

高等专科学校教材

数据结构

潘道才 编

成都电讯工程学院出版社

高等专科学 校教材

数 据 结 构

潘道才 编

成都电讯工程学院出版社

· 1988

内 容 提 要

本书是由电子工业部“大专计算机专业教材编审委员会专业基础课编审小组”组织评选和推荐出版的教材。

本书介绍了各种数据结构的基本概念、逻辑特性和存储结构，对各种结构定义了相应的运算。书中以框图形式用自然语言描述算法，清晰易懂，利于学习。对大多数算法还给出了相应的PASCAL程序，具有实用意义。各章都附有习题和上机题目，便于教学和实践。

本书可作为大专类计算机类、硬件专业的教材，也可作为使用计算机的广大科技人员自学或参考。

高等专科学校教材

数 据 结 构

潘道才 编

※

成都电讯工程学院出版社出版

成都市师范学校印刷厂印刷

四川省新华书店经销

※

开本 850×1168 1/32 印张 12 字数 300千字

版次 1988年4月第一版 印次 1988年4月第一次印刷

印数：1—7,000册

中国标准书号：ISBN7-81016-022-2/TP·3

(15452·30)

定价：3.60元

出 版 说 明

根据国务院关于高等学校教材工作分工的规定，我部承担了全国高等学校、中等专业学校工科电子类专业教材的编审、出版的组织工作。由于各有关院校及参与编审工作的广大教师共同努力，有关出版社的紧密配合，从1978年至1985年，已编审、出版了两轮教材，正在陆续供给高等学校和中等专业学校教学使用。

为了使工科电子类专业教材能更好地适应“三个面向”的需要，贯彻“努力提高教材质量，逐步实现教材多样化，增加不同品种、不同层次、不同学术观点、不同风格、不同改革试验的教材”的精神，我部所属的七个高等学校教材编审委员会和两个中等专业学校教材编审委员会，在总结前两轮教材工作的基础上，结合教育形势的发展和教学改革的需要，制订了1986~1990年的“七五”（第三轮）教材编审出版规划。列入规划的教材、实验教材、教学参考书等近400种选题。这批教材的评选推荐和编写工作由各编委会直接组织进行。

这批教材的书稿，是从通过教学实践、师生反映较好的讲议中经院校推荐，由编审委员会（小组）评选择优产生出来的。广大编审者、各编审委员会和有关出版社为保证教材的出版和提高教材的质量，作出了不懈的努力。

限于水平和经验，这批教材的编审、出版工作还会有缺点和不足之处，希望使用教材的单位，广大教师和同学积极提出批评建议，共同为不断提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

前 言

本教材系按电子工业部制定的工科电子类专业教材 1986~1990年编审出版规划,由大专计算机专业教材编审委员会专业基础课编审小组组织征稿、评选、推荐出版的。

本教材由上海科技专科学校潘道才副教授担任主编,西北电讯工程学院朱儒荣副教授担任主审。

本课程的参考学时数为72学时,其主要内容为数据的逻辑结构、物理结构以及对每种结构所定义的运算及应用。其中包括线性表、串、数组、广义表、树型结构、图结构、文件结构以及它们的存储。对于同一种数据结构,讨论了不同的存储结构及相应的算法和PASCAL程序。本教材偏重于数据结构的应用,所以除在章节中通过实例介绍数据结构的应用外,还专门讨论了查找和排序的方法,在第九章则举例讨论了综合应用数据结构的例子。在每章中都为学生提供了一定份量的习题并标明了应上机调试的题目。

本教材按电子工业部电子工业职工高等院校“计算机软件及其应用专业”的教学大纲进行编号。考虑到数据结构的发展及全日制高等专科学校的需要,适当的增加了一些内容。例如,汉字串、B树、广义表等。对于计算机硬件专业,可以适当减少一些内容,建议删减的章节前面加了“•”号。本教材考虑了专科教育的针对性、实践性和应用性,因此,可以作为高等专科学校、职工大学、职业大学、夜大学、函授大学等高等院校的大专类计算机软、硬件专业的教材或参考书。书中采用自然语言描述算法,给出框图,通俗易懂,所以也适于使用计算机的广大科技人员自学。

参加本教材审阅工作的还有南京大学金志权副教授、上海科技大学袁荣喜副教授，他们为本书提出了许多宝贵意见，值此表示诚挚的感谢。由于编者水平有限，书中难免还存在一些缺点和错误，殷切希望广大读者批评指正。

