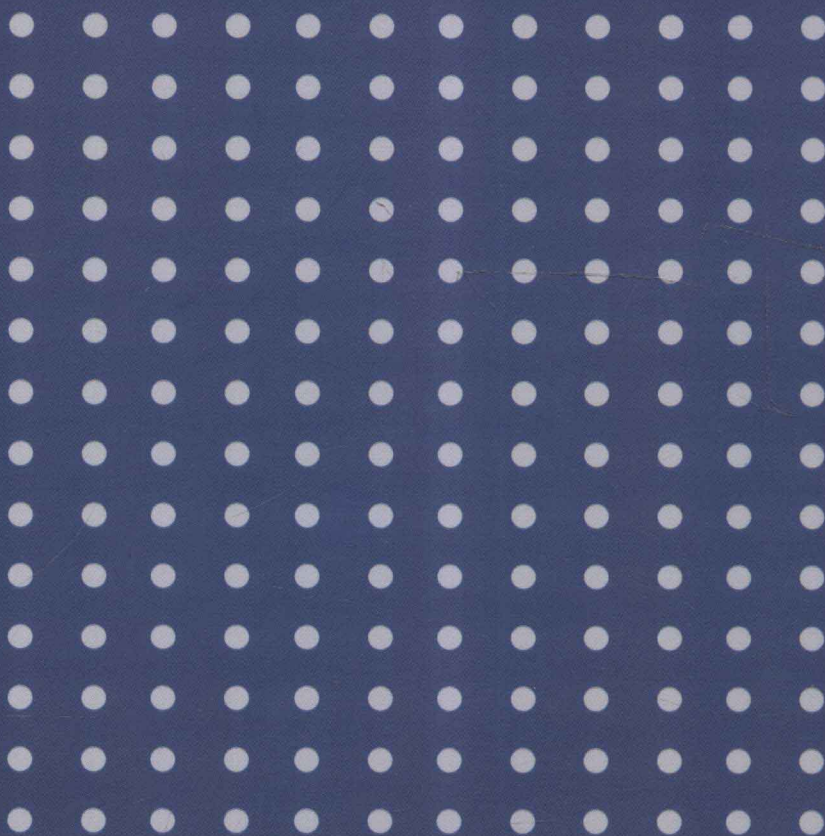


重点大学计算机专业系列教材

数据结构基础教程(C语言)

叶小平 陈瑛 编著



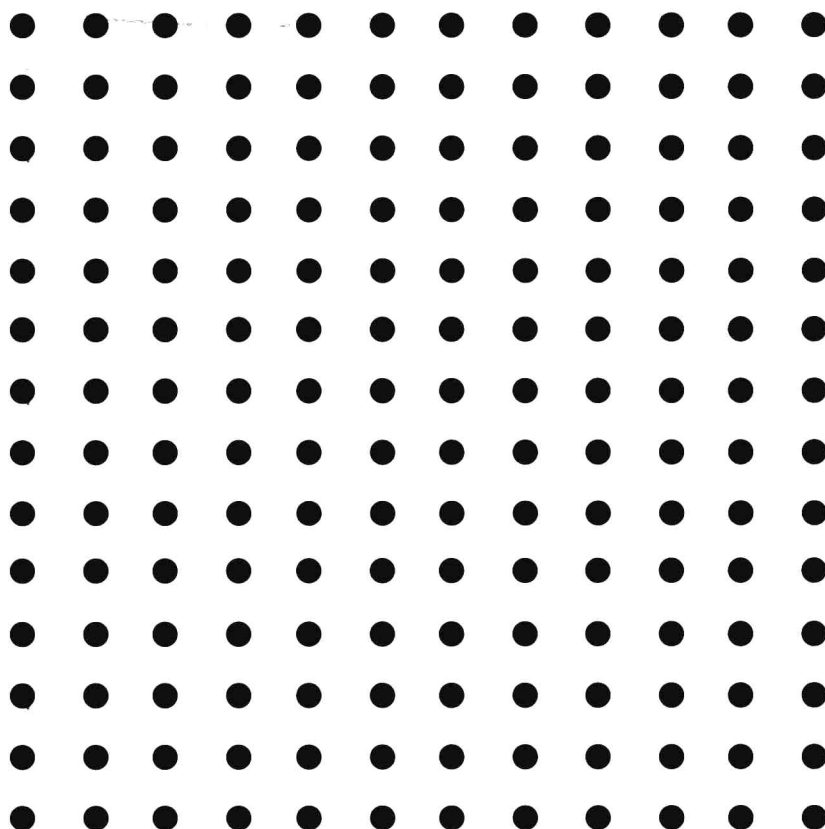
清华大学出版社



重点大学计算机专业系列教材

数据结构基础教程(C语言)

