

21 SHIJIGAOENGZHIYEJIAOYU
GUIHUAJIAOCAI

数据结构



21世纪高等职业教育规划教材（2年制）

SHUJUGIEGOU

● 主编 王元安



▲ 中国财政经济出版社

21 世纪高等职业教育规划教材 (2 年制)

数 据 结 构

主编 王元安

审稿 张学云

中国财政经济出版社

图书在版编目 (CIP) 数据

数据结构/王元安主编. —北京: 中国财政经济出版社, 2005.8

21 世纪高等职业教育规划教材. 2 年制

ISBN 7-5005-8469-5

I. 数… II. 王… III. 数据结构-高等学校: 技术学校-教材 IV.
TP311.12

中国版本图书馆 CIP 数据核字 (2005) 第 085629 号

中国财政经济出版社出版

URL: <http://www.cfeph.cn>

E-mail: cfeph@cfeph.cn

(版权所有 翻印必究)

社址: 北京市海淀区阜成路甲 28 号 邮政编码: 100036

发行电话: (010) 88190616/54 88190655 (传真)

北京财经印刷厂印刷 各地新华书店经销

787×960 毫米 16 开 15.75 印张 244 000 字

2005 年 12 月第 1 版 2005 年 12 月北京第 1 次印刷

定价: 19.00 元

ISBN 7-5005-8469-5/TP·0114

(图书出现印装问题, 本社负责调换)

本教材的正版图书封底上贴有“中国财政经济出版社 教育分社”防伪标识。根据标识上提供的查询网站、查询电话和查询短信, 输入揭开防伪标识后显示的产品数字编号, 即可查询本书是否为正版图书。版权所有, 翻印必究, 欢迎读者举报。举报电话: 010—88190654。

出版说明

为了进一步贯彻落实《国务院关于大力推进职业教育改革与发展的决定》和全国职业教育工作会议的精神,适应二年制高等职业教育发展的趋势,满足各类职业技术学院专业教学的实际需要,我们组织编写了21世纪高等职业教育规划教材。该系列教材涵盖了二年制高等职业教育教学中所需的公共课(包括文化基础课、思想政治课)、财务会计、市场营销、电子商务、金融与证券、国际贸易、旅游饭店与管理、文秘等专业主干课程,从2005年秋季开学起,这些教材将陆续提供给各类职业技术学院使用。

该系列教材是根据教育部提出的“以综合素质培养为基础,以能力培养为主线”为指导思想,结合二年制高等职业教育的教学培养目标而编写的。新教材全面贯彻素质教育思想,从社会发展对高技术应用性人才的需求出发,在内容的构建上结合专业岗位(群)对职业能力的需要来确定教材的知识点、技能点和素质要求点,并注重新知识、新技术、新工艺、新方法的应用,注重对学生的创新精神和实践能力的培养。新教材在理论体系、组织结构和阐述方法等方面均作了一些新的尝试,以适应高等职业教育教学改革,满足各类高等职业技术学院教学需要。在此,我们真诚的希望各类职业技术学院在教材的使用过程中,能够总结经验,及时提出修改意见和建议,使之不断完善和提高。

前言

对于高等职业院校计算机应用类专业的学生而言,“数据结构”是计算机专业的基础课程。其主要内容有:数据结构和算法设计与分析的基本知识,各种基本数据结构的定义、存储结构、相应的算法以及应用,基本的数据结构与算法的关系等。开设这门课程是为了培养计算机专业的学生结合实际情况,设计有效的算法和数据结构的能力,并为其后续课程的学习以及软件设计水平的提高打下良好的基础。

为了适应国家对职业技术教育的最新要求,本教材在内容组织和编排上力求体现“先理论,后应用,理论与应用相结合”的原则,强调理论够用为度,侧重基本概念的讲解和相应算法的运用,通过精选的概念、例题和上机实例,综合运用课堂讲授、在多媒体教室中采用电子教案授课以及上课时边讲边演示等教学方法与教学手段,使学生由浅入深、循序渐进地对所学知识加以融会贯通。与本书配套的辅导教材为中国财政经济出版社出版的《数据结构习题与上机实验》一书,其中配有学生综合上机实验和综合分阶段分单元的测试题。