编 者

目 录

第一章 绪 论	1
1.1 什么是数据结构	1
1.2 数据结构发展概况	3
1.3 关于算法描述和分析的说明	4
习题	6
第二章 线性表及其应用	7
2.1 基本概念和常用术语	7
2.2 线性表的存储结构(一)——向量	9
2.2.1 顺序分配	10
2.2.2 向量的插入和删除	11
2.3 线性表的存储结构(二)——线性链表	15
2.3.1 链式分配	15
2.3.2 线性链表的插入和删除	17
2.4 栈和队列	22
2.4.1 栈及其存储结构	23
2.4.2 栈的应用	28
2.4.3 队列及其存储结构	33
2.5 循环线性链表和双向链表	42
2.5.1 循环线性链表	42
2.5.2 双向链表和循环双向链表	44
2.6 多项式相加问题	49
习题	55
第三章 串	58
3.1 串的定义和特性	58
3.2 串的存储结构	59

3.2.1 串的顺序存储结构.....	59
3.2.2 串的链式存储结构.....	62
3.2.3 串变量的存储.....	63
3.3 串的运算.....	64
3.3.1 求串的长度.....	65
3.3.2 串的联接.....	66
3.3.3 求子串.....	67
3.3.4 搜索子串在串中的序号.....	69
3.3.5 插入子串和删除子串.....	72
3.3.6 串的置换.....	75
3.4 文本编辑.....	77
3.5 汉字串.....	79
习题	85
第四章 数组和广义表	87
4.1 数组的顺序分配.....	87
4.2 稀疏矩阵的三元组表示法.....	92
4.3 稀疏矩阵的十字链表表示法	100
4.4 矩阵相加运算	105
4.5 迷宫问题	107
4.6 广义表	112
习题.....	115
第五章 树型结构及其应用.....	118
5.1 树型结构的基本概念	118
5.2 树的存储结构	121
5.3 二叉树	124
5.3.1 二叉树的定义和性质.....	125
5.3.2 二叉树的存储结构.....	128
5.4 二叉树与树、森林之间的转换	130
5.4.1 二叉树与树之间的转换.....	130

5.4.2	二叉树与森林之间的转换	133
5.5	遍历二叉树	135
5.5.1	前根遍历	136
5.5.2	中根遍历	137
5.5.3	后根遍历	141
5.6	线索树	144
5.6.1	建立线索树	145
5.6.2	检索结点	149
5.6.3	插入结点	153
5.7	树的应用	156
5.7.1	二叉排序树	156
5.7.2	哈夫曼树及其应用	164
5.7.3	判定树	174
	习题	177
第六章	图结构及其应用	180
6.1	图结构的基本概念	180
6.2	图的存储结构	184
6.2.1	邻接矩阵	184
6.2.2	邻接链表	186
6.3	遍历图及求图的连通分量	188
6.3.1	深度优先搜索法	188
6.3.2	广度优先搜索法	191
6.3.3	求图的连通分量	193
6.4	生成树	194
6.4.1	生成树的概念	194
6.4.2	最小生成树	195
6.5	最短路径	199
6.5.1	从某个源点到其余各顶点的最短路径	200
6.5.2	每一对顶点间的最短路径	202
6.6	有向无回图及其应用	206

6.6.1 有向无回路图.....	208
6.6.2 拓扑排序.....	210
*6.6.3 关键路径.....	218
习题.....	224
第七章 查 找	228
7.1 顺序表的查找	228
7.1.1 顺序查找.....	228
7.1.2 折半查找.....	231
7.2 树表查找	235
7.2.1 二叉查找树和平衡树.....	236
*7.2.2 B 树.....	242
7.3 哈希表及其查找	244
7.3.1 什么是哈希表.....	244
7.3.2 构造哈希函数的基本方法.....	246
7.3.3 解决冲突的几种方法.....	249
习题.....	258
第八章 排 序	260
8.1 插入排序	261
8.1.1 直接插入排序.....	261
8.1.2 希尔排序.....	264
8.2 交换排序	268
8.2.1 冒泡排序.....	268
8.2.2 快速排序.....	270
8.3 选择排序	274
8.3.1 直接选择排序.....	274
8.3.2 堆排序.....	277
8.4 归并排序	285
*8.5 基数排序	290
8.6 外排序	297

8.6.1 外存信息的特性·····	297
8.6.2 文件及其组织·····	299
8.6.3 外排序的基本方法·····	302
习题·····	305
第九章 数据结构应用实例·····	307
9.1 存储管理·····	307
9.1.1 问题的提出·····	307
9.1.2 动态存储分配和回收·····	309
9.1.3 不用单元收集和紧凑存储·····	317
9.2 银行业务活动的模拟·····	319
9.3 学生成绩管理·····	324
9.3.1 数据结构的描述·····	324
9.3.2 各子程序的功能和实现·····	326
习题·····	341
附录·····	343

第一章 绪 论

“数据结构”是一门随着计算机科学的发展而逐渐形成的新兴学科。

早期的电子计算机主要用于数值计算。后来，随着计算机软、硬件的发展，计算机的应用范围越来越广泛。其加工处理的对象也就从简单的数字、字符发展到具有一定结构关系的数据。为了更有效地使用计算机，设计出高效率、高可靠性的准确程序，需要对数据的组织、数据内部的结构关系，以及在计算机内如何表示和操作进行深入的研究。这样，便促进了“数据结构”学科的发展。

