

高等学校计算机应用规划教材

数据结构

(C 语言版) (第二版)

		梁海英	李淑梅	主 编
刘艳玲	罗 琳	赵方珍	罗兰花	副主编

清华大学出版社

北 京

内 容 简 介

本书基于作者多年的教学经验,从实用的角度出发,对线性和非线性数据结构的顺序和链式存储及其操作进行了详细讲解,在教给学生数据结构设计和算法设计的同时,培养学生分析问题、解决问题和总结问题的能力。书中的每一章均配有实战练习及大量习题,实现了理论与实践相结合,让学生学以致用。本书免费提供电子课件、源代码及习题答案,全部案例均已在 Visual C++ 6.0 环境中成功运行。

本书既可作为普通高校计算机类专业和电子信息类相关专业的教材,也可作为计算机类专业考取硕士研究生或博士研究生的参考教材。

本书配套的电子课件、习题答案和实例源代码可以到 <http://www.tupwk.com.cn/downpage> 网站下载,也可以通过扫描前言中的二维码下载。

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: C 语言版/梁海英,李淑梅主编. —2 版. —北京:清华大学出版社,2021.2

高等学校计算机应用规划教材

ISBN 978-7-302-57397-5

I. ①数… II. ①梁… ②李… 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 ②TP312.8

中国版本图书馆 CIP 数据核字(2021)第 017755 号

责任编辑:胡辰浩

封面设计:高娟妮

版式设计:孔祥峰

责任校对:成凤进

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:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:大厂回族自治县彩虹印刷有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:16.75 字 数:429 千字

版 次:2017 年 8 月第 1 版 2021 年 4 月第 2 版 印 次:2021 年 4 月第 1 次印刷

定 价:69.00 元

产品编号:089358-01

前言

数据结构不仅是一般程序设计的基础，而且是设计和实现编译程序、操作系统、数据库系统及其他系统程序的重要基础。目前比较权威的数据结构教材大多是考研指定教材，难度比较大，不太适合应用型本科、三本及专科学生使用。为此，我们编写了这本教材，通过引入大量案例，将复杂的理论问题直观化、简单化。这种案例驱动式教学，更有利于这个层次的学生接受。

我们基于多年丰富的教学经验及素材积累，精心编写此书，目的是让初学者能循序渐进地掌握各种数据结构及操作，力求透彻、全面、易学、易用，充分调动学生的学习积极性。书中使用 C 语言定义各种数据结构、描述算法。本书对每种数据结构和算法的剖析都遵循由浅入深的原则，并配以实用的案例和图示，配有相应的 C 语言源代码，适合具有 C 语言基础的数据结构初学者。

全书共分 8 章，对于常用的数据结构，如线性表、栈、队列、特殊矩阵、广义表、树、二叉树、图等进行深入讲解，使学生能够全面地理解基本概念、逻辑结构、存储结构、操作运算、实现算法以及案例应用，进而利用比较法讲解各种查找和排序的方法，并对各种算法的性能进行分析，以便在不同的应用场合选取合适的方法。

本书由梁海英和李淑梅任主编，刘艳玲、罗琳、赵方珍和罗兰任副主编，全书由贺州学院梁海英教授统稿。

在本书的编写过程中，得到了多位同行的热心帮助和支持，参加本书内容编写、程序调试、课件制作、习题收集、答案制作、内容审校等工作的老师有姚建盛、巫湘林、殷玉玲、曾霖、马文成、任秀丽、谭晓东、罗志林、陈冠萍、袁淑丹等，在此向他们表示衷心的感谢！

由于编者水平有限，书中难免存在不妥之处，敬请读者谅解并提出宝贵意见。我们的电话是 010-62796045，邮箱是 992116@qq.com。

本书配套的电子课件、习题答案和实例源代码可以到 <http://www.tupwk.com.cn/downpage> 网站下载，也可以通过扫描下方二维码下载。



编者
2020 年 10 月

目 录

第1章 绪论	1
1.1 数据结构概述	1
1.2 常用术语和基本概念	3
1.3 数据类型	6
1.4 算法和算法复杂度	6
1.4.1 算法的重要性	7
1.4.2 时间复杂度	8
1.4.3 空间复杂度	10
1.5 本章实战练习	11
1.6 本章小结	13
1.7 习题1	13
第2章 线性表	17
2.1 线性表概述	17
2.2 线性表的顺序存储及运算的实现	18
2.2.1 线性表的顺序存储	18
2.2.2 顺序表的基本操作	19
2.3 线性表的链式存储及运算的实现	24
2.3.1 单链表	24
2.3.2 循环链表	33
2.3.3 双向链表	34
2.4 本章实战练习	37
2.4.1 顺序表的常用操作	37
2.4.2 单链表的常用操作	40
2.4.3 通讯录管理	42
2.5 本章小结	47
2.6 习题2	49
第3章 栈和队列	55
3.1 栈	55

3.1.1 栈的定义	55
3.1.2 栈的顺序存储与操作	56
3.1.3 栈的链式存储与操作	60
3.2 队列	61
3.2.1 队列的定义	61
3.2.2 队列的顺序存储与操作	62
3.2.3 队列的链式存储与操作	65
3.3 本章实战练习	67
3.3.1 top为指针且指向栈顶元素的 下一个位置	67
3.3.2 top为整数且指向栈顶元素的 当前位置	70
3.3.3 栈的应用——数制转换	71
3.3.4 顺序队列的基本操作	73
3.3.5 循环队列设置不同队空与队 满条件的解决方案	75
3.3.6 链队列的基本操作	77
3.4 本章小结	80
3.5 习题3	81
第4章 特殊矩阵和广义表	87
4.1 特殊矩阵及其压缩存储	87
4.1.1 特殊矩阵的压缩存储	88
4.1.2 稀疏矩阵及其压缩存储	90
4.2 广义表	93
4.2.1 广义表的定义	93
4.2.2 广义表的存储结构及实现	93
4.3 本章实战练习	95
4.4 本章小结	98
4.5 习题4	99