教学内容及学时分配

课程内容	教学 要求	重点 (☆)	难点 (Δ)	学时 安排
第 1 章 绪言	A	算法的概念、描述方法以及评价标准	算法评价标准	4
第 2 章 线性表	A	线性表的顺序存储结构、线性表的链式存储结构	循环链表	10 (上机 2)
第 3 章 栈和队列	A	栈的存储结构及其基本操作、队列存储结构及其基本操作	模式匹配	6 (上机 2)
第 4 章 串	A	全面领会串的概念及运算, 掌握串的存储结构及对应算法的实现	串模式匹配运算	6 (上机 2)
第 5 章 数组与广义表	B	特殊矩阵的压缩存储 广义表的定义、术语、存储结构、运算	十字链表	6 (上机 2)
第 6 章 树与二叉树	A	二叉树的遍历、线索算法、哈夫曼树及其应用	线索算法、哈夫曼树及其应用	6 (上机 2)
第 7 章 图	B	图的存储结构 图的遍历操作	图的几个典型应用问题	8 (上机 2)
第 8 章 查找	A	顺序查找、折半查找动态查找表及查找算法: 二叉排序树	哈希表及查找算法、二叉排序树	8 (上机 2)
第 9 章 排序	A	快速排序、堆排序、归并排序、希尔排序	快速排序、堆排序	8 (上机 2)
第 10 章 文件	C	文件的基本概念 文件的各种组织方式和操作的实现	文件在外存储器中的表示及组织方式	6 (上机 2)
机动及考试				4
合 计				72

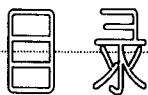
上表中的教学要求：A - 熟练掌握；B - 掌握；C - 了解。本课程总共 72 学时，其中讲授 50 学时，上机实习 18 学时，机动与考试 4 学时。

在本书编写中，李沁蓉编写了本书的第 2 章和第 3 章，丁应远编写了第 4 章、第 8 章和第 9 章，贾步忠编写了第 6 章和第 7 章，林世琼编写了第 10 章，王元安任主编，编写了第 1 章与第 5 章并负责全书的统稿和定稿工作。

由于时间仓促，编写任务重，书中难免有不妥和谬误之处，敬请读者和专家予以谅解指正。

编 者

2005 年 8 月



第1章 绪 论	(1)
1.1 数据结构的概念	(2)
1.2 算法的描述	(11)
小 结.....	(17)
第2章 线性表	(18)
2.1 线性表的定义和基本运算	(19)
2.2 线性表的顺序存储结构	(21)
2.3 线性表的链式存储结构	(27)
2.4 线性表的综合应用与上机演示程序	(40)
小 结.....	(44)
第3章 栈和队列	(47)
3.1 栈	(48)
3.2 队列	(55)
3.3 栈和队列的综合应用与上机演示程序	(61)
小 结.....	(68)
第4章 串	(69)
4.1 串的基本概念	(70)
4.2 串的定长顺序存储及基本运算	(73)
4.3 串的堆分配存储结构表示	(78)

4.4	串的块链存储结构	(82)
4.5	串操作应用举例	(84)
小 结		(85)
第5章	数组与广义表	(87)
5.1	多维数组	(88)
5.2	矩阵的压缩存储	(92)
5.3	广义表	(103)
小 结		(108)
第6章	树与二叉树	(111)
6.1	树的概念	(112)
6.2	二叉树	(115)
6.3	树、森林与二叉树的关系	(128)
6.4	哈夫曼树	(135)
小 结		(137)
第7章	图	(139)
7.1	图的概念	(140)
7.2	图的存储结构	(144)
7.3	图的遍历	(150)
7.4	图的生成树	(155)
7.5	最短路径	(158)
7.6	拓扑排序	(161)
小 结		(164)
第8章	查找	(166)
8.1	基本概念与术语	(167)
8.2	静态查找表	(170)
8.3	动态查找表	(179)
8.4	哈希表查找	(185)
小 结		(194)