1.1 什么是数据结构

什么是数据结构 (data structures)？在计算机科学中，数据是计算机程序加工处理的对象。抽象地说，数据是对客观事物所进行的描述，而这种描述是采用了计算机系统能够识别、存储和处理的形式来进行的。这种描述一般可以归结为数字，字符或符号的集合。我们把组成数据的基本单位称为数据元素，所以，数据是数据元素的集合。例如，一个计算几何图形面积的程序，其加工处理的数据是实数和整数；一个编译程序所处理的数据是源程序。

数据结构中要研究的数据不是一、二个孤立的数据，而是大量的相互关联的数据。数据元素之间存在的相互关系就叫做结构。数据元素之间抽象化的相互关系称为数据的逻辑结构。这种相互关系可用一组运算及其规则来描述。通常，把这种逻辑结构

简称为数据结构。例如，有一本电话号码簿，它记录了 n 个用户的名字和相应的电话号码。若把这种数据设计成一维数组。它的每个数据元素是一个数对 (a_i, b_i) 。其中， a_i 和 $b_i (1 \leq i \leq n)$ ，分别表示某用户的名字和对应的电话号码。则有一维数组：

$$[(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n)]$$

对于一维数组，我们可以定义一些运算及相应的运算规则。如进行插入或删除某个数据元素等。这样，就构成了上述电话号码的数据结构。

在设计算法和程序时，不仅要考虑数据的逻辑结构及其运算，而且要考虑这种结构在计算机存储器中的存储方式。这就是数据的物理结构或称存储结构。一种逻辑结构可以通过映象得到与它相应的存储结构。本书中将要讨论的逻辑结构有线性的，包括栈、队列和串等；有非线性的，包括树和图结构。它们分别通过顺序和非顺序映象，可以得到不同的存储结构。

在数据结构中，往往涉及数据类型和数据对象的概念，它们之间是不同的。

数据类型是指某种程序设计语言中所允许使用的变量种类。如 FORTRAN 语言中，变量的数据类型有整型、实型、布尔型、双精度型、复型等。在 PASCAL 语言中，除了有标准的整型、实型、布尔型、字符型，以及由用户自己定义说明的纯量类型，还允许使用由用户定义的构造类型。一个数据类型不仅定义了一个变量可以设定的值的集合，而且还规定了对变量允许进行的一组运算和运算规则。例如，PASCAL 语言中的布尔型变量的值是 TRUE（真）和 FALSE（假），允许对布尔型变量进行的运算则有 AND（与）、OR（或）、NOT（非）等。

数据对象是指某种数据类型的数据元素的集合，是数据的一个子集。如整数的数据对象是集合 $D = \{0, \pm 1, \pm 2, \dots\}$ ，字母字

符的数据对象是集合 $C = \{A, B, \dots, Z\}$ 。

由此可见，数据结构、数据类型、数据对象是三个不同的概念。

1.2 数据结构发展概况

六十年代初期，国内外的大学中，都没有独立的“数据结构”课程。但数据结构的有关内容已散见于操作系统、编译原理和表处理语言等课程之中。1968年，美国一些大学的计算机科学系的教育计划中，明确规定“数据结构”为一门课程。但仍没有明确规定该课程的内容范围。当时，数据结构几乎和图论，特别是表和树的理论是同义语。随后，数据结构这个概念被扩充到包括网络、代数、集合论、关系等现在称之谓“离散数学结构”的那些内容。它与现在的数据结构的某些内容混在一起，总称为“数据结构”。由于数据必须在计算机中进行处理，因此，不能局限于研究数据本身的数学概念，还必须考虑数据的物理结构。这就进一步扩大了数据结构的内容。自从1968年，美国研究计算机科学的著名教授D.E.Knuth所著的“计算机程序设计技巧”（The Art of Computer Programming）问世以后，逐渐将数据的有关数学概念独立出来，形成了现在的“离散数学结构”，而把数据的逻辑结构、物理结构以及每种结构所定义的运算组成了“数据结构”课程的主要内容。近年来，由于数据库系统的不断发展，在数据结构课程中又增加了文件管理，特别是大型文件的组织等方面的内容。

数据结构与数学、计算机硬件、特别是计算机软件有着密切的关系。它是计算机专业的一门核心课程，是编译原理、操作系统、数据库、人工智能等课程的基础。同时又广泛应用于信息科学、系统工程、应用数学以及各种工程技术领域。