叶小平 陈瑛 编著



清华大学出版社

北京

内 容 简 介

本书是数据结构课程的基本教材。全书分为3个部分：课程概述、基于内存的数据结构(线性结构、树结构、图结构和集合)和基于文件的数据结构(文件的组织与查找)。本书的编写思路 and 材料组织具有体现整体结构、注重本质关联、突出关键细节和强化实例讲解等特点，同时书中的基本算法和实例实现程序都经过VC平台调试运行，实现教材学习到实验操作的无缝对接。本书还配有相应学习指导书供相关人员选用。

本书可供高等院校计算机信息科学与技术及其相关专业本科教学使用，同时也适合于非计算机专业人员自学使用。

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

数据结构基础教程：C语言/叶小平，陈瑛编著. --北京：清华大学出版社，2012.8

(重点大学计算机专业系列教材)

ISBN 978-7-302-28840-4

I. ①数… II. ①叶… ②陈… 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(2012)第102470号

责任编辑：魏江江 王冰飞

封面设计：常雪影

责任校对：焦丽丽

责任印制：宋 林

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>, <http://www.wqbook.com>

地 址：北京清华大学学研大厦A座 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：010-62770175, zhiliang@tup.tsinghua.edu.cn

课 件 下 载：<http://www.tup.com.cn>, 010-62795954

印 装 者：北京嘉实印刷有限公司

经 销：全国新华书店

开 本：185mm×260mm

印 张：19

字 数：466千字

版 次：2012年8月第1版

印 次：2012年8月第1次印刷

印 数：1~3000

定 价：29.50元

产品编号：041138-01

出版说明

随着国家信息化步伐的加快和高等教育规模的扩大,社会对计算机专业人才的需求不仅体现在数量的增加上,而且体现在质量要求的提高上,培养具有研究和实践能力的高层次的计算机专业人才已成为许多重点大学计算机专业教育的主要目标。目前,我国共有 16 个国家重点学科、20 个博士点一级学科、28 个博士点二级学科集中在教育部部属重点大学,这些高校在计算机教学和科研方面具有一定优势,并且大多以国际著名大学计算机教育为参照系,具有系统完善的教学课程体系、教学实验体系、教学质量保证体系和人才培养评估体系等综合体系,形成了培养一流人才的教学和科研环境。

重点大学计算机学科的教学与科研氛围是培养一流计算机人才的基础,其中专业教材的使用和建设则是这种氛围的重要组成部分,一批具有学科方向特色优势的计算机专业教材作为各重点大学的重点建设项目成果得到肯定。为了展示和发扬各重点大学在计算机专业教育上的优势,特别是专业教材建设上的优势,同时配合各重点大学的计算机学科建设和专业课程教学需要,在教育部相关教学指导委员会专家的建议和各重点大学的大力支持下,清华大学出版社规划并出版本系列教材。本系列教材的建设旨在“汇聚学科精英、引领学科建设、培育专业英才”,同时以教材示范各重点大学的优秀教学理念、教学方法、教学手段和教学内容等。

本系列教材在规划过程中体现了如下一些基本组织原则和特点。

(1) 面向学科发展的前沿,适应当前社会对计算机专业高级人才的培养需求。教材内容以基本理论为基础,反映基本理论和原理的综合应用,重视实践和应用环节。

(2) 反映教学需要,促进教学发展。教材要能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。规划教材建设的重点依然是专业基础课和专业主干课;特别注意选择并安排了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重点大学

计算机专业教学内容和课程体系改革成果的教材。

(4) 主张一纲多本,合理配套。专业基础课和专业主干课教材要配套,同一门课程可以有多本具有不同内容特点的教材。处理好教材统一性与多样化的关系;基本教材与辅助教材以及教学参考书的关系;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,择优落实。在制订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、评审确定主编。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。建立一支高水平的以老带新的教材编写队伍才能保证教材的编写质量,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。

