33
本章小结	38

综合练习	39
第 3 章 栈和队列	42
3.1 栈	42
3.1.1 栈的定义及操作特性	42
3.1.2 栈的顺序存储结构及其基本运算的实现	44
3.1.3 栈的链式存储结构及其基本运算的实现	46
3.1.4 栈的应用举例	47
3.2 队列	50
3.2.1 队列的定义及操作特性	50
3.2.2 队列的顺序存储结构及其基本运算的实现	51
3.2.3 队列的链式存储结构及其基本运算的实现	54
3.2.4 队列的应用举例	55
3.3 栈和队列的应用案例	56
本章小结	58
综合练习	58
第 4 章 串	60
4.1 串的定义及基本概念	60
4.2 串的存储结构	62
4.2.1 串的顺序存储结构——定长顺序串	62
4.2.2 串的链式存储结构——链串	63
4.2.3 堆	64
4.3 串的基本运算	65
4.4 串的模式匹配算法	69
4.4.1 Brute-Force 算法	69
4.4.2 KMP 算法	71
4.4.3 next 数组及 nextval 数组	72
4.5 串的应用案例	74
本章小结	75
综合练习	76
第 5 章 数组和广义表	77
5.1 数组	77
5.1.1 数组的基本概念	77
5.1.2 数组的存储结构	79
5.2 特殊矩阵的压缩存储	81
5.2.1 对称矩阵	82
5.2.2 三角矩阵	82

5.2.3	对角矩阵	83
5.3	稀疏矩阵	84
5.3.1	定义	84
5.3.2	稀疏矩阵的存储方法	84
5.4	稀疏矩阵的应用案例	87
5.5	广义表	88
5.5.1	广义表的定义及逻辑结构	88
5.5.2	广义表的存储	90
5.5.3	广义表的基本运算	92
	本章小结	92
	综合练习	93
第 6 章	树和二叉树	95
6.1	树的基本概念	95
6.1.1	树的定义	95
6.1.2	树的基本术语	98
6.1.3	树的逻辑表示方法	99
6.2	二叉树	100
6.2.1	二叉树概念及特点	100
6.2.2	二叉树性质	104
6.2.3	二叉树的存储结构	106
6.2.4	二叉树的基本运算及其实现	109
6.3	二叉树的遍历	110
6.3.1	二叉树遍历的概念	110
6.3.2	二叉树遍历的方法	110
6.3.3	二叉树遍历的应用	111
6.3.4	二叉树遍历递归算法	111
6.3.5	二叉树遍历非递归算法	112
6.4	线索二叉树	116
6.4.1	线索二叉树的定义	116
6.4.2	线索二叉树的存储结构	116
6.4.3	线索二叉树的遍历	118
6.5	树和森林	119
6.5.1	树的存储结构	119
6.5.2	树、森林与二叉树的转换	123
6.5.3	树和森林的遍历	125
6.6	哈夫曼树	126
6.6.1	哈夫曼树概述	126
6.6.2	哈夫曼树的构造	127

6.6.3 哈夫曼编码·····	129
6.7 树的应用案例·····	130
本章小结·····	131
综合练习·····	132
第7章 图 ·····	134
7.1 图的基本概念·····	134
7.1.1 图的定义·····	134
7.1.2 图的基本术语·····	135
7.2 图的存储结构·····	138
7.2.1 邻接矩阵·····	138
7.2.2 邻接表·····	141
7.2.3 十字链表·····	143
7.2.4 邻接多重表·····	144
7.2.5 图的存储方法的比较·····	145
7.3 图的遍历·····	146
7.3.1 图的遍历的概念·····	146
7.3.2 深度优先搜索遍历·····	146
7.3.3 广度优先搜索遍历·····	147
7.3.4 非连通图的遍历·····	150
7.4 图的应用·····	151
7.4.1 最小生成树·····	151
7.4.2 最短路径·····	153
7.4.3 拓扑排序·····	157
7.4.4 AOE网与关键路径·····	160
7.5 图的应用案例·····	164
本章小结·····	164
综合练习·····	164
第8章 查找 ·····	167
8.1 查找的基本概念·····	167
8.2 静态查找·····	168
8.2.1 顺序查找·····	169
8.2.2 二分查找·····	171
8.2.3 分块查找·····	174
8.3 动态查找·····	175
8.3.1 二叉排序树·····	176
8.3.2 平衡二叉树·····	182
8.3.3 B-树·····	188