第5章 树	101
5.1 树的概念	101
5.1.1 树的定义	101
5.1.2 树的基本术语	102
5.2 二叉树	104
5.2.1 二叉树的定义	104
5.2.2 二叉树的性质	106
5.3 二叉树的存储结构	108
5.3.1 二叉树的顺序存储	108
5.3.2 二叉树的链式存储与操作	109
5.4 二叉树的遍历	111
5.4.1 遍历算法	111
5.4.2 线索二叉树	115
5.5 树与森林	118
5.5.1 树和森林的存储	118
5.5.2 二叉树、树和森林的转换	122
5.5.3 树和森林的遍历	123
5.6 哈夫曼树	123
5.6.1 哈夫曼树的定义	124
5.6.2 哈夫曼树的存储定义	124
5.6.3 哈夫曼树的构造算法	124
5.6.4 哈夫曼编码	126
5.7 本章实战练习	127
5.7.1 二叉树的基本操作	127
5.7.2 线索二叉树的操作	132
5.7.3 哈夫曼树的构造	136
5.8 本章小结	137
5.9 习题5	137
第6章 图	147
6.1 图的定义和基本术语	147
6.1.1 图的定义	147
6.1.2 图的基本术语	148
6.2 图的存储与操作	151
6.2.1 邻接矩阵	151
6.2.2 邻接表	154
6.3 图的遍历	157
6.3.1 深度优先遍历算法	158
6.3.2 广度优先遍历算法	159

6.4 图与最小生成树	162
6.4.1 生成树和森林的概念	162
6.4.2 最小生成树	163
6.5 最短路径	167
6.5.1 单源点到其余各顶点的最短路径	167
6.5.2 任意源点之间的最短路径	170
6.6 AOV网与拓扑排序	170
6.6.1 AOV网	170
6.6.2 拓扑排序	171
6.7 AOE网与关键路径	172
6.7.1 AOE网	172
6.7.2 关键路径	173
6.8 本章实战练习	176
6.8.1 图的邻接矩阵操作	176
6.8.2 图的邻接表操作	180
6.8.3 利用邻接矩阵实现连通图的 深度优先遍历	184
6.8.4 利用邻接表实现连通图的 深度优先遍历	186
6.8.5 利用邻接矩阵实现连通图的 广度优先遍历	188
6.8.6 利用邻接表实现连通图的 广度优先遍历	191
6.8.7 普里姆最小生成树算法	194
6.8.8 迪杰斯特拉最短路径算法	196
6.9 本章小结	199
6.10 习题6	199
第7章 查找	209
7.1 查找的基本概念	209
7.2 静态查找算法	211
7.2.1 顺序查找	211
7.2.2 折半查找	213
7.3 动态查找表	216
7.4 哈希表	219
7.4.1 哈希表的定义	219
7.4.2 哈希函数的构造	220
7.4.3 处理冲突的方法	221
7.4.4 哈希表的查找和性能	223

7.5 本章实战练习	224	8.3 交换排序算法	238
7.5.1 顺序查找算法	224	8.3.1 冒泡排序	238
7.5.2 折半查找算法	225	8.3.2 快速排序	239
7.5.3 二叉排序树查找算法	226	8.4 选择排序算法	241
7.6 本章小结	228	8.4.1 直接选择排序	242
7.7 习题7	228	8.4.2 堆排序	243
第8章 排序	233	8.5 归并排序算法	245
8.1 排序的基本概念	233	8.6 排序算法的比较	247
8.2 插入排序算法	234	8.7 本章实战练习	248
8.2.1 直接插入排序	234	8.8 本章小结	253
8.2.2 折半插入排序	236	8.9 习题8	253
8.2.3 希尔排序	237	参考文献	257

第 1 章

绪 论

本章主要介绍数据结构的相关概念以及它所研究的问题与内容，目的是让读者对数据结构有个大致的了解，为后续内容的学习提供必要的基础知识。本章介绍的主要内容包括数据结构概述、常用术语和基本概念、数据类型、算法和算法复杂度。

1. 总体要求

掌握数据结构的基本概念；掌握数据的逻辑结构和存储结构之间的关系；了解算法的五个特征；掌握通过计算语句执行次数来估算算法时间复杂度的方法。

2. 相关知识点

相关术语：数据、数据元素、数据项、数据结构。数据的逻辑结构：集合结构、线性结构、树结构和图结构。数据的物理结构：顺序存储结构、链式存储结构、索引结构和散列结构等。算法的五个特征、时间复杂度及空间复杂度。

3. 学习重点

数据的逻辑结构和存储结构及其之间的关系；算法时间复杂度、空间复杂度及其计算。

1.1 数据结构概述

图灵奖获得者 N. Wirth(沃斯)提出“数据结构(Data Structure)+算法(Algorithm)=程序(Program)”。计算机算法与数据的结构密切相关，算法无不依附于具体的数据结构，数据结构直接关系到算法的选择和效率。运算由计算机来完成，这就需要设计能够实现相应插入、删除和修改操作的算法。也就是说，数据结构还需要给出每种结构类型所定义的各种运算的算法。

一般来说，用计算机来解决一个具体问题时，大致需要经过以下几个步骤：首先，要从具体问题中抽象出一个适当的数学模型，然后设计一个解此数学模型的算法，最后编写程序、进行调试和运行直到得到最终解答。