教材编委会

前言

教育部高校计算机科学与技术教学委员会指出“数据结构与算法是计算机学科本科教学计划中的骨干基础课程,对学生基本的计算机问题求解能力的培养具有重要意义。作为一门必修课程,该课程既是对以往课程的深入和扩展,也是为将来更加深入学习其他专业课程打下基础。课程中所学习的排序问题的算法以及基本的树、图等数据结构,是计算机学科的基本功。B⁺树、散列等高级数据结构,也是数据库、操作系统、编译原理、计算机网络等后续课程的基础。”上述论述充分说明了“数据结构与算法”课程在整个计算机学科领域中的地位 and 意义。

计算机学科通常可以用来涵盖计算机科学、计算机技术和计算机工程 3 个方面,而科学性、技术性和工程性实际上是从不同侧面反映了计算机学科的本质特征,也是从不同角度和层面描述刻画和处理实现各类计算问题。数学基础的建立和数学思想方法的运用是学科科学性的重要标志。首先,对于计算机学科而言,构造性数学(如离散数学)是其主要理论的展开框架。数据结构以集合、树和图等数学概念为框架支柱,以形式化语言表述数据模型,以整体统一视野研究抽象数据类型,这些都集中体现了数学对计算机应用的指导,充分展示着计算机学科的科学性态。其次,计算机技术的灵魂是算法,在计算机技术发展过程中,发现或创立一种有效的算法通常都被视作是对计算机学科的一种实质性贡献。数据结构课程的核心内容是计算模型的建立(建立计算对象的逻辑与存储模型)、计算方法的设计和计算效率的评估,特别是其中大量充满人类智慧和科学美感的实例算法,这些对于学习者计算机技术基本素质技巧的培养和训练有着至关重要的意义,发挥着不可替代的作用。另外,计算机工程是对实际问题中相应科学分析结论和技术设计方案的实现,其中的基本前提就是程序设计语言复杂而灵活的运用。数据必须在计算机中进行处理,计算机处理数据的前提是数据的存储,数据存储和相应数据操作的实现需要灵活高效地操纵高级程序设计语言,因此,“数据结构与算法”课程使得学习者获得综合的体验机会和难得的训练环境。数据结构在一门课程当中自然而完美地整合了计算问题的数学分析建模、基本算法构建和

基于高级程序设计语言的具体实现,从而体现出课程的重要意义和强大魅力。

教材建设是教学改革的基本组成部分,也是提高教学质量、培养高素质人才的重要前提和评估指标之一。在本书编写过程中,我们借鉴国内外优秀教材和相关材料,努力做到下述几个方面。

理念决定成败 人们进行某项工作,特别是完成某些比较困难而又具有意义的工作,都需要一种内在的驱动力量,否则就难以保持坚忍不拔的意志去面对和克服出现的问题、困难与挫折。这种内在的驱动力量在日常生活中就是信仰,在学习和研究过程中就是理念。相对于其他计算机课程,“数据结构与算法”有着数学抽象性、算法多样性和程序语言实现灵活性等显著特点,这些都会让学习过程出现不少困难与艰辛。怎样在这些问题面前保持正确的理念导向,从而采取相应的积极正确应对措施呢?我们的理解是,需要从整个计算机学科的全局上明确课程的重要地位。首先,“数据结构”课程可看作是众多计算机前修基础课程的汇集处与深化点。研究数据的集合、线性表、树和图等组织形式和相互之间的关系,是“离散数学”课程的深化与扩展;研究各种算法的设计与实现,为“程序设计语言”等课程提供各类设计方法与技巧在更高层面上和更广阔语境中的总结与升华;研究算法的时间与空间复杂性,在具体平台上展开“计算概论”等课程中算法可行性和复杂度的基本原理,同时也为在应用环境中理解掌握“概率统计”原理方法提供实战训练。其次,“数据结构”也是计算机专业“编译原理”、“计算机网络”、“数据库技术”和“图形图像处理”等重要应用课程的基础,单就树结构来说,就有操作系统中多级菜单的层次树、算法程序设计中的递归树、数据库查询优化中的语法分析树以及数据挖掘中的决策树等。明确了所学课程对于“前驱”课程基点的升华和对于“后继”课程学习的支撑,就明确了所学课程对于专业人才培养的重要意义,从而就有助于从本质上不断增强和升华一种明确的学习理念,提升学习的内在驱动力量。

