

21 世纪高等学校应用型规划教材



计 算 机 系 列

新编数据结构 教程

XINBIAN
SHUJU
JIEGOU
JIAOCHENG

马春江 李慧勇 孟繁军 编著



中国电力出版社
www.infopower.com.cn

21 世纪高等学校应用型规划教材



计算机系列

新编数据结构 教程

XINBIAN
SHUJU
JIEGOU
JIAOCHENG

马春江 李慧勇 孟繁军 编著



中国电力出版社

www.infopower.com.cn

内容提要

数据结构是计算机以及相关专业的一个重要专业理论基础课程。本书系统地介绍了线性表、栈、队列、串、二维数组、广义表、树、森林、二叉树、图等十大结构,为后续的操作系统、编译原理、数据库原理等课程奠定了良好的基础。本书还结合查找和排序两大应用领域深入探讨了各类数据结构和程序设计技巧的综合应用,最后对文件系统进行了初步的介绍。书中还把程序设计的三大控制结构(顺序、分支、循环)、函数调用、递归技术、界面设计、菜单管理、算法设计与分析等程序设计技巧或基础知识做了较为详细的介绍,提出了计算机编程“点式思维”的思想方法。

本书的主要特点为篇章布局合理,难度递增幅度较小,语言通俗易懂,图示充分、提示性强,案例与分析翔实,利于教学或自学,特别注重数据结构的来龙去脉和背景知识的研讨,算法设计尽量符合软件工程规范并且给出了部分可执行源代码。

本书适合作为大学本科和高职高专计算机科学与技术、计算机应用、软件工程以及相关专业数据结构课程的教材,也可以作为各类计算机程序设计培训班和计算机爱好者自学使用。

图书在版编目(CIP)数据

新编数据结构教程/马春江等编著. —北京:中国电力出版社, 2006

21 世纪高等学校应用型规划教材·计算机系列

ISBN 7-5083-4101-5

I. 新... II. 马... III. 数据结构-高等学校-教材 IV. TP311.12

中国版本图书馆 CIP 数据核字(2006)第 158459 号

丛 书 名: 21 世纪高等学校应用型规划教材·计算机系列

书 名: 新编数据结构教程

出版发行: 中国电力出版社

地 址: 北京市三里河路 6 号

邮政编码: 100044

电 话: (010) 68362602

传 真: (010) 68316497, 88383619

本书如有印装质量问题, 我社负责退换

服务电话: (010) 88515918 (总机)

传 真: (010) 88518169

E-mail: infopower@cepp.com.cn

印 刷: 北京同江印刷厂

开本尺寸: 185×260

印 张: 17

字 数: 420 千字

书 号: ISBN 7-5083-4101-5

版 次: 2006 年 3 月北京第 1 版

印 次: 2006 年 3 月第 1 次印刷

印 数: 0001—4000 册

定 价: 25.00 元

版权所有, 翻印必究

创新是计算机界发展的原动力

(自序)

计算机的出现原本就是人类的奇迹，能够使用二进制把现实生活中的各种信息进行高速和正确地处理，实在是一件很令人从心底震撼的事情。而计算机从 1946 年 2 月 15 日在美国发明出的第一台 ENIAC 起，到今天也才近 60 年的时间，但是已经发生了突飞猛进的变化。不论是硬件设计，还是软件开发，都早已不知道升级了多少代了。而这其中，创新意识始终是计算机界能够飞速发展的一个原动力。

第一个案例是互联网的出现，早期的计算机能够在本机中完成计算任务就很了不起了，之后的数据共享经过软盘、磁带、光盘等中间介质也可以做到，但是人们不满足于现状，从大、中、小型机的 CPU 共享，利用终端进行联网，到实际上平等的计算机之间互连成局域网，再到更大空间的城域网，最后到遍布全世界的互联网，可以把不同的计算机硬件平台、操作系统平台、种族、国家轻松地联系在一起，甚至可以进行全球化的分布式多机并行计算，达到“计算”能力的“极致”，这就是创新带来的巨大变化。

第二个案例是游戏软件的开发，除了操作系统等系统应用软件的开发，没有任何一类软件的开发会比游戏软件对于计算机的发展贡献更大了。因为早期计算机游戏不论是功能还是界面都很简单，如今已经发展到可以利用互联网技术使得成千上万的不同国度、不同性别、不同文化程度的人坐在一起玩游戏，这真是一件令人惊叹不已的奇迹。在这些软件的开发中，出现了很多的创新。从含有声音、图片、影像等形式多媒体技术到现今的利用互联网远程图像传输、语音传输、异地互动等大型游戏的开发，更是导致了許多新主板、显卡、开发工具、外围设备和相关产业的产生。