寻求数学模型的实质是分析问题，从中提取操作的对象，并找出这些操作对象之间含有的关系，然后用数学的语言加以描述。当人们用计算机处理数值计算问题时，所用的数学模型是用数学方程描述的，因此程序设计者的主要精力集中于程序设计技巧上，而不是数据的存储和组织上。然而，计算机应用的更多领域是“非数值型计算问题”，它们的数学模型无法用数学方程描述，而是要用数据结构描述。解决此类问题的关键是设计出合适的数据结构，非数值型

计算问题的数学模型是用线性表、树、图等结构来描述的。

【例 1.1】 学生信息登记表。

每年新生入学都会用类似表 1.1 所示的二维表进行信息登记，以便完成各种数据的统计。二维表(即线性表)是经常用到的数学模型。

表 1.1 学生信息登记表

学号	姓名	性别	民族	籍贯	出生日期
20210108001	苏宏	男	汉族	南宁市	20030121
20210108002	梁琪	女	满族	四平市	20030913
20210108003	韦华	男	壮族	桂林市	20030815
20210108004	覃婷	女	汉族	广州市	20031203

【例 1.2】 食堂就餐排队管理问题。

在食堂就餐排队管理中，为了保证公平，满足先来先得的原则，排队就餐的算法应该是“先到的同学先打饭离开”。相应地，就餐排队管理模型应该是一个“队列”，即就餐窗口的服务人员应该先为排在“队头”的同学提供就餐服务。当陆续有同学再来就餐时，应该让他主动在“队尾”排队等候，如图 1.1 所示。队列是经常用到的一种数学模型。

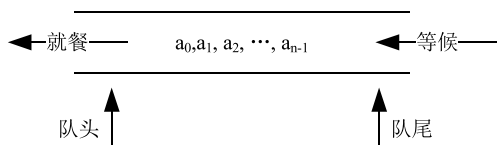


图 1.1 学生就餐管理

【例 1.3】 大学组织架构。

为了方便分级管理，一所大学下面会分成多个学院，每个学院会有多个系，每个系会包含多个班级，如图 1.2 所示。因此，若将一个学校的组织架构画在一张图上，则可得到一棵倒立生长的“树”。“树根”是学校，而所有的“叶子”就是具体的班级，寻找班级的过程就是从“树根”沿“树杈”到某个“叶子”的过程。“树”可以是某些非数值型计算问题的数学模型，它也是一种数据结构。

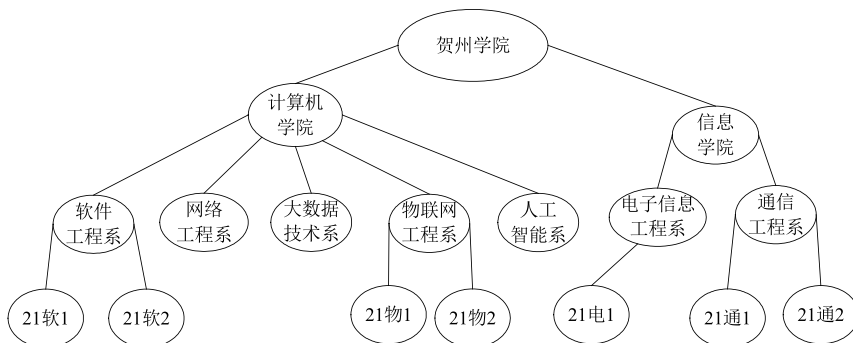


图 1.2 大学组织架构图

综合以上3个例子可见,描述这类非数值型计算问题的数学模型不再是数学方程,而是诸如二维表、队列和树之类的数据结构。因此,简单来说,在非数值计算的程序设计问题中,数据结构是一门研究计算机的操作对象及其相互之间的关系和运算等的学科。

1.2 常用术语和基本概念

要学好“数据结构”这门课程,必须明确各种概念及其相互之间的关系。本节只介绍一些常用的术语和基本概念,其他的相关术语和概念将在以后各章节中陆续讲解。

1. 数据

数据(Data)是信息的载体,是可以被计算机识别、存储并加工处理的描述客观事物的信息符号的总称。在计算机学科中,数据是指所有能输入计算机中,并能被计算机程序所处理的符号的集合,它是计算机程序加工处理的对象。因此,客观事物包括数值、字符、声音、图形、图像等,它们本身并不是数据,只有通过编码变成能被计算机识别、存储和处理的符号后才是数据。

2. 数据元素和数据项

数据元素(Data Element)是在某一问题中作为整体进行考虑和处理的基本单位。在计算机中存储数据时,都是以一个数据元素为单位的。数据元素有时也被称为节点(Node)、顶点(Vertex)、记录(Record)等。一个数据元素由若干数据项组成。

数据项(Data Item),也称项(Item)或域(Field),是描述数据的不可分割的、含有独立意义的最小单位。

例如,例1.1描述一个学生信息的数据元素可由下列6个数据项构成:学号、姓名、性别、民族、籍贯、出生日期。

3. 数据结构

数据结构(Data Structure)由数据和结构两部分构成。其中,数据部分是指数据元素的集合;结构就是关系,结构部分是指数据元素之间关系的集合。因此,数据结构就是指数据元素的集合及数据元素之间关系的集合。概括地讲,数据结构就是指相互之间有一种或多种特定关系的数据元素的集合。在计算机上要处理数据,就要保存数据及它们之间的关系。在这里,关系就是数据的逻辑结构,它指反映数据元素之间的逻辑关系的数据结构,其中的逻辑关系是指数据元素之间的前后件关系,而与它们在计算机中的存储位置无关。

1) 数据的逻辑结构

数据的逻辑结构(Logic Structure)是从具体问题抽象出来的数学模型,与数据在计算机中的具体存储没有关系。数据的逻辑结构独立于计算机,是数据本身所固有的特性。从逻辑上可以把数据结构分为线性结构和非线性结构,主要包括集合、线性、树和图结构。