关联主导内容 学习理念和学习驱动力量的缺失还可能在于对整个学习内容茫无头绪,不明确课程“到底”讲什么、怎样讲,所涉及内容之间如何进行梳理,所学课程众多内容模块之间怎样形成“链接”。因此,本书在材料组织和内容展开等方面就特别注意突出这种关联或链接的知识建构。实际上,“数据结构”课程的要点是研究如何为数据对象建立4种逻辑组织结构:线性表、树、图和集合,其中前三者主要是基于内存的逻辑组织形式,而集合既可以基于内存也可以基于外存。由于外存中数据组织没有像内存中具有多种形式之分,因此外存中数据集合就统称为文件。对于数据集合而言,主要采用4种存储方式:顺序、链式、索引和散列,前两种是基于内存的,索引和散列主要基于外存(索引也可以基于内存,如基于内存的有二叉查找树,基于外存的有B⁺树)。基于线性表、树和图的数据操作形式多样,内容丰富,通常采用ADT进行统一描述;基于集合的数据操作相对单纯,对于内存数据来说,主要有排序和检索,对于外存来说,主要是检索。掌握了的知识应该是理解了的知识,理解了的知识实际上就是具有清晰关联和严谨结构的知识。

细节成就品质 文学作品的生命在于具体情节与生动细节。作为一门重要的专业基础课程,在明确大局、梳理关联的前提下,注重基本概念的实际区别和实现技术的关键细节就关乎到相应的学习品质,即通常所说的讲“透”或讲到“点子”上。实际上,“数据结构与算法”

作为众多前修课程的深化与发展,其自身品质在很大程度上都是凭借各种算法中精细构思和程序细节而实现的,如果剥离或弱化了其中具有闪光点的细节,那就不易与“离散数学”或“程序设计”课程相区分了。这些细节可分为概念内涵细节与技术实现细节。例如,概念内涵细节就有数据项与数据记录,数据元素与数据对象,数据类型与基本数据类型、抽象数据类型,数组与顺序表的联系与区别,头结点、头指针与表头结点,串和数组为什么通常采取顺序存储,二叉树与树的联系与区别,内存中的数据表与外存中的文件等;技术实现细节就有给定存储结构中数据插入/删除时相应指针(地址或下标)调换操作在程序设计语言(C或C++)中的具体描述和实现,栈在递归中所起到的存储工作记录的作用,循环队列在杨辉三角形显示中的作用,稀疏矩阵十字链表存储的行/列循环链表头结点与整个链表头结点,树的前序、中序和后序遍历与图的深度与广度优先遍历等。类似的细节在课程中相当丰富多彩,需要细心揣摩和认真梳理。当然,细节应当是相应内容模块的“点子”,讲细节的前提是要讲到“点子”上,需要围绕所涉及主题,通过细节将其“讲透”,注意避免枝蔓丛生,掩盖或冲淡主旨。

实例驱动兴趣 如同数学课程中定理的证明过程与实例的求解描述是其核心内容一样,数据结构中的算法细致讲解与实例验证说明也是本课程的主导和灵魂,抽掉了算法讲解与实例验证,就像一本数学教材只罗列定义与定理一样失去了意义。实际上,不仅算法和实例讲解部分蕴含着课程的重点,而且更可能是课程的难点。在对相关概念与方法基本了解的基础上,只有从直观上感知,才能从本质上把握。对于学生来说,一个基于程序语言实现的重要算法的细致讲解胜过对算法原理的抽象描述,一个好的实例验证更是胜过千言万语的抽象说教。对于老师来说,一个关键细节的程序语言描述和一个恰到好处的实例说明,也会带来教学热情的骤然高涨,智慧灵感的瞬间显现。课程中的实例讲解可以分为两个部分:一是实例中体现的算法(如栈在递归中的应用实例、队列在作业排队中的应用实例以及 Huffman 树等重要应用实例);再就是算法中的实例解释验证(如各种排序和查找算法等的基本实例)。对于前一种情形,需要强化问题引入的背景分析,注重相应算法的关键点展示;对于后一种情形,需要借鉴现有优秀教材的做法,比较详细介绍算法的图示解释,特别强调图示解释中的描述到程序设计语言的转换,同时教材中所有程序都是实际可运行的编码,以实现课程到(验证模仿型)实验的无缝连接。