第三个案例是智能机器人的发展，早期的机器人只能从事简单而单调的工作，如加工印刷电路板、给汽车外壳喷漆等。而人们依然不满足于现状，努力地开发出各种带有一定智能的计算机，如进行水下作业、危险场所的工作等，即使到了今天，人们还在努力地开发出更先进的机器人。曾经看到过国外的机器人可以跳舞，机器鱼可以在水里自由游动，感到非常的惊奇，因为它在机械的动作中已经融入了“美”，但是动作的速度基本是恒定的，而中国人最新研发的机器人已经可以做出武术的套路，如剑术。因为不光动作很美观和规范，还要有速度的变化、力度的展现等更高深的境界，所以中国人在这个领域中用自己的实力达到了令全球瞩目的成就。相信在不久的将来，真的会出现在医院里可以听懂病人的话、可以根据现场自动做出决策和实际动作来帮助病人的机器人。这绝不是空想，而是创新所带来的必然结果。

实际上，不光是计算机界，在现实生活中，任何一个领域乃至人类，要想进步，都要靠我们的创新意识和具体扎实的工作。新的时代需要更多创新的人、创新的思想和创新的产品。

数据结构是程序开发的基础，是进入程序设计殿堂的敲门砖。虽然目前有很多版本的数据结构教材，但是我努力把这一本书写成充满创新和激情的形式，在其中您可以体会到我的良苦用心和精雕细刻，可以感受到数据结构的“形式美”和“力度美”，它是我多年来勤恳工作、认真教学和仔细思考的结晶，在付出了很多艰辛之后我终于有了今天的收获，也希望您能分享和品味这其中学习的乐趣。

马春江
2005 年岁末

前言

从您阅读本书开始,我和您就一起踏上了征服数据结构“痛并快乐着”的征程,说它很艰难,是因为大多数人在学习它的时候都会感到过于抽象和高深,但是希望您在阅读本书的时候,会感到愉快、兴奋、刺激和激动,您可能会多次被科学家们的奇思妙想所震撼,您可能会突然发现程序设计原来是这样的引人入胜,您可能会情不自禁地希望上机去实现一些算法,体会一下超越之后的成就感,所以我希望您在学习本书的时候是充满乐趣的。

我从事数据结构的教學快二十年了,据每一届学生反馈,数据结构是计算机专业的一门很难学好的课程,普遍反映太难理解和难于实现。我自己也在反思,从高级语言的学习直接过渡到这样抽象和有一定理论深度的课程,任何人肯定都会不适应。2005年在一次全国教学交流会议上,我有幸认识了天津的王立柱老师,他用了多年的时间,把C、C++以及数据结构进行了全面的整合,给我留下了深刻的印象。这促使我开始动笔,把我在多年教学中的心得体会写出来,这就是本书的来历。

本教材和其他的相比有大量的细节差异。本书的体系架构是基于学生易于理解和老师易于组织教学、可以进行互动讨论模式的。本书一定会使您感到学习数据结构不再是一个拦路虎,而且也会激发出您进行深入学习程序设计理论的兴趣和动手进行程序设计的原动力。

本书的内容主要分为三个方面,第一是对数据结构中逻辑结构的讨论,第二是对数据结构中存储结构的讨论,第三是对数据结构的应用特别是一些基础算法设计进行讨论。同时为了后续课程的学习,特意选择了一些与后续课程相关的内容。

本书的写作力求在以下几个方面有所突破,为教师组织教学和学生自学教材内容提供更强的针对性和方便性。

(1)整体框架设计更加易于学习。本书的创新点之一为打破“分类编写”,我把排序技术中5种基本的经典算法,提前到数据结构正式开始之前,既复习了高级语言的相关知识,又初步体现了数据结构的作用,还把数据结构相关的程序设计、效率分析或算法设计风格通过实例进行了讲解,使得学生能够有模仿的基础,有分析和训练的资料。通过排序算法的多样性也会激发学生的好奇心和求知欲望。而在第14章里再讨论更抽象、更奇特的排序方法或利用数据结构才能完成的排序,这样更能体现由浅入深、螺旋式上升的学习方法。而像字符串高级匹配技术没有放在“串”一章而是放在“查找技术”一章也是基于降低学生学习难度的考虑。对于算法的时间效率分析,我采用了尽量从简的方式进行处理,程序设计的另外一门核心课《算法设计与分析》中会把时空效率分析作为重点知识之一进行详细讨论。