(1) 集合结构。

集合结构(Set Structure)中的数据元素除了“同属于一个集合”的关系外,再无其他关系。如整数集、字符集等。

(2) 线性结构。

线性结构(Linear Structure)中的数据元素之间存在“一对一”的关系，如数组、队列等。

(3) 树结构。

树结构(Tree Structure)中的数据元素之间存在“一对多”的关系。比如例 1.3 中的大学组织架构等。

(4) 图结构。

图结构(Graphic Structure，也称网状结构)中的数据元素之间存在“多对多”的关系，如城市交通图等。

图 1.3 是上述 4 种结构的关系图。其中，树结构和图结构又称为非线性结构。由于集合中数据元素之间的关系是非常松散的，因此常用其他几种结构来描述集合。

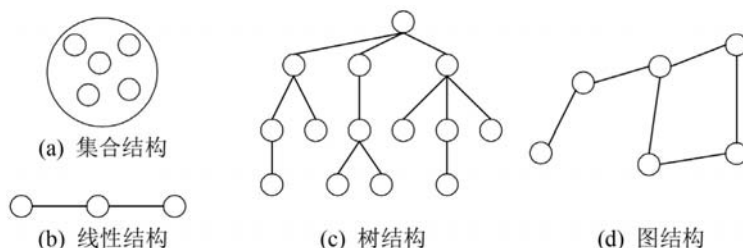


图 1.3 4 种基本数据结构关系图

数据结构由相互之间存在着一种或多种关系的数据元素的集合和该集合中数据元素之间的关系组成。数据结构的形式化定义记为： $\text{Data_Structure}=(D,R)$ 。其中， D 是数据元素的有限集合， R 是数据集合 D 中所有元素之间的关系的有限集合。

【例 1.4】定义集合 $D=\{3, 6, 9, 18, 27\}$ 的数据结构。

$DS_1=(D,R_1)$ ，其中 R_1 定义为 D 上的“>”(大于)关系，则数据结构 DS_1 可以表示为如图 1.4(a)所示的形式，该结构为线性结构。 $DS_2=(D,R_2)$ ，其中 R_2 定义为 D 上的“整除”关系，则 $R_2=\{(3,6),(3,9),(3,18),(3,27),(6,18),(9,18),(9,27)\}$ ，数据结构 DS_2 可以表示为如图 1.4(b)所示的形式，该结构为图结构。

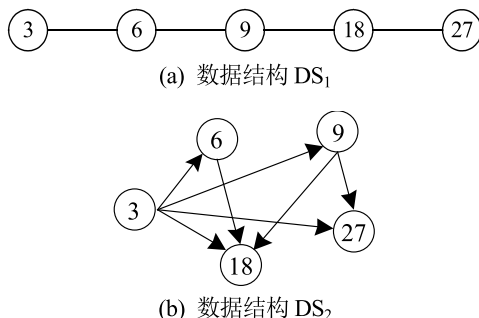


图 1.4 集合 D 上定义的两个数据结构

从上面的例子可以看出，即使是由相同元素构成的集合，只要定义的关系不同，也不是同一数据结构。数据结构不仅描述了结构中的元素，还描述了这些元素之间的关系。数据结构的定义仅是对操作对象的一种数学描述，结构中定义的关系是数据元素之间的逻辑关系。

数据结构可以分为逻辑上的数据结构和物理上的数据结构。数据结构的形式化定义为逻辑

结构。物理结构为数据在计算机中的存储，它包括数据元素的存储和关系的存储。在计算机中存储信息的最小单位是二进制的位(bit)，可以用一个由若干位组合起来形成的一个位串来存储一个数据元素。因此可将位串看成是数据元素在计算机中的存储形式。

2) 数据的物理结构(Physical Structure)

数据的物理结构，又称为存储结构(Storage Structure)，是指数据的逻辑结构在计算机存储空间中的存放形式。数据的物理结构是数据结构在计算机中的存储，它包括数据元素的机内存储和关系的机内存储。由于具体实现的方法有顺序、链式、索引、散列等多种，因此一种数据结构可存储成一种或多种存储结构。

数据元素的机内存储方法：用节点存储数据元素。当数据元素由若干数据项组成时，节点中与各数据项对应的部分称为数据域(Data Field)。

关系的机内存储方法：数据元素之间的关系的机内存储可以分为顺序存储和非顺序存储，这样就可以得到两种不同的存储结构，即顺序存储结构和链式存储结构。顺序存储借助元素在存储器中的相对位置来存储数据元素之间的逻辑关系。链式存储借助指示元素存储位置的指针(Pointer)来存储数据元素之间的逻辑关系。

(1) 顺序存储结构(Sequence Storage Structure)。

顺序存储结构通过数据元素在计算机存储器中的相对位置来存储数据元素的逻辑关系，一般把逻辑上相邻的数据元素存储在物理位置相邻的存储单元中，它是一种最基本的存储方法，一般采用数组来实现。

(2) 链式存储结构(Linked Storage Structure)。

链式存储结构对逻辑上相邻的两个数据元素不要求其存储位置必须相邻，元素间的逻辑关系通过指针来存储，一般采用链表来实现。链式存储结构中的数据元素称为节点，在节点中附设地址域(Address Domain)来存储与该节点相邻的节点的地址，从而实现节点间的逻辑关系。图 1.5 给出了图 1.4 中数据结构 DS_1 的不同存储方式。