第 9 章 排序	(195)
9.1 基本概念	(196)
9.2 插入排序	(197)
9.3 交换排序	(200)
9.4 选择排序	(205)
9.5 二路归并排序	(211)
9.6 各种排序方法的比较	(215)
小 结	(216)
 第 10 章 文件	(217)
10.1 文件的基本概念	(218)
10.2 顺序文件	(222)
10.3 索引文件	(224)
10.4 索引顺序文件	(228)
10.5 直接存取文件	(233)
10.6 多关键字文件	(236)
小 结	(239)

第1章

绪 论

主要内容

- (1) 数据和数据结构等名词和术语。
- (2) 算法描述的方法及从时间和空间角度分析算法的方法。

本章要点

(1) 熟悉各名词和术语的含义；掌握各种基本概念，特别是数据结构的三个方面（逻辑结构、存储结构和数据运算）以及这三个方面的相互关系；熟悉数据结构的四类基本结构（集合结构、线性结构、树形结构、图状结构或网状结构）和四种基本的存储方法（顺序存储方法、链式存储方法、索引存储方法、散列存储方法）。

(2) 算法描述的方法（本教材采用 C 语言）及算法与程序的区别。

(3) 了解算法的五大要素（输入、输出、有穷性、确定性、可行性）。

(4) 了解估算算法的时间和空间复杂度的方法。

1.1

数据结构的概念

1.1.1 数据类型和数据元素

数据 (Data) 是信息的载体。它能够被计算机识别、存储和加工处理, 是计算机程序加工的“原料”。数据按其性质可分为以下几种:

- (1) 数值型数据, 如描述人的身高、体重、年龄等。
- (2) 字符型数据, 如描述人的姓名、住址、性别等。
- (3) 声音数据。
- (4) 图像数据。

数据类型 (Data Type) 是程序设计语言中所允许的变量的种类, 也是变量可取的值和可以进行运算的集合。数据类型又分为基本数据类型和构造数据类型两大类, 如 C 语言中有五种基本数据类型, 分别是字符 (char)、整数 (int)、浮点数 (float)、双精度浮点数 (double) 和无值 (void)。C 语言中的其他数据类型, 如数组、结构体、联合等, 都是以这些基本数据类型为基础构造而成的。

数据元素 (Data Element) 是数据的基本单位。数据元素也称元素、结点、顶点、记录。通常, 一个数据元素可以由若干个基本数据项 (也可称为字段、域、属性) 组成, 数据项是具有独立含义的不可再分的最小标识单位。在数据结构中, 通常将一个数据元素作为一个整体进行考虑和处理。

1.1.2 什么是数据结构

数据结构 (Data Structure) 是指数据之间的相互关系, 即数据的组织形式。

1. 数据结构一般包括以下三方面内容:

- (1) 数据元素之间的逻辑关系, 也称**数据的逻辑结构**。
- (2) 数据元素及其关系在计算机存储器内的表示, 称为**数据的存储结**

构。

(3) 数据的运算，即对数据施加的操作。

我们可以将数据结构定义为：按某种逻辑关系组织起来的一批数据，按一定的存储方式把它们存储在计算机的存储器中，并在这些数据上定义了一个运算的集合，就叫一个数据结构。简言之，数据结构 = {逻辑关系 + 存储方式 + 运算}。

2. 数据的逻辑结构

数据的逻辑结构是从逻辑关系上描述数据，与数据的存储无关，是独立于计算机的。数据的逻辑结构可以看做是从具体问题抽象出来的数学模型。

为了增加对数据结构的感性认识，下面举例来说明有关数据结构的概念。

【例 1.1】 某班学生成绩表，如表 1-1 所示。

表 1-1 学 生 成 绩 表

学号	姓名	数学	计算机基础	总分
1000	张三	70	80	150
1001	李四	90	65	155
1002	王二	80	90	170
1003	王五	40	60	100
.....				