(2)对基础的重视度更高。本书的创新点之二是避免开始就系统介绍数据结构,在学习数据结构课程中,许多学生感到原理还可以理解,但是一旦要求上机编程实现,总是很难动手,很明显他们缺乏的不是高级语言的知识,而是编程的方法论。在本书的开端,我把程序设计中必须了解的方法论进行了介绍,首创性地提出了人类的“面式思维”和计算机的“点式思维”的关系等概念,并且进行了对比。递归机制是数据结构中常用的程序设计技巧,在

线性表中的链表存储结构定义部分就已经开始出现递归,所以把它也放在正式的数据结构之前进行了介绍,为正式的数据结构拓宽了视野,打下了良好基础。

(3) 教材使用的语言更加通俗,本书的创新点之三为避免过分强调专业术语,避免那种英文教材不通顺的翻译,避免晦涩难懂,在本来就很抽象的数据结构知识中尽量减少人为的难度,我的写作目标就是“通俗化”、“口语化”和“亲切化”,把数据结构的知识作为重点凸显出来。本书的目标之一就是深入浅出、充分讨论、利于教学或自学。

(4) 范例选择更多更广。本书的创新点之四是避免以理论为主、范例较少、案例更少、很少理论结合实际的书写方式,因为那样容易使学生失去学习的兴趣和目标。本书中提供了大量的范例和案例分析,使您能充分理解到数据结构在工作中、生活中的巨大作用,同时也提出了很多挑战课题,使您能充分发挥自己的主观能动性,积极主动地研究学习而不是被动地记住一些死板的理论。

(5) 分析过程更加全面深入。本书的创新点之五为避免直接给出结论、较少进行讨论的写法,因为那样的写法使得数据结构充满了神秘莫测的色彩,使学生感到知其然却不知其所以然。本书中我力求把每一个结论和每一种新的数据结构的来龙去脉介绍清楚,甚至也讨论一些过程性结论,使您能够感受到事物发展的脉络和规律,也就真正地进入了数据结构的殿堂。

(6) 算法设计细节更加符合规范。本书力求避免在算法设计上过多的随意性,比如变量名的起法,数据结构的算法通常比高级语言中的练习题更难理解,所以大量使用诸如 h 、 p 、 q 、 i 、 j 、 k 、 x 、 y 、 z 等类似这样的单变量名,必然会增加阅读难度,另外还有规范缩格、结构空行、算法注释等诸多细节。虽然在数据结构中提供的并不是程序而是算法,但是正是这些细节使学生日积月累地形成良好的开发习惯,所以在本书中算法细节更注重软件工程设计规范,也便于理解和改编。

(7) 适合对象更加广泛。我希望本书能够满足各类不同需求的目标,数据结构教材通常过于抽象和复杂,自学很难。计算机发展到今天,突飞猛进日新月异,很多青年学子虽然学习的不是计算机专业,甚至没有进入大学校门的机会,但是依然渴望深入学习数据结构等重要的理论知识,那么我的书对于他们来说,希望是一个机会,在撰写过程中,我力求使任何一个具备高中基础知识、高级语言语法基础的人都能够较容易地自学下去。同时我还提供了较难的、还需要进一步研究的课题,所以这本书将适合各种层次的、渴望学习数据结构的读者。本书的目标之一就是要尽最大的努力使不论是学生自学还是老师授课都很容易进入角色。而每一小节前括号中的索引也可以使您较容易知道正在讨论什么内容,以决定是否需要阅读。由于学时数的限制,广义表和文件通常都是选学的内容,所以我注明了特殊标记*,供老师根据总学时数和学生的接受程度灵活处理,也供学有余力的学生自学和研究。每一章的最后,我都提供了数目不等的挑战课题,有的较为容易,而有的很难在短期内完成,还有的可能需要其他课程的相关知识,特此说明。

首先我要感谢我的父母、我的妻子在编写本书过程中给予的很多关心和支持。

我还要感谢我在清华大学攻读研究生课程时的指导老师石纯一教授的鼓励,他在阅读了本书书稿后的评价“你花了很多心血,有观点且有体会,真是不容易。值得祝贺!”给了我很大的鼓舞,在此表示发自内心的感激之情。

我还要特别感谢的是钱新恩教授、姜木林教授、史旅华副教授给予我精神上的鼓励和帮

助，本书在编写过程中，还得到了本院付勇智高工、梅琴老师等人的支持，还有我的学生马哲江、王康、吴倩等人在编辑校对方面和程序编写方面做了很多实际工作，

本书能得以出版，还要感谢很多朋友的大力支持。青岛酒店管理学院计算机技术学院的李慧勇院长和内蒙古师范大学计算机信息与工程学院的孟繁军老师也参与了部分内容的编写，天津师范大学管理学院的王立柱老师也提出了很多中肯的建议。