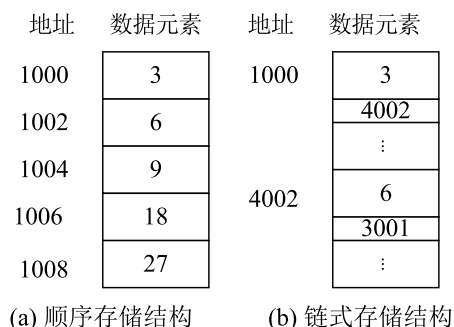


图 1.5 数据结构 DS_1 存储结构示意图

3) 数据的运算

算法的设计取决于数据的逻辑结构，而算法的实现依赖于数据采用的存储结构。数据的存储结构实质上是它的逻辑结构在计算机存储器中的实现，为了全面地反映一个数据的逻辑结构，它在存储器中的映像包括两方面的内容，即数据元素之间的信息和数据元素之间的关系。不同的数据结构有其相应的若干运算。数据的运算是在数据的逻辑结构上定义的操作算法，如查找、

插入、删除、更新和排序等。数据的运算是数据结构的一个重要方面，讨论任何一种数据结构时都离不开对该结构上数据的运算及其实现算法的探讨。

1.3 数据类型

1. 数据类型

数据类型(Data Type)是高级程序设计语言中的概念，是数据的取值范围和对数据进行操作的范围。数据类型规定了程序中对象的特性。程序中的每个变量、常量或表达式的结果都应该属于某种确定的数据类型。一方面，在程序设计语言中，每一个数据都属于某种数据类型。类型显式或隐式地规定了数据的取值范围、存储方式以及允许进行的运算。数据类型是在程序设计中已经实现的数据结构。另一方面，在程序设计过程中，当需要引入某种新的数据结构时，总是借助编程语言所提供的数据类型来描述数据的存储结构。

2. 本书在用 C 语言描述时的约定

(1) C 语言的数组元素的下标从“0”开始，为此，在表示数据结构时，数据元素的序号也从 0 开始。

(2) 数据元素的类型约定为 ElemType。具体的类型可以由用户在使用时定义：

```
typedef int ElemType /*定义所用数据类型为 int*/
```

(3) 数据存储结构用类型定义(typedef)描述，例如：

```
typedef struct{
    ElemType *elem;
    int length;
    int listsize;
}SqList; /*定义名为 SqList 的线性表采用顺序存储结构的类型定义*/
```

(4) 算法以函数形式描述：

```
类型标识符 函数名(形参表)
/*算法说明*/
{语句}
```

1.4 算法和算法复杂度

解决实际问题需要找出解决问题的方法。用计算机解决实际问题，就要先给出解决问题的算法，再依据算法编写程序完成要求。算法是指在有限的时间范围内，为解决某一问题而采取的方法和步骤的准确、完整的描述，它是一个有穷的规则序列，这些规则决定了解决某一特定问题的一系列运算。算法是程序设计的精髓，算法的设计取决于数据的逻辑结构，算法的实现取决于数据的物理结构。

1.4.1 算法的重要性

1. 算法的五个特征

1) 有穷性

一个算法必须总是(对任何合法的输入值)在执行有穷步之后结束,且每一步都可在有穷的时间内完成。这也是算法与程序的最主要区别,程序可以无限地循环下去,如操作系统的监控程序在机器启动后就一直监测着操作者的鼠标动作和输入的命令。

2) 确定性

算法中的每一条指令都必须有明确的含义,不应使读者产生二义性;并且在任何条件下,算法只有唯一的一条执行路径,即对于相同的输入只能得到相同的输出。

3) 可行性

一个算法是可以被执行的,即算法中的每个操作都可以通过已经实现的基本运算执行有限次来完成。

4) 有输入

根据实际问题的需要,一个算法在执行时可能要接收外部数据,也可能无须外部输入。所以一个算法应有零个或多个输入,这取决于算法本身要实现的功能。

5) 有输出

一个算法在执行完毕后,一定要有一个或多个结果或结论。这就要求算法一定要有输出,这些输出是同输入有着某些联系的量。

通常,解决同一个问题,不同的人有不同的想法,即使是同一个人,在不同的时间里可能对同一个问题的理解也不完全相同。算法是依据个人的理解和想法人为设计出来的求解问题的步骤,不同的人或同一个人不同的时间里设计出来的算法也不尽相同。那么,究竟哪种算法设计得好呢?如何衡量一个算法的好坏呢?

2. 算法效率的度量标准

通常,在设计算法时应该考虑从以下几个方面来度量算法的效率。

1) 正确性

正确性(Correctness),满足预先规定的功能和性能的要求,这是算法设计最基本的要求,算法应严格地按照特定的规格说明进行设计,要能够解决给定的问题。但是,“正确”一词的含义在通常的用法中有很大的区别,大体上可分为以下4个层次。

- (1) 依据算法所编写的程序中不含语法错误。
- (2) 程序对于几组输入数据能够得到满足规格要求说明的结果。
- (3) 程序对于经过精心挑选较为苛刻的几组输入数据也能够得到令人满意的结果。
- (4) 程序对于所有符合要求的输入数据都能得到正确的输出。

对于大型软件需要进行专业测试,一般情况下,通常以第(3)个层次的要求作为衡量算法正确性的标准。

2) 可读性

可读性(Readability)是指一个算法应当思路清晰、层次分明、简单明了、易读易懂。设计算法的主要目的是解决实际问题,在设计实现一个项目时,往往不是一个人独立完成的。为了达到可读性的要求,在设计算法时,一般要使用有一定意义的标识符来命名变量、函数等,以便能够“见名知义”。其次,可以在算法的开头或指令的后面加注释来解释算法和指令的功能。

3) 健壮性

健壮性(Robustness)是指一个算法应该具有很强的容错能力,当输入不合法的数据时,算法应当能做适当的处理,使得不至于引起严重的后果。当输入不合法的数据时,算法能做出相应的响应或进行适当的处理,避免带着非法数据执行,从而导致莫名其妙的结果。