数据结构在计算机科学中有着十分重要的地位。它有自己的理论、研究对象和应用范围、而且其研究的内容正在不断扩充和深化。因此, 作为一门课程、一本教材, 因受特定的对象和时期的限制, 而不能全面地反映数据结构的全貌。但值得注意的是, 作为一门学科, 数据结构是新兴的, 正值方兴未艾, 蓬勃发展的阶段。

1.3 关于算法描述和分析的说明

数据结构是一门实践性强的学科。通过该课程的学习, 应使学生能运用数据结构的技巧更好地进行算法和程序设计。所以, 我们在讨论各种数据结构的基本运算时, 都给出了相应的算法。对于算法的描述, 我们力求做到通俗易懂和适于自学, 所以采用文字框图进行图示。读者在掌握和理解了框图所示的设计思想后, 可以较方便地使用自己熟悉的算法语言来编制源程序。另外, 考虑到用PASCAL语言来编写程序可以较好地体现结构程序设计的一些原则, 并且简明易懂, 容易学习, 且具有实用价值, 所以本书中大部分算法在给出文字框图的同时还给出了用PASCAL语言编写的源程序片断。其中包括常量说明、类型说明以及用过程或函数形式给出的子程序。但对一些较长的程序则收在附录中。(源程序中出现的 $\leq$ 即 $\leq$, $\neq$ 即 $\neq$, $\geq$ 即 $\geq$)。书中虽然给出了程序片断, 但程序本身已在计算机上调试通过。限于时间和作者的水平, 算法的实现并不一定精炼。

关于PASCAL语言的内容, 读者可以参阅有关书籍, 这里不作介绍。本书框图中使用的图形符号见图1.1。

1. 椭圆框: 框中用文字标明算法的“开始”或“结束”。

2. 矩形框: 框内描述某些操作, 如赋值、组织循环、输入和输出等。统称为操作框。

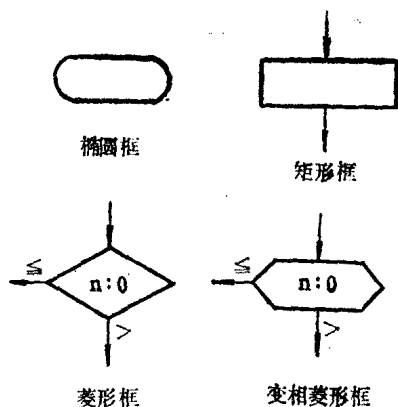


图 1.1 框图所用图形

3. 菱形框（包括变相菱形框）：是判别框。框中符号“:”表示比较。例如 $n:0$ 表示 n 与 0 相比较，比较的结果写于框外连接线条的旁边。带箭头的线条表示算法或程序的走向，写于其旁的判断结果就是算法或程序的分支走向应满足的条件。

另外，对算法（或程序）的分析和评价通常较复杂。一般需考虑正确性、简单性、最优性、运算量及占用存储量等诸多因素。为了简化讨论，我们仅采用其中的两条标准。一是用问题的某个参数的函数来估算其存储量；其二是数据运算研究所需的计算量。假设问题的参数为 n 。此参数可以是矩阵的阶、线性表的长度、图的顶点数等显示该问题规模大小的参数。那末，所选数据结构上执行某种操作所需要的存储量及计算量是 n 的什么函数呢？我们引进记号‘O’，对这些函数作数量级的估算。如，对于下述三个简单程序段：

- (1) $x := x + 1;$
- (2) FOR $i := 1$ TO n DO $x := x + 1;$
- (3) FOR $i := 1$ TO n DO
 FOR $j := 1$ TO n DO $x := x + 1;$

设程序段1中, 语句 $x := x + 1$ 不包含在任何循环体之中, 则此语句只执行一次, 计算量可记为 $O(1)$ 。在程序段2中, 上述赋值语句在FOR循环之中, 所以要执行 n 次, 其执行时间和 n 成正比, 计算量可记为 $O(n)$ 。在程序段3中, $x := x + 1$ 语句要执行 $n \times n$ 次, 其执行时间和 n^2 成正比, 故计算量可记为 $O(n^2)$ 。对于算法或程序所需的存储量, 也可以作类似分析。然而, 要事先对一个算法的计算量作仔细的分析很复杂, 而且也不是本课程的主要内容, 所以书中对一些算法的性能评价只根据算法中执行次数最多的语句来估算该算法的计算量的数量级。

习 题

1. 举例说明什么叫数据、数据结构和数据类型?
2. 数据结构研究的主要内容是什么?
3. 分析下列程序的计算量

```
PROGRAM EX;
  VAR i, j, p: integer;
  BEGIN
    FOR i:=1 TO 9 DO write(i:4);
    FOR i:=1 TO 2 DO writeln;
    FOR i:=1 TO 9 DO
      BEGIN
        FOR j:=1 TO i DO
          BEGIN
            p:=i*j;
            write(p:4);
          END;
        FOR j:=1 TO 2 DO writeln;
      END
    END
  END
```