最后希望读者不吝批评、指正。

马春江
2005 年岁末

目 录

前 言

第 1 章 数据结构基础知识	1
1.1 引言	1
1.2 数据结构的基本概念和术语	4
1.3 逻辑结构	5
1.4 存储结构	5
1.5 数据结构的基本操作	6
1.6 算法和算法效率分析基础	7
1.7 数据存储的特点	10
1.8 程序设计基本流程控制结构	11
1.9 指针、链表的概念和基本操作	14
1.10 面式思维和点式思维	15
1.11 程序设计的基本技巧	16
*1.12 抽象数据类型	16
1.13 本章总结	17
第 2 章 基本的排序算法及程序设计	20
2.1 引言	20
2.2 排序算法设计与讨论	21
2.3 排序的程序设计	27
2.4 排序的应用	36
2.5 本章总结	37
第 3 章 递归技术及算法设计	38
3.1 引言	38
3.2 递归的基本概念	39
3.3 简单递归问题	39
3.4 复杂递归问题	41
3.5 本章总结	42
第 4 章 线性表的构造和实现	45
4.1 引言	45
4.2 逻辑结构的讨论	45
4.3 线性表的顺序存储	47
4.4 线性表的链接存储	49

4.5	线性表链接存储的变形.....	51
4.6	主要算法设计与讨论.....	53
4.7	线性表的其他算法设计.....	60
4.8	本章总结	63
第 5 章	栈的构造和实现.....	65
5.1	引言	65
5.2	逻辑结构的讨论.....	65
5.3	栈的顺序存储.....	66
5.4	栈的链接存储.....	67
5.5	主要算法设计与讨论.....	68
5.6	栈的应用	71
5.7	本章总结	76
第 6 章	队列的构造和实现.....	77
6.1	引言	77
6.2	逻辑结构的讨论.....	77
6.3	队列的顺序存储.....	78
6.4	队列的环状顺序存储.....	80
6.5	队列的链接存储.....	81
6.6	主要算法设计与讨论.....	81
6.7	队列的应用.....	83
6.8	本章总结	86
第 7 章	串的构造和实现.....	87
7.1	引言	87
7.2	逻辑结构的讨论.....	87
7.3	串的顺序存储.....	91
7.4	串的链接存储.....	92
7.5	串的索引存储.....	92
7.6	主要算法设计与讨论.....	94
7.7	串的应用	98
7.8	本章总结	99
第 8 章	二维数组的构造和实现.....	100
8.1	引言	100
8.2	逻辑结构的讨论.....	100
8.3	二维数组的行序或列序优先顺序存储.....	101
8.4	特殊矩阵的压缩存储.....	103
8.5	稀疏矩阵的压缩存储.....	104
8.6	稀疏矩阵的十字链表存储.....	105
8.7	二维数组结构的算法设计.....	107
8.8	本章总结	116

*第9章 广义表的构造	117
9.1 引言	117
9.2 逻辑结构的讨论	117
9.3 存储结构: 链接存储	119
9.4 表结构的应用	121
9.5 本章总结	123
第10章 树和森林的构造	125
10.1 引言	125
10.2 逻辑结构的讨论	125
10.3 树的顺序存储	129
10.4 树的链接存储	129
10.5 树的顺序和链接联合存储	130
10.6 树的应用	133
10.7 本章总结	136
第11章 二叉树的构造和实现	137
11.1 引言	137
11.2 逻辑结构的讨论	137
11.3 二叉树的顺序存储	139
11.4 二叉树的链接存储	140
11.5 二叉树的根序遍历和算法设计	141
11.6 二叉树的层次遍历和算法设计	145
11.7 其他主要算法设计与讨论	146
*11.8 线索二叉树	147
11.9 二叉树的应用	152
11.10 树、森林和二叉树的关系	161
11.11 本章总结	163
第12章 图的构造和实现	165
12.1 引言	165
12.2 逻辑结构的讨论	165
12.3 图的顺序存储	169
12.4 图的链接存储	171
12.5 遍历操作的算法设计	176
12.6 图的应用之一: 通信网最小生成树的研究和算法设计	180
12.7 图的应用之二: 公路网最短路径的研究和算法设计	185
12.8 图的应用之三: AOV网与拓扑排序的研究和算法设计	189
12.9 本章总结	192
第13章 查找技术及算法设计	195
13.1 引言	195
13.2 初级查找技术: 基于静态数据结构的查找	196