4) 高效率

运行时间(Running Time)是指算法在计算机上运行所花费的时间,它等于算法中每条语句执行时间的总和。一般来说,执行时间越短,性能越好。依据算法编写的程序运行速度较快。

5) 低存储

占用空间(Storage Space)是指算法在计算机上存储所占用的存储空间,包括存储算法本身所占用的存储空间、算法的输入及输出数据所占用的存储空间和算法在运行过程中临时占用的存储空间。依据算法编写的程序在运行时所需的内存空间较小。

对于一个系统设计人员来说,前3项很容易实现。在使用软件时,人们更加注重于软件的运行速度,而后两项恰恰是影响速度的主要因素。

1.4.2 时间复杂度

1. 时间复杂度(Time Complexity)的定义

一个算法执行所耗费的时间,从理论上是不能算出来的,必须上机运行测试才能知道。但我们不可能也没有必要对每个算法都上机测试,只需知道哪个算法花费的时间多,哪个算法花费的时间少就可以了;并且一个算法花费的时间与算法中语句的执行次数成正比,哪个算法中语句的执行次数多,它花费的时间就多。一个算法中的语句执行次数称为语句频度或时间频度。一般情况下,算法中基本操作重复执行的次数是问题规模 n 的某个函数,用 $T(n)$ 表示,若有某个辅助函数 $f(n)$,使得当 n 趋近于无穷大时, $T(n)/f(n)$ 的极限值为不等于零的常数,则称 $f(n)$ 是 $T(n)$ 的同数量级函数,记作 $T(n)=O(f(n))$,称 $O(f(n))$ 为算法的渐进时间复杂度,简称时间复杂度。其中大写字母 O 为 Order(数量级)的第一个字母, $f(n)$ 为函数形式,如 $T(n)=O(n^2)$ 。

2. 时间复杂度的计算方法

在计算不同结构的算法的时间复杂度时有以下几个简单的方法。

(1) 对于一些简单的输入输出语句或赋值语句,近似认为需要 $O(1)$ 时间。

(2) 对于顺序结构,需要依次执行一系列语句所用的时间可采用大 O 下“求和法则”。

求和法则:是指若算法的两个部分的时间复杂度分别为 $T_1(n)=O(f(n))$ 和 $T_2(n)=O(g(n))$,则 $T_1(n)+T_2(n)=O(\max(f(n),g(n)))$ 。

(3) 对于选择结构,如 `if` 语句,它的主要时间耗费是在执行相应分支中的语句时所用的时

间，需要注意的是检验条件也需要 $O(1)$ 时间。

(4) 对于循环结构，循环语句的运行时间主要体现在多次迭代中执行循环体以及检验循环条件的时间耗费，一般可用大 O 下“乘法法则”。

乘法法则：是指若算法的两个部分的时间复杂度分别为 $T_1(n)=O(f(n))$ 和 $T_2(n)=O(g(n))$ ，则 $T_1(n) \times T_2(n)=O(f(n) \times g(n))$ 。

其中，求解循环结构算法的时间复杂度的具体步骤如下。

① 找出算法中的基本语句。

算法中执行次数最多的那条语句就是基本语句，通常是最内层循环的循环体。

② 计算基本语句的执行次数的数量级。

只需计算基本语句执行次数的数量级，这就意味着只要保证基本语句执行次数的函数中的最高次幂正确即可，可以忽略所有低次幂和最高次幂的系数。这样能够简化算法分析，并且使注意力集中在最重要的一点上：增长率。

③ 用大 O 记号表示算法的时间性能。

将基本语句执行次数的数量级放入大 O 记号中。

注意：

如果算法中包含嵌套的循环，则基本语句通常是最内层的循环体；如果算法中包含并列的循环，则将并列循环的时间复杂度相加。

例如：

```
for (i=1; i<=n; i++)
    x++;
for (i=1; i<=n; i++)
    for (j=1; j<=n; j++)
        x++;
```

若第一个 `for` 循环的时间复杂度为 $O(n)$ ，第二个 `for` 循环的时间复杂度为 $O(n^2)$ ，则整个算法的时间复杂度为 $O(n+n^2)=O(n^2)$ 。

(5) 对于复杂的算法，可以将它分成几个容易估算的部分，然后利用求和法则和乘法法则计算整个算法的时间复杂度。

【例 1.5】 分析如下程序段的时间复杂度。

```
s=0;
for(i=1;i<=n;i++)
    s=s+i;
```

分析：

`s=0`; 执行 1 次

`i=1`; 执行 1 次

`i<=n`; 执行 $n+1$ 次

`s=s+i`; 执行 n 次

`i++`; 执行 n 次

总的执行次数为 $3(n+1)$ 次，因此该算法的时间复杂度为 $T(n)=O(3(n+1))=O(n)$ 。

在分析算法的时间性能时，常用最基本语句的执行次数来估算。所谓最基本语句通常是指

最深层循环体中的语句，也是执行频率最快的语句。它的执行次数反映了整个算法的基本时间性能。如例 1.5 中的 $s=s+i$ 和 $i++$ 均被执行了 n 次，所以 $T(n)=O(n)$ 。

3. 算法时间复杂度的时间量级

实际上，算法的时间量级有多种形式，见表 1.2，其对应的函数曲线见图 1.6。

表 1.2 算法的时间量级分类表

名称	时间复杂度 $T(n)$	说明
常量阶	$O(1)$	与问题规模无关的算法
线性阶	$O(n)$	与问题规模相关的单重循环
平方阶	$O(n^2)$	与问题规模相关的二重循环
立方阶	$O(n^3)$	与问题规模相关的三重循环
指数阶	$O(e^n)$	较为复杂
对数阶	$O(\log_2 n)$	折半查找算法
复合阶	如 $O(n \log_2 n)$	堆排序算法
其他	不太确定	过于复杂