8.3.4 B ⁺ 树	192
8.4 哈希查找	193
8.4.1 哈希表的基本概念	193
8.4.2 哈希函数的构造方法	194
8.4.3 哈希冲突解决方法	196
8.4.4 哈希表上的运算	198
本章小结	200
综合练习	201
第9章 排序	203
9.1 排序的基本概念	203
9.1.1 基本概念	203
9.1.2 排序方法的分类	204
9.1.3 待排序序列的存储方法	204
9.2 插入排序	204
9.2.1 直接插入排序	204
9.2.2 希尔排序	206
9.3 交换排序	208
9.3.1 冒泡排序	208
9.3.2 快速排序	209
9.4 选择排序	212
9.4.1 直接选择排序	212
9.4.2 堆排序	213
9.5 归并排序——二路归并排序	219
9.6 基数排序	221
9.7 各种排序方法的综合比较	224
9.8 外部排序简介	225
本章小结	226
综合练习	226
第10章 文件	228
10.1 有关文件的基本概念	228
10.1.1 文件及其类别	228
10.1.2 记录的逻辑结构和物理结构	229
10.1.3 文件的操作(运算)	229
10.1.4 文件的物理结构	230
10.2 顺序文件	230
10.3 索引文件	233
10.3.1 索引非顺序文件	233

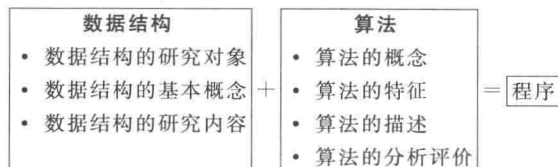
10.3.2 索引顺序文件	235
10.4 散列文件	239
10.5 多关键字文件	240
10.5.1 多重表文件	241
10.5.2 倒排文件	242
本章小结	243
综合练习	243
参考文献	245

第 1 章

绪论

利用计算机进行数据处理是计算机应用的一个重要领域。计算机处理的数据一般包括数值数据和非数值数据。瑞士计算机科学家 N. Wirth 教授曾经提出这样一个等式：算法 + 数据结构 = 程序，可见数据结构在程序设计中的重要地位。

下面是本章知识结构图：



1.1 数据结构概论

1.1.1 数据结构的研究对象

计算机科学是一门研究用计算机进行信息表示和处理的科学，计算机对信息的加工处理不仅仅是单一的数值计算，大量解决的是如图像、声音、字符等复杂的非数值数据，非数值数据问题的结构复杂，数据元素之间的相互关系一般无法用数学的一般公式来描述，必须分析待处理的对象的特性以及各处理对象之间存在的关系，即只有设计出合理的数据结构，才能有效地解决问题。

例 1.1 对于学生成绩信息管理的问题，可以在计算机中建立一张按学号顺序排列的学生成绩信息表(表 1-1)，按照一定的算法就可以自动检索学生的成绩信息。在这种文档管理的信息表中，每个学生的成绩信息之间存在着简单的线性关系，故称为线性结构。若将每个学生的完整信息浓缩成一个点元素，用带箭头的线表示点元素之间的关系，其抽象的拓扑结构，如图 1-1 所示，可见线性结构中数据元素之间是一一对一的关系。

表 1-1 学生成绩信息表

学 号	姓 名	高等数学(分)	计算机导论(分)	英语(分)
09401	陈 宇	85	69	92
09402	赵 明	87	73	74
09403	李晓影	92	64	84
...



图 1-1 线性结构拓扑图

例 1.2 对于 Windows 磁盘管理问题主要讨论的是在磁盘上如何组织和实现对所有文件和文件夹的管理,其逻辑结构可以用树形结构来描述,在这种结构中,文件夹和文件之间的层次关系被清晰地体现出来,故树形结构又称为层次结构,如图 1-2 所示。