13.3	中级查找技术：基于动态数据结构的查找.....	200
13.4	高级查找技术：基于哈希表结构的查找.....	205
13.5	特殊查找技术：基于字符串结构的查找.....	210
13.6	查找的应用.....	214
13.7	本章总结	215
第 14 章	复杂排序技术及算法设计	216
14.1	引言	216
14.2	希尔排序技术.....	216
14.3	快速排序技术.....	217
14.4	树形选择排序技术.....	219
14.5	堆排序技术.....	220
14.6	基数排序技术.....	223
14.7	归并排序技术.....	226
*14.8	外部排序技术简介	228
14.9	本章总结	230
*第 15 章	文件的基本原理和应用	231
15.1	引言	231
15.2	存储结构之一：顺序文件.....	233
15.3	存储结构之二：索引文件.....	234
15.4	存储结构之三：索引顺序存取方法文件.....	236
15.5	存储结构之四：虚拟存储存取方法文件.....	237
15.6	存储结构之五：直接存取文件（散列文件）	239
15.7	存储结构之六：多重表文件.....	240
15.8	存储结构之七：倒排文件.....	241
15.9	文件系统的实用操作命令.....	242
15.10	文件应用的程序实现.....	252
15.11	本章总结.....	255
参考文献		259

第 1 章 数据结构基础知识

本章主要介绍数据结构的基本概念和计算机程序设计的一些基础知识，为后续的数据结构理论奠定基础。主要内容有数据结构的基本概念和术语、几种主要的逻辑结构、存储结构与基本操作的名称和特点，讨论了算法和算法效率分析，并且把数据结构需要的基础知识进行了归纳，供读者参考。

1.1 引言

在开始学习数据结构时您可能会不大明白数据结构有什么重要性，分不清楚它和高级语言之间有什么区别，看到书中出现的一些算法，也许会认为数据结构就是讨论高级语言和编程，实际上它和高级语言之间既有联系又有很大区别。下面首先通过几个范例来说明数据结构的重要性和特点。

【范例介绍 1-1】**【搭鸡窝和盖高楼的对比】**一个农民在自家的院子里搭建的鸡窝和建筑工人盖的高楼都是房子，这是相同点，那么不同点呢？您可能找出几十种不同点，但是这里最关注的是内部的结构，搭建一个鸡窝可以很随意，不行了推倒重来就行，但是盖一座大楼就不能这样随意。很显然，我们可以说大楼是建筑工人建造的，但是更可以说是设计人员构建的，因为在动工之前，所有的设计细节都已经考虑好，比如哪个地方有个门窗、哪个地方是楼梯、哪个地方安放避雷针等，显然不可能边造边想，发现这间屋子没有门于是砸个门，发现这层楼没有楼梯于是补建个楼梯。那么为什么要在开工之前把所有细节都设计完毕呢？因为它的结构太复杂，如果不考虑充分，可能会导致大楼彻底报废。本范例说明量变会引起质变。当我们面临的任务或系统足够复杂时，事先的整体设计就成为首要的工作。有些人在用高级语言编程时，不进行整体设计，边想边做，也不按照一些规范书写（如缩格制、空行以及注释等），他们通常认为只要能运行就没有什么大问题。但是一旦该开发任务涉及的数据之间的关系比较复杂，就可能出现开发失败的结果，所以我们在软件开发时必须首先考虑各种情况，尤其是数据的关系（也就是数据结构）的设计。

【范例介绍 1-2】**【图书管理系统的计算机化】**在没有计算机之前，图书馆就已经存在。对于成千上万的图书，试想一下有一个超大的礼堂，里面堆着几万本书像一座小山，您要借书，管理员说里面肯定有，您进去找吧。即使您有耐心，也许花几天您也找不到呀，显然我们必须使用一些管理思想。实际上图书馆把书进行编号，之后再分类，再按照楼层、房间、书架号等信息进行放置，在您提出借书要求后，只要提供书号，管理人员就会在最短的时间里找出那本书，这就是管理的效果。虽然管理人员能很快找到那本书，却不能提供更深入的服务，比如，原来的几本被谁借去、什么时候能还回来、图书馆中某个著者编著的所有书有哪些等等。但如果启用计算机，这些问题就很容易解决，但是对于成千上万的数据，如果需要数个小时、数天甚至更多的时间才能得到查询结果，那么即使这个结果是准确的，也可能没有实