一个算法的时间复杂度可能存在最好情况和最坏情况，通常要以算法的平均时间复杂度来进行算法分析。但是算法的平均时间复杂度取决于各种数据出现的概率，难以进行分析。所以，往往借助于最坏时间复杂度来进行算法分析与评价。

一般地，对于足够大的 n ，常用的时间复杂度存在如下顺序：

$$O(1) < O(\log_2 n) < O(n) < O(n \log_2 n) < O(n^2) < O(n^3) < \dots < O(2^n) < O(3^n) < \dots < O(n!)$$

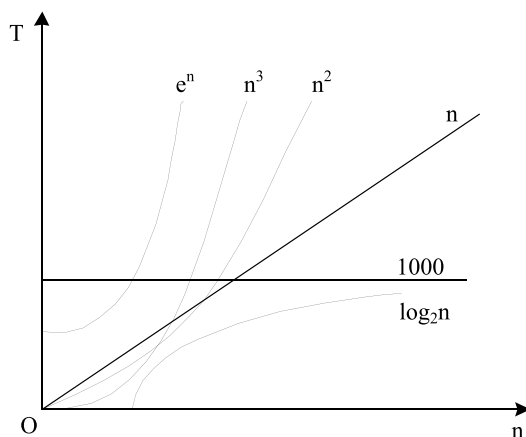


图 1.6 常见函数的增长率

1.4.3 空间复杂度

空间是指执行算法所需要的存储空间，算法所对应的程序在运行时所需的存储空间包括固定部分和可变部分。固定部分所占空间与所处理的数据结构外数据的大小和数量无关，或者称

与该问题的实例的特征无关，主要包括程序代码、常量、简单变量等所占的空间；可变部分所占空间与该算法在某次执行中处理的特定数据的大小和规模有关。例如，100 个数据元素的排序算法与 1000 个数据元素的排序算法所需的存储空间显然是不同的。

1. 空间复杂度(Space Complexity)的定义

与算法的时间复杂度类似，空间复杂度也是关于问题规模 n 的一个函数，当问题规模 n 趋近于无穷大时的空间量级就称为算法的渐进空间复杂度，简称空间复杂度。可以将空间复杂度作为算法所需存储空间的度量，记作 $S(n)=O(f(n))$ 。

2. 空间复杂度的计算方法

那么，算法的空间需求有哪些呢？大体上，依据算法所编写的程序除了需要存储空间来寄存程序本身所用的指令和数据外，还需要一些对数据进行操作的工作单元和存储一些为实现计算所需信息的辅助空间。通常，程序所占的空间变化不大，所以在此主要考虑算法的辅助空间需求。

【例 1.6】分析下面程序的时间复杂度和空间复杂度。

```
RevArray(int a[],int n)
{
    int i,j,*b;
    b=(int *)malloc(sizeof(int)*n);
    for(i=0,j=n-1;i<n;i++j--)
        b[j]=a[i];
    for(i=j=0;i<n;i++,j++)
        a[i]=b[i];
    free(b);
}
```

算法空间复杂度的分析：

因为基本语句就是循环内的赋值语句，共执行了 $2n$ 次，所以 $T(n)=2n=O(n)$ 。而算法的辅助空间是一个与问题规模同量级的一维数组 b 所占用的空间，另外再加上两个控制变量 i 和 j ，共 $n+2$ 个，所以 $S(n)=n+2=O(n)$ 。

1.5 本章实战练习

- 在数据结构中，从逻辑上可以把数据结构分为()。
 - 动态结构和静态结构
 - 紧凑结构和非紧凑结构
 - 线性结构和非线性结构
 - 内部结构和外部结构
- 设有数据逻辑结构为：

```
B=(K,R)
K=(k1,k2,...,k9)
R={<k1,k2>,<k1,k8>,<k2,k3>,<k2,k4>,<k2,k5>,<k3,k9>,<k5,k6>,<k8,k9>,<k9,k7>}
```

试画出这个逻辑结构的图示。

3. 算法应该具有什么特征?
4. 算法的衡量标准或依据是什么?
5. 分析下列程序片段的时间复杂度。

(1) 程序片段 1:

```
sum=0;
for(i=0;i<N;i++)
    sum++;
```

(2) 程序片段 2:

```
sum=0;
for(i=0;i<N;i++)
    for(j=0;j<N;j++)
        sum++;
```

(3) 程序片段 3:

```
sum=0;
for(i=0;i<N;i++)
    for(j=0;j<N*N;j++)
        sum++;
```

(4) 程序片段 4:

```
sum=0;
for(i=0;i<N;i++)
    for(j=0;j<i;j++)
        sum++;
```

(5) 程序片段 5:

```
sum=0;
for(i=0;i<N;i++)
    for(j=0;j<i;j++)
        for(k=0;k<j;k++)
            sum++;
```

实战练习参考答案

1. C。
2. 该逻辑结构示意图如图 1.7 所示。

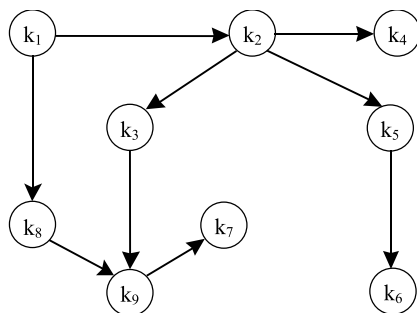


图 1.7 逻辑结构图

3. 一个算法应该具有 5 个特征：有穷性、确定性、可行性、有输入、有输出。
4. 算法是通过时间复杂度和空间复杂度进行度量的。其中，算法的时间复杂度是指运行算法时所需要消耗的时间；算法的空间复杂度是指算法在计算机内执行时所需存储空间的度量。
5. (1) $O(n)$ 。
(2) 是嵌套循环，其时间复杂度为 $O(n^2)$ 。
(3) 关键语句 `sum++` 执行次数为 $N*N*N$ ，因此，时间复杂度为 $O(n^3)$ 。
(4) 关键语句 `sum++` 执行次数为 $1+2+3+\cdots+N=N(N+1)/2$ ，因此，时间复杂度为 $O(n^2)$ 。
(5) $O(n^3)$ 。