本书在逻辑上可以分为两块:基于内存的数据结构和基于外存的文件结构。前者包括数据的线性结构(第2章线性表、第3章栈与队列、第4章数组、串与广义表)、数据的树形结构(第5章二叉树、第6章树与森林)、数据的图型结构(第7章图)和数据的集合结构(第8章查找、第9章排序);后者包括文件的组织与查找(第10章文件)。其中,第1~4章和第10章主要由陈瑛编写,其余各章主要由叶小平编写,全书由叶小平统稿。

在教材编写与课程实践中,笔者参考了国内外众多数据结构与算法方面的优秀教材,其中大多列举在书后的参考文献中。在此,我们对这些教材的编著者表示衷心感谢。另外,在教材编写过程中,虽然有如上所述的一些基本考虑,并试图将这些考虑体现在教材的内容组织与编写当中,但限于作者水平和其他条件约束,疏漏和不周之处在所难免,希望专家学者、选用此书的教师和同学不吝赐教。

清华大学出版社计算机与信息分社的领导和编审老师给予本书编写以很大的支持与细心指导,在此也表示衷心的感谢!

本书的程序都在常规平台(如 VC)上进行了验证运行。另外,本书还配有相应的学习与实验指导书供选用者参考。

本书得到华南师范大学增城学院教材立项资金资助(2011005)。

作 者

2012 年 5 月

目录

第 1 章 绪论	1
1.1 数据	1
1.1.1 数据的基本概念	1
1.1.2 数值型数据与非数值型数据	2
1.2 数据项与数据元素	2
1.3 数据类型与抽象数据类型	3
1.3.1 数据类型	3
1.3.2 抽象数据类型	4
1.4 数据模型与数据结构	6
1.4.1 数据逻辑结构	6
1.4.2 数据存储结构	7
1.5 数据操作与算法	8
1.5.1 数据运算	9
1.5.2 算法及其基本特征	9
1.5.3 算法设计与分析	11
1.6 数据结构课程的地位与本书内容体系	13
1.6.1 数据结构课程的地位	13
1.6.2 本书的内容组织	14
1.6.3 课程学习建议	15
本章小结	16
第 2 章 线性表	19
2.1 线性表的概念	19
2.1.1 线性表的逻辑结构	19
2.1.2 线性表的 ADT 描述	20

2.2	线性表的顺序存储	21
2.2.1	顺序存储结构	21
2.2.2	基于顺序存储的基本操作	23
2.3	线性表的链式存储	28
2.3.1	单链表的概念	28
2.3.2	单链表的基本操作	30
2.3.3	循环链表	35
2.3.4	双向链表	37
2.3.5	静态链表	39
2.3.6	单链表的应用	43
2.4	线性表存储结构的比较	48
	本章小结	49
第3章	栈与队列	51
3.1	栈	51
3.1.1	栈的基本概念	51
3.1.2	栈的顺序存储结构	52
3.1.3	栈的链式存储结构	55
3.1.4	栈的应用	57
3.2	队列	70
3.2.1	队列的基本概念	70
3.2.2	顺序队列与循环队列	71
3.2.3	队列的链式存储结构	75
3.2.4	队列的应用	77
	本章小结	79
第4章	数组、串与广义表	81
4.1	数组	81
4.1.1	二维数组	82
4.1.2	矩阵的顺序表示与实现	82
4.1.3	特殊矩阵的压缩存储	83
4.1.4	稀疏矩阵的压缩存储	86
4.2	串	93
4.2.1	串及其相关概念	94
4.2.2	串的基本操作	94
4.2.3	串的存储结构	95
4.2.4	串的模式匹配	97

4.3 广义表	104
4.3.1 广义表的基本概念	104
4.3.2 广义表的存储结构	106
4.3.3 广义表的基本操作	107
本章小结	110
第5章 二叉树	112
5.1 二叉树及其基本性质	112
5.1.1 二叉树的基本概念	112
5.1.2 满二叉树和完全二叉树	114
5.1.3 二叉树的基本性质	115
5.2 二叉树的存储	116
5.2.1 二叉树的顺序存储	117
5.2.2 二叉树的链式存储	118
5.3 二叉树的遍历	121
5.4 线索二叉树	131
5.4.1 线索与线索二叉树	131
5.4.2 创建线索二叉树	132
5.4.3 线索二叉树的遍历	135
5.5 二叉树的应用	136
5.5.1 等长编码与非等长编码	136
5.5.2 Huffman 树的构造思想	137
5.5.3 基于顺序存储的 Huffman 树构造	138
5.5.4 Huffman 编码	141
本章小结	144
第6章 树与森林	146
6.1 树及其相关概念	146
6.1.1 树	146
6.1.2 结点及其相关概念	148
6.2 树的存储结构	150
6.2.1 父结点表示法存储	150
6.2.2 子结点表示法存储	151
6.2.3 左子结点/右兄弟结点表示法存储	152
6.3 树的遍历	153
6.3.1 层次遍历	153
6.3.2 先序遍历	156