用的价值。本范例是说明有时即使功能能够实现，但是运行效率太低，也没有实用价值，故必须探讨如何利用一些数据结构提高查找操作的时间效率。

【范例介绍 1-3】【用计算机实现人机对弈（下棋、打扑克等）】很多朋友喜欢计算机是因为利用计算机可以玩一些娱乐性的游戏。作为计算机专业工作者，您需要更深层的思考而不仅仅是从中取乐。如各种棋类的对弈是如何实现的。先不谈如何行棋，仅仅是界面实现就很困难，如棋盘和棋子如何实现，显然不能是简单的图片，因为还要考虑棋子的移动。而且既然要对弈，那么计算机怎么知道哪些方向可以走，哪些地方可以落子，对方的哪些棋子可以被消灭，而自己的哪些棋子已经被威胁，进一步又如何决策进攻或防守呢？这些问题甚至还有更多的细节是如何解决的？现在您可以体会到即使是界面也是很难设计的，如果没有数据结构的帮助，将很难编出精美好玩的各种游戏。

【范例介绍 1-4】【图形处理的本质】如果您用过一些图形处理软件，如 Photoshop，您就会被它神奇的功能所震撼，因为它可以轻易地对图片进行放大、缩小、移动、旋转、剪裁、拼接、变色等操作。实现这些功能其实很简单，这就是数学的力量，因为这些操作都是通过矩阵运算而来的，那么如果您面临巨大的矩阵相乘，您考虑过如何存储，又如何运算吗？因为希望很快能看到变形的结果，也就很关注变化的速度即使用的时间，您有什么好的加快速度的思路呢？面临巨大的数据量，您又没有好的存储方法能节省存储量？这个范例是要说明很多数学原理是非常有用的，使用计算机处理也是必然的，因而数据结构也就显得非常重要。

【范例介绍 1-5】【计算机计算表达式的难题】计算机能计算表达式，这好像是天经地义的，要不然为什么叫计算机呢？实际上虽然您使用计算机进行计算很简单，但是当初计算机科学家们实现它却是很困难的，让我们一起来体会一下。试想，如果目前的表达式是 $2+1$ ，计算机敢做吗，不行，因为可能是 $2+11$ ，所以必须往后读，即使是遇到运算符，还是不能做，因为可能是 $2+11\times 3$ ，根据表达式的运算规则，必须先计算 11×3 ，可是表达式可能是 $2+11\times 3^2$ ，那么能不能先算平方呢？不能，因为原来的表达式可能是 $2+11\times 3^{(1+1)}$ ，既然有括号，那么刚才的问题就会反复出现，如 $2+11\times 3^{(1+1\times 2)}$ ，即使是中间的部分计算结果出来了，还是有问题，如果原式为 $2+11\times 3^{(1+1\times 2)}-6$ ，那么现在该退回去做加法呢，还是往后走去做减法呢？过去的数据和运算符存到哪里去了，后面的数据又如何存储，如果这些数值用一些变量存储，那么启用多少变量呢？变量之间如何管理呢？内部的逻辑关系是什么呢？这些问题如果不能很好的解决，那么想让计算机进行“计算”那真是天方夜谭，好在计算机科学家们已经为我们巧妙地解决了这些问题，目前大家都能使用计算机做一些日常数值计算，更别说火箭上天、宇宙飞船的运行更要依靠大量的计算。在后面的讨论中我将揭秘计算机是如何进行表达式计算，此处的范例是要说明计算机要实现的很多功能仅仅靠编程技巧是很难解决的，必须依赖一些数据结构的帮助。

【范例介绍 1-6】【痕迹检测、DNA 检测、文字识别、人像识别等技术的实现前提】现代科技为我们带来了很多人闻所未闻、想也想不到的科技成果，比如公安系统能够使用指纹、血液等信息追查罪犯、医院能够用 DNA 做亲子鉴定、计算机可以进行文字或语音识别等。而这些技术必须解决一个很困难的问题，那就是必须存储超大量的数据从而进一步进行快速查找，这些数据量的总量远远超过我们想象。那么面对海量的数据，计算机是如何存储的，又是如何存储数据之间的关系的，而且又是如何实现上述功能呢？这个范例是要说明数据结构对查找技术能起到巨大的作用。

【数值计算问题的特点】在计算机发展的初期，人们使用计算机的目的主要是处理数值计算问题，如天气预报计算、军事需要计算、大型工程计算等。使用计算机解决一个具体问题一般需要经过以下几个步骤：首先从具体问题抽象出一个数学模型，然后设计或选择一个求解此数学模型的算法，最后编写程序进行调试和测试，直至得到最终的结果。数值计算问题的特点是数据量不太大，但是计算量巨大，数据之间的关系不大复杂，容易处理。其主要工作是努力提高运算速度和提高精度。由于早期所涉及的运算对象是简单的整型、实型或布尔类型数据，所以编程者的主要精力集中在程序设计的技巧上，而无须重视数据结构。