我们把表 1-1 称为一个数据结构，表中的每一行是一个数据元素（或记录、结点），它由学号、姓名、各科成绩及总分成绩等数据项组成。在该表中，数据元素之间是一种先后的位置关系，对于表中任一个结点，与它相邻且在它前面的结点（亦称为直接前趋）最多只有一个；与表中任一结点相邻且在其后的结点（亦称为直接后继）也最多只有一个。表中只有第一个结点没有直接前趋，故称为开始结点；也只有最后一个结点没有直接后继，故称之为终端结点。这样一种关系我们称之为“线性结构”，如图 1-1 所示。

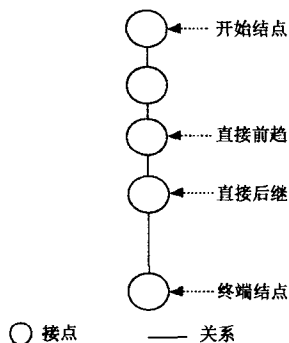


图 1-1 线性结构

【例 1.2】 公司机构组织图，如图 1-2 所示。

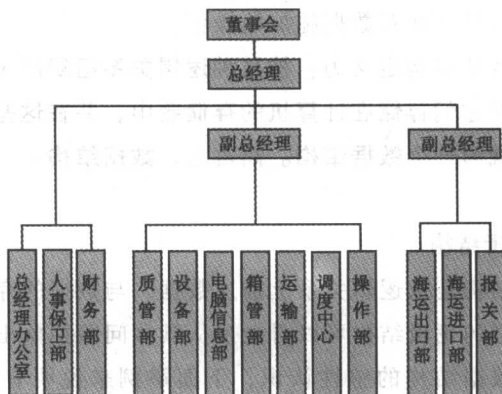


图 1-2 公司机构组织图

在图 1-2 中，我们将每一个机构看做是一个数据元素，则数据元素之间的关系要比表 1-1 要复杂许多。数据元素之间的关系是一对多的关系，即一个数据元素向上和一个元素相连（称为父结点），向下和多个数据元素相关（称为子结点），最上面的结点没有父结点称为根结点，最下层的结点没有子结点称为叶子。我们将这种关系称为“树结构”，如图 1-3 所示。

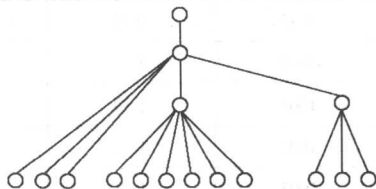


图 1-3 树结构

在不产生混淆的前提下，常将数据的逻辑结构简称为数据结构。根据数据元素之间的不同特性，通常有下列四类基本的结构，如图 1-4 所示。

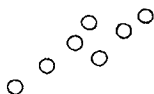
(1) **集合结构**：结构中的数据元素之间除了同属于一个集合的关系外，无任何其他关系。

(2) **线性结构**：结构中的元素存在着一对一的线性关系。若结构是非空集，则有且仅有一个开始结点和一个终端结点，并且所有结点都最多只有一个直接前趋和一个直接后继。线性表是一个典型的线性结构。栈、队列、串等都是线性结构，有惟一的前趋结点，有惟一的后继结点。

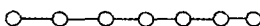
(3) **树形关系**：结构中的元素存在着一对多的层次关系，有惟一的父结点，可有多个子结点。

(4) **图状结构或网状结构**：结构中的元素存在着多对多的任意关系。一个结点可能有多个直接前趋和直接后继。

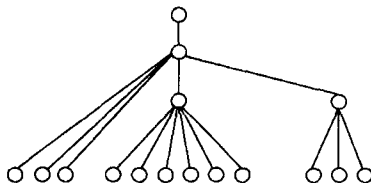
集合



线性结构



树



图

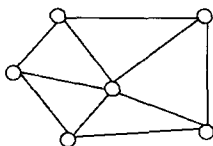


图 1-4 数据结构的四种基本类型

由于集合的关系非常松散，因此可以用其他的结构代替它，故数据的逻辑结构可概括如下：

逻辑结构 { 线性结构——线性表、栈、队、字符串、数组、广义表
非线性结构——树、图

3. 数据的四种基本存储结构

数据元素及其关系在计算机存储器内的表示，称为数据的存储结构 (Storage Structure)，又称为物理结构。

数据的存储结构是逻辑结构用计算机语言的实现 (亦称为映像)，它包括数据元素的表示和数据元素之间关系的表示。数据元素的表示在其他课程中已有表述，在此，我们仅讨论数据元素之间关系的表示方法。