其抽象的拓扑结构,如图 1-3 所示,可见树形结构中数据元素之间是一对多的关系。

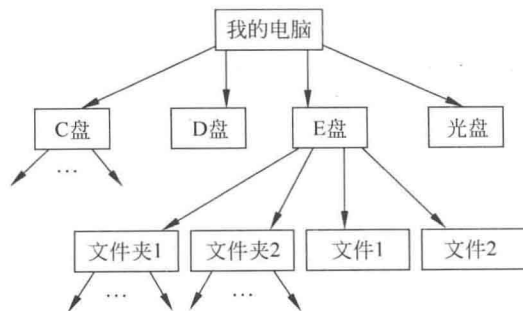


图 1-2 Windows 磁盘管理示意图

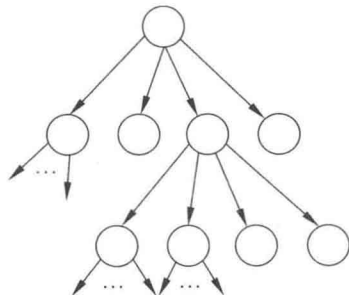


图 1-3 树形结构拓扑图

例 1.3 一个教学计划包含多门课程,这些课程之间有些必须按照先修和后续的次序来安排,有些课程则没有这样的次序要求可以任意地安排,描述这类数据元素之间的关系的结构称为图形结构,又称为图状结构。表 1-2 为计算机软件方面的一些课程设置编排表,其抽象的拓扑结构,如图 1-4 所示,图形结构中数据元素之间是多对多的关系。

表 1-2 课程设置编排表

课程编号	课程名称	先修课程
c ₁	高等数学	无
c ₂	计算机导论	无
c ₃	离散数学	c ₁
c ₄	程序设计语言 C	c ₁ 、c ₂
c ₅	数据结构	c ₃ 、c ₄
c ₆	计算机原理	c ₂ 、c ₄
c ₇	数据库原理	c ₄ 、c ₅ 、c ₆

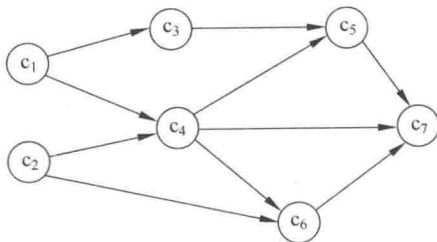


图 1-4 图形结构拓扑图

综上所述,数据结构主要是研究非数值计算的程序设计问题中出现的计算机操作对象,以及它们之间的关系和操作的学科。

1.1.2 数据结构的基本概念

为了更好地理解数据结构这门课程,首先来了解和掌握数据结构中的一些基本概念和术语。

数据(Data):指所有能输入到计算机中并能被计算机程序识别和处理的符号集合。

数据一般分为数值型数据和非数值型数据。

数值数据:包括整数、实数或复数等。主要用于工程计算、科学计算。

非数值数据:包括字符、文字、图形、图像、语音等。用于情报检索、企业管理、人工智能、远程教育、远程医疗、电子商务、电子图书馆和办公自动化等诸多领域。

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。一个数据元素可以由一个或若干个数据项组成。数据元素也被称为结点或记录。

数据项(Data Item):是数据的具有独立意义的不可分的最小单位,它是对数据的数据元素属性的描述,又称为字段或域。

如表 1-1 所示的学生成绩信息表,整张表是学生的成绩数据,除去表头之外,表中的每一行代表每个学生的完整成绩信息,称为记录,又称为数据元素,而每一个数据元素又是由不同科目的成绩(即数据项)构成的,又称做字段,如图 1-5 所示。

学号	姓名	高等数学(分)	计算机导论(分)	英语(分)
09401	陈宇	85	69	92
09402	赵明	87	73	74
09403	李晓影	92	64	84

图 1-5 数据结构基本概念示意图

数据对象(Data Object):是具有相同性质的数据元素的集合,它是数据的子集。如上例所示,一个班级的成绩表可以看做一个数据对象。

数据类型(Data Type):是具有相同性质的计算机数据的集合及定义在这个数据集合上的一组操作的总称。

例如,C语言中的整型数据类型是集合 $C = \{0, \pm 1, \pm 2, \pm 3, \dots\}$ 及定义在这个集合上的加、减、乘、除、取余运算等一组操作。

构成数据元素的每个数据项属于某个确定的基本数据类型,数据类型可以分为两类:原子类型和结构类型。

(1) 原子类型:如果一个数据元素由一个数据项构成,这个数据元素的类型就是这个数据项的数据类型,其值在逻辑上是不可分解的。例如,C语言中的基本数据类型(整型、实型、字符型)、指针类型和空类型。

(2) 结构类型:如果一个数据元素由多个不同类型的数据项组成,这个数据元素的类