【非数值计算问题的特点】随着计算机应用领域的扩大和软、硬件的发展，非数值计算问题显得越来越重要。据统计，当今处理非数值计算性问题（如数据处理）占用了 90% 以上的机器时间。这类问题涉及到的数据结构相当复杂，数据元素之间的关系已经无法用数学方程式加以描述，故非数值计算问题的特点是数据量往往巨大，且关系复杂，但是计算量却不大。因此，解决这类问题的关键不再是数学分析和计算方法，而是要设计出合适的数据结构。这类非数值计算问题的数学模型不再是数学方程，而是诸如表、树、图之类的数据结构。因此，可以说数据结构课程主要是研究非数值计算的程序设计问题中所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。学习数据结构的目的是为了了解计算机处理对象的特性，将实际问题中所涉及的处理对象在计算机中表示出来，并对它们进行处理。与此同时，通过算法训练来提高学生的逻辑思维能力，通过程序设计的技能训练来促进学生的综合应用能力和专业素质的提高。

【数据结构】计算机科学就是一门研究数据表示和数据处理的科学。数据是计算机化的信息，它是计算机可以直接处理的最基本和最重要的对象。在进行科学计算、数据处理、过程控制以及对文件的存储、检索及数据库技术等计算机应用领域中，都有对数据处理的过程。因此，要设计出一个结构良好、效率很高的程序，必须研究数据的特性及数据间的相互关系及其对应的存储表示，并利用这些特性和关系设计出相应的算法，进一步编出程序。

【数据结构研究的主要内容】数据结构课程主要讨论软件开发过程中的设计阶段，同时设计编码和分析阶段的若干基本问题也会被综合讨论。此外，为了构造出好的数据结构并实现，还需考虑数据结构及其实现的评价与选择。因此，数据结构的内容包括 3 个层次的 5 个“要素”，见表 1-1。

表 1-1 数据结构研究的主要内容

方 面 层 次	数据表示	数据处理
抽象层	逻辑结构	基本操作
实现层	存储结构	算法
评价层	各种存储结构的比较及算法分析	

通过以上讨论，您是否可以初步感受到数据结构的重要性？如果觉得感受还不是很深刻，那么就随着讨论的深入慢慢体会计算机科学家的智慧和各种数据结构的神奇魅力吧！很显然，您仅仅了解和掌握高级语言的知识或语法是不够的，因为即使您背熟所有的语法细节，但是需要用计算机来解决很多类似上面范例的问题时，您还会遇到很难克服的困难，您缺乏的不是语法知识，而是解决相关问题的思想方法，这其中就有很多是基于数据结构方面的。

1.2 数据结构的基本概念和术语

类似任何一门新的学科或课程, 数据结构也有一些新的概念和术语, 熟记这些概念和术语可以方便后面的学习。

【术语介绍】【数据 (Data)】 首先来讨论一下什么是数据。通常人们谈及数据, 就会想到 0~9 或由它们组合的数字, 而实际上这些应该被称为数值。在现实生活中, 人们经常谈到“信息”这个概念, 而一旦这些信息被计算机存储, 就是所谓的“数据”。要注意的是“数据”并不能完全覆盖“信息”, 比如日常生活中的“眉目传情”, 这样的情感表达是一种信息, 但是并不能被计算机识别、存储从而处理。所以“数据”是信息的一种载体, 它能够被计算机识别、存储和加工处理。计算机科学中, 数据可以是数值数据, 也可以是非数值数据。数值数据是一些整数、实数或复数, 主要用于工程计算、科学计算和商务处理等; 而非数值数据包括字符、文字、图形、图像、语音等。而其中的文字又包括英文、中文、其他国家的文字、中国少数民族的文字等类型。

【数据元素 (Data Element)】 数据元素是数据的基本单位。在不同的场合下, 可把数据元素称为元素、结点、顶点、记录等。

【数据项 (Data Item)】 有时一个数据元素还可以由若干个数据项组成, 例如, 人员信息的每一个数据元素就是一个人的记录。它可能包括编号、姓名、性别、籍贯、出生日期、电话、手机、通信地址、电子邮件地址等数据项。这些数据项再次被分为两种: 一种叫做基本项, 如性别、籍贯等, 这些数据项在数据处理时不再分割; 另一种叫做组合项, 如出生日期可以分为年、月、日等更小的项。