数据的存储结构可用以下四种基本存储方法来实现。

(1) **顺序存储方法**。该方法把逻辑上相邻的结点存储在物理位置上相邻的存储单元里，结点间的逻辑关系由存储单元的邻接关系来体现，即顺序存储结构是借助数据元素在存储器中的相对位置来表示数据元素之间的关系。由此得到的存储表示称为顺序存储结构，通常借助程序语言的数组描述。

该方法主要应用于线性的数据结构。顺序存储结构的主要特点是：

①存储空间必须是连续的，即一块连续的存储空间。

②逻辑上相连，必然物理位置上相连。

③存储空间利用率高。可通过首地址直接计算得到数据元素的地址，从而实现对数据元素的随机存取。

④特别适用于线性结构。

【例 1.3】 对于表 1-1 的学生成绩表，假设每一个数据元素占用 1 个存储单元，该表在内存起始地址为 1000，则它的顺序存储结构如图 1-5 所示。

地址	
1000	(1000, 张三, 70, 80, 150)
1001	(1001, 李四, 90, 65, 155)
1002	(1002, 王二, 80, 90, 170)
1003	(1003, 王五, 40, 60, 100)
1004	

图 1-5 顺序存储结构

(2) **链式存储结构**。链式存储结构是借助于指示数据元素存储地址的指针来表示数据元素之间的关系。在链式存储结构中，每个数据元素增加若干个数据项，该数据项用于存放后一个或若干子数据元素的地址，通过这些数据项(又称为“指针”)来表示数据元素之间的关系。链式存储结构的特点是：

①存储空间不必是连续的。该方法不要求逻辑上相邻的结点在物理位置上亦相邻，结点间的逻辑关系由附加的指针字段表示。

②存储空间利用率不高。

③不能随机存取，所以检索等运算速度较顺序存储慢。

【例 1.4】 对于表 1-1 的链式存储结构如图 1-6 所示。

(3) **索引存储方法**。该方法通常在储存结点信息的同时，还建立附加的索引表。索引表由若干索引项组成。若每个结点在索引表中都有一个索引项，则该索引表称之为**稠密索引**。若一组结点在索引表中只对应一个索引

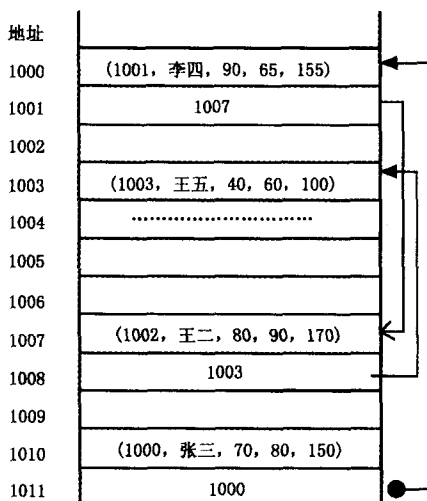


图 1-6 链式存储结构

项，则该索引表称为**稀疏索引**。索引项的一般形式是：(关键字、地址)，关键字是能惟一标识一个结点的那些数据项。稠密索引中索引项的地址指示结点所在的存储位置；稀疏索引中索引项的地址指示**一组结点的起始存储位置**。

(4) **散列存储方法**。该方法的基本思想是：根据结点的关键字直接计算出该结点的存储地址。

以上四种基本存储方法既可单独使用，也可组合起来对数据结构进行存储映像。

同一逻辑结构采用不同的存储方法，可以得到不同的存储结构。选择何种存储结构来表示相应的逻辑结构，视具体要求而定，主要考虑运算方便及算法的时空要求。

4. 数据运算

数据运算是**对数据施加的操作**，某一种运算算法的设计取决于选定的逻辑结构，而算法的实现依赖于采用的存储结构。

数据的运算定义在数据的逻辑结构上，每种逻辑结构都有一个运算的集合。最常用的检索、插入、删除、更新、排序等运算实际上只是在抽象的数据上所施加的一系列抽象的操作。所谓抽象的操作，是指我们只知道这些操作是“做什么”，而无需考虑“如何做”。只有确定了存储结构之后，才考虑如何具体实现这些运算。常见的数据运算有：