1.6 本章小结

数据结构是计算机存储、组织数据的方式。数据结构是指相互之间存在一种或多种特定关系的数据元素的集合。一个数据结构是由数据元素依据某种逻辑关系组织起来的。对数据元素间逻辑关系的描述称为数据的逻辑结构；数据必须在计算机内存储，数据的存储结构是数据结构的实现形式，是其在计算机内的存储；此外，讨论一个数据结构必须同时讨论在该类数据结构上执行的运算才有意义。一个逻辑数据结构可以有多种存储结构，且各种存储结构会影响数据处理的效率。通常情况下，精心选择的数据结构可以带来更高的运行效率或者存储效率。

1. 数据结构的 4 种基本逻辑结构是集合结构、线性结构、树结构、图结构。
2. 数据结构的 4 种基本物理结构是顺序结构、链式结构、索引结构、散列结构。
3. 一个算法应该具有 5 个特征：有穷性、确定性、可行性、有输入、有输出。
4. 算法通过时间复杂度和空间复杂度进行度量。
5. 算法的时间复杂度是指运行算法时所需要消耗的时间。
6. 算法的空间复杂度是指算法在计算机内执行时所需存储空间的度量。

1.7 习题 1

一、选择题

1. 算法分析的目的是()。
A. 找出数据结构的合理性
B. 研究算法中的输入和输出的关系
C. 分析算法的效率以求改进
D. 分析算法的易懂性和文档性
2. 数据结构中，与所使用的计算机无关的是数据的()结构。
A. 存储
B. 物理
C. 逻辑
D. 物理和存储
3. 对一个算法的评价，不包括如下()方面的内容。
A. 健壮性
B. 并行性
C. 正确性
D. 可读性

4. 算法分析的两个主要方面是()。
 - A. 空间复杂性和时间复杂性
 - B. 正确性和简明性
 - C. 可读性和文档性
 - D. 数据复杂性和程序复杂性
5. 算法指的是()。
 - A. 计算机程序
 - B. 解决问题的计算方法
 - C. 排序算法
 - D. 解决问题的有限运算序列
6. 线性结构表示数据元素之间存在一种()关系。
 - A. 一对多
 - B. 多对多
 - C. 多对一
 - D. 一对一
7. 某程序的时间复杂度为 $(3n+n\log_2n+n^2+8)$, 其数量级表示为()。
 - A. $O(n)$
 - B. $O(n\log_2n)$
 - C. $O(n^2)$
 - D. $O(\log_2n)$
8. 计算机算法指的是()。
 - A. 计算方法
 - B. 排序方法
 - C. 解决问题的有限运算序列
 - D. 调度方法
9. 计算机算法必须具备输入、输出和()5个特性。
 - A. 可行性、可移植性和可扩充性
 - B. 可行性、确定性和有穷性
 - C. 确定性、有穷性和稳定性
 - D. 易读性、稳定性和安全性
10. 以下属于逻辑结构的是()。
 - A. 顺序表
 - B. 哈希表
 - C. 有序表
 - D. 单链表

二、填空题

1. 数据结构按逻辑结构可分为两大类, 它们分别是_____和_____。
2. 数据的物理结构主要包括_____和_____两种情况。
3. 数据结构是指数据及其相互之间的_____。当节点之间存在 M 对 $N(M:N)$ 的关系时, 称这种结构为_____。
4. 数据的逻辑结构被分为_____、线性结构、树结构和_____4种。
5. 线性结构中元素之间存在_____关系, 树结构中元素之间存在_____关系。
6. 数据结构的形式定义为 (D,R) , 其中 D 是_____的有限集合, R 是 D 上_____的有限集合。
7. 在图结构中, 每个节点的前驱节点数和_____节点数可以_____。
8. 在非数值计算的程序设计问题中, 数据结构是一门研究计算机的_____及相互之间的_____和运算等的学科。
9. 在线性结构中, 第1个节点_____前驱节点, 其余每个节点有且只有1个前驱节点; 最后1个节点_____后继节点, 其余每个节点有且只有1个后继节点。
10. 数据结构包括数据的_____、数据的_____和数据的运算这3个方面的内容。
11. 在树结构中, 根节点没有_____节点, 其余每个节点有且只有_____个前驱节点。
12. 数据的存储结构可用4种基本的存储方法表示, 它们分别是_____、_____、链式和散列。

三、判断题

1. 算法的优劣与算法描述语言无关，但与所用的计算机有关。 ()
2. 数据的物理结构是指数据在计算机内的实际存储形式。 ()
3. 数据元素是数据的最小单位。 ()
4. 数据的逻辑结构是指数据的各数据项之间的逻辑关系。 ()

四、简答题

1. 分析下面程序段的时间复杂度。

```
x=0;
for(i=1; i<n; i++)
    for(j=1; j<=n-i; j++)
        x++;
```

2. 分析下面程序段的时间复杂度。

```
i=1;
while(i<=n) i=i*3;
```

3. 分析下面程序段的时间复杂度。

```
s=0;
for(i=0; i<n; i++)
    for(j=0; j<n; j++)
        s+=B[i][j];
sum=s;
```

4. 分析下面程序段的时间复杂度。

```
for(i=0; i<n; i++)
    for(j=0; j<m; j++)
        a[i][j]=0;
```