6.3.3 后序遍历	157
6.4 森林	157
6.5 树与二叉树的转换	158
6.5.1 树转换为二叉树	159
6.5.2 二叉树还原为树	159
6.5.3 森林与二叉树的转换	160
本章小结	161
第7章 图	163
7.1 基本概念与相关描述	163
7.1.1 图的基本概念	163
7.1.2 图的相关概念	164
7.2 图的存储	167
7.2.1 基于邻接矩阵的存储	167
7.2.2 基于邻接表的存储	169
7.3 图的遍历	172
7.3.1 深度优先遍历	173
7.3.2 广度优先遍历	174
7.3.3 简单路径与长度最短的路径	177
7.4 生成树与最小生成树	178
7.4.1 图的生成树	178
7.4.2 最小生成树	180
7.5 最短路径	187
7.6 有向无环图及应用	192
7.6.1 拓扑排序	192
7.6.2 关键路径	195
本章小结	202
第8章 查找	204
8.1 数据查找	204
8.2 基于线性表的查找	206
8.2.1 顺序查找	206
8.2.2 分块查找	207
8.2.3 二分查找	209
8.3 基于二叉树的查找	212
8.3.1 二叉查找树的概念	212
8.3.2 基于二叉查找树的查找	213

8.3.3	二叉查找树的插入与创建	214
8.3.4	二叉查找树的删除	216
8.3.5	平衡二叉树	218
8.4	基于散列表的查找	221
8.4.1	常用散列函数的构造	222
8.4.2	散列冲突处理	224
	本章小结	227
第9章	排序	229
9.1	数据排序	229
9.1.1	排序的基本概念与分类	229
9.1.2	排序算法的稳定性	230
9.2	插入排序	231
9.2.1	直接插入排序	231
9.2.2	二分插入排序	235
9.2.3	表插入排序	237
9.2.4	Shell 插入排序	239
9.3	交换排序	241
9.3.1	冒泡排序	241
9.3.2	快速排序	243
9.4	选择排序	247
9.4.1	直接选择排序	247
9.4.2	堆排序	248
9.5	归并排序	253
9.6	外排序	256
9.6.1	外排序的基本步骤	256
9.6.2	败者树的 k 路归并算法	258
9.6.3	k 路归并算法的实现	259
	本章小结	261
第10章	文件	266
10.1	文件的基本概念	266
10.1.1	文件的分类	266
10.1.2	文件的结构与操作	268
10.2	顺序文件	270
10.2.1	顺序文件的存储结构	270
10.2.2	基于磁带/磁盘的顺序存储	271

10.3	索引文件	271
10.3.1	索引的概念及操作	271
10.3.2	ISAM 文件	273
10.3.3	VSAM 文件	275
10.4	动态索引 B_ 树	277
10.4.1	B_ 树	277
10.4.2	B ⁺ 树	283
10.5	散列文件	284
10.6	多关键字文件	285
10.6.1	多重表文件	286
10.6.2	倒排文件	287
	本章小结	288
	参考文献	290

绪 论

第 1 章

1968 年,Knuth 在其所著的《计算机程序设计艺术》第一卷“基本算法”中较为系统地阐述了数据的逻辑结构和存储结构及其操作,开创了数据结构的学科体系。同年,数据结构作为一门独立课程开始出现在高校计算机专业课程设置当中。使用计算机处理数据的前提是将数据“组织”起来,数据的“组织”形式可以看做数据模型或数据“结构”。计算机处理的数据通常可分为“数值”型和“非数值”型两种基本类型。数值型数据的数据模型多是数学公式或数学方程;非数值型数据的数据模型主要是线性、树和图型结构。本书将要学习的数据结构课程主要研究非数值型数据模型或结构以及建立在其上相应数据操作。数据结构是计算机学科的最重要的核心课程之一,它是设计和实现各种应用软件的必备基础。