【数据对象 (Data Object)】 数据对象也可以称作数据元素类 (Data Element Class), 是具有相同性质的数据元素的集合。在某些具体问题中, 数据元素具有相同的性质, 属于同一数据对象, 即数据元素是数据对象的一个实例。

【数据结构 (Data Structure)】 数据结构是指互相之间存在一种或多种关系的数据元素的集合。在使用计算机处理任何问题时, 我们不是仅仅关注数据本身, 而一定会关注数据元素之间的关系, 因为这些“关系”就是我们需要的“信息”, 这种数据元素之间的关系就是所谓的“结构”。

【数据结构的形式定义】 数据结构是一个二元组

$$\text{Data_Structure} = (D, R)$$

其中, D (Data 的头字母) 是数据元素的有限集, R (Relation 的头字母) 是 D 上关系的有限集。

【数据的逻辑结构和数据的存储结构】 数据的“逻辑结构”可以看作是从具体问题抽象出来的数学模型, 是数据本身的关系而已, 它与数据的存储并无关系, 有点像“纸上谈兵”。但是我们研究数据结构的目的是为了在计算机中实现对它们的存储和其他操作, 为此还需要研究如何在计算机中表示逻辑结构。而计算机中的存储方法就是“存储结构”, 有时也称为数据的物理结构。它主要研究的是逻辑结构在计算机中的实现方法, 包括: 元素的表示和元素之间关系的表示。所以对数据结构目前可以理解为两个层面, 即“逻辑结构”和“存储结构”。但是在不引起误会的情况下, 有时也会把逻辑结构或存储结构直接称为数据结构。下面将通过专门的小节继续展开讨论这两个术语。

1.3 逻辑结构

既然逻辑结构主要是数据之间可能存在的关系，那么下面就先考察一下现实生活中有什么样的关系被我们关注。

【逻辑结构】根据数据元素之间可能存在的关系，可将逻辑结构分为以下 4 种基本结构：线性结构、树形结构、图形结构和集合结构。

【线性结构】该结构的数据元素之间（从一个方向而言）只有一维的关系，即线性关系，也称为“一对一”关系。

【树形结构】该结构的数据元素之间存在分支的关系，也称为“一对多”关系。这可以从一个军队的组成体系来理解：一个团有多个营，一个营有多个连，一个连有多个班，一个班有多名战士等。

【图形结构】该结构的数据元素之间可以有任意的关系，也称为“多对多”关系，图形结构有时也称作网状结构。这可以从城市的交通来理解，从一个地点到另外一个地点可能存在很多路径，也可能绕了一圈，又回到原地。

【集合结构】集合结构主要强调元素的“整体性”和元素的“存在性”，而不再关心数据元素之间的关系，在操作上主要关注“是否存在其中”，显然集合是元素关系很松散的一种结构，但是有的情况下这种结构依然是很重要的。比如，在高级语言设计中，如何判断程序设计者使用了哪些变量，以及哪些是定义过的，哪些是没有定义过的，为检查语法错误提供一些参考。

图 1-1 表示了 4 种逻辑结构的原理示意图。

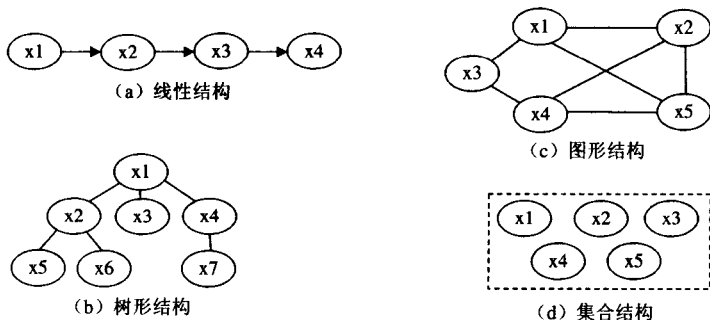


图 1-1 4 种逻辑结构的原理示意图

1.4 存储结构

考察了现实生活中有什么样的关系被我们关注后，下一步就必须知道计算机中能提供哪些存储方案供我们选择。

【存储结构】常用的数据存储结构有 4 种：顺序存储、链接存储、索引存储和哈希存储。

【顺序存储】顺序存储方法就是启用一批物理位置相邻的存储单元，然后将逻辑上相邻的元素依次存储在其中。不过这是狭义的顺序存储，后面会看到变形的顺序存储。顺序存储结构是一种最基本的存储表示方法，可以借助于高级语言中的“数组”机制来实现。

【链接存储】借助于高级语言中的指针机制可以实现链表，而这种链接存储方法对逻辑上