在学习过程中,需要注意以下问题:

- 数据、数据元素和数据项的基本概念。
- 数据类型和抽象数据类型的基本概念及其联系与区别。
- 数据结构的基本概念,数据逻辑结构的 4 种基本类型,数据存储结构的两种基本类型。
- 理解数据存储结构除了存储数据元素本身,更重要的是要存储数据元素之间的关系。
- 算法的基本概念与特性,如何评价算法的效率。
- 认识和理解数据结构在计算机学科中的地位与意义。

1.1 数据

理解数据结构的意义和作用的前提是掌握数据基本概念与相关特征。

1.1.1 数据的基本概念

数据是一个“元概念”,不能给出严格意义上的定义。对于各种元概念,从实际应用角度考虑,通常都是基于其若干基本特征进行“描述性”的界定。

从广义上来讲,数据可以看做是客观对象本身或其某一特征的一种“符号化”表示与描述,或者说,数据是反映客观对象的符号集合。需要注意,不是所有符号集合都能进入计算机并由计算机进行处理的。因此,从狭义上来讲,能够进入计算机并能由计算机进行处理的符号集合才是计算机意义下的数据。在本课程以及其他计算机领域,人们通常多是在这种狭义意义之下描述和使用数据的。数据可以有多种表现形式,例如,可以是数值、字符和字符串等,也可以是经过编码的图像、声音、视频和多媒体等。实际上,数据概念本身是随着计算机发展而不断扩展的概念。在发展初期,由于计算机主要用于数值计算,数据主要指的就是整数、实数等数值型数据;在计算机用于文字处理时,数据主要指的是由英文字母(包括汉字等)组成的字符串数据;随着计算机硬件和软件技术的不断发展,扩大了计算机的应用领域,诸如经过编码处理的表格、图形、图像、声音等也都属于数据的范畴。

下面从狭义角度给出(计算机)数据的基本描述。

数据(Data)是对客观事物的描述,是信息的载体,是能够被计算机识别、存储和加工处理的各种形式的总和。数据是计算机程序加工的原料,通常需要进行解释才能明确其意义,即从数据中获取信息。例如,“37”是数据,此时的数据只是描述,在没有特定的背景之下,人们通常难以确定“37”是“年份”还是“年龄”,是人的摄氏体温还是某两个地点之间的距离。即是说,单独的“37”只是数据,而通常只有赋予其意义之后才能称为“信息”,这也可以看做是“数据”与“信息”这两个元概念的区别与联系。

1.1.2 数值型数据与非数值型数据

从数据处理角度来看,计算机处理的数据可分为数值型数据和非数值型数据两种情形。

(1) 数值型数据:主要包括整数、实数和复数等可以“直接”计算的数据,这些数据可以通过数制转换“直接”进入计算机。计算机可以对数值型数据进行加、减、乘、除等运算,主要应用于计算方法、工程计算和商务处理等领域中的数值计算问题,例如,利用数值分析方法解代数方程的程序的处理对象都是整数和实数。数值型数据的基本特征是数据量通常不大,数据管理较为简单,但相应处理方法却比较复杂。研究通过计算机进行数值型数据处理的主要是数值计算(计算方法)等课程。

(2) 非数值型数据:主要包括字符、文字、图像、音频和视频等,这些数据需要经过各种比较复杂的编码方式才能进入计算机,计算机系统对这些数据主要不是进行“直接计算”,而是着眼于它们之间的各种关系进行存储和处理,其主要应用于文字处理、多媒体管理以及更为复杂的逻辑推理等领域中的数据的组织、管理和查找等与“直接计算”无关的问题。非数值型数据的基本特征是数据处理方法相对简单,而数据间关系的计算机存储与管理却相当复杂,涉及的数据量也极为庞大(海量数据)。研究这些问题的计算机学科就是即将学习的“数据结构”。数据结构课程的主要任务就是研究程序设计过程中计算机的操作对象(数据)的组织结构以及对这些结构化数据操作的实现,而不是单纯的数据值的具体计算。

1.2 数据项与数据元素

在数据结构框架内,为了更有效描述和使用数据概念,从技术角度考虑还需要引入数据项、数据元素和数据对象等相关概念。