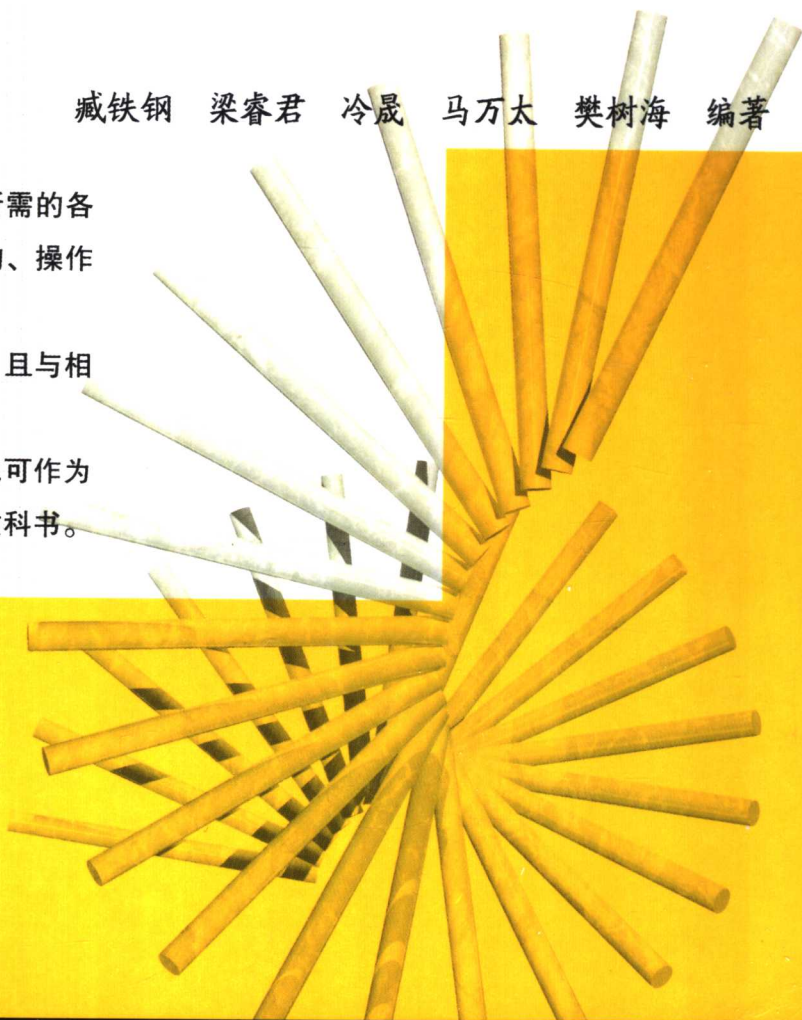


21世纪大学计算机基础规划教材

软件开发技术基础

臧铁钢 梁睿君 冷晟 马万太 樊树海 编著

- 本书全面涵盖了工程软件开发所需的各类知识内容，主要包括数据结构、操作系统、数据库原理及软件工程。
- 实例和习题丰富，具有代表性，且与相关知识要点结合紧密。
- 本书可供软件开发人员参考，也可作为大中专院校工程类专业学生的教科书。



中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE



21 世纪大学计算机基础规划教材

软件开发技术基础

臧铁钢 梁睿君

编著

冷 晟 马万太 樊树海



内 容 简 介

本书主要为了满足非计算机专业的工程类科技人员及研究人员,以及大专院校工程专业类的学生对工程软件开发技术的需要,介绍了工程软件系统的基本概念、数据结构、数据处理方法、操作系统原理、工程数据库和软件工程学等内容,并通过实例全面介绍了工程软件开发的过程。每章均配有习题。

本书内容丰富、全面,实用性强,兼顾了工程软件开发的各个方面,可作为工程专业的软件课程教材,也可作为从事计算机工程应用的科技人员的参考书。

图书在版编目(CIP)数据

软件开发技术基础/臧铁钢等编著. —北京:中国铁道出版社, 2005. 2

(21世纪大学计算机基础规划教材)

ISBN 7-113-06389-6

I. 软… II. 臧… III. 软件开发—高等学校—教材 IV. TP311.52

中国版本图书馆CIP数据核字(2005)第011582号

书 名: 软件开发技术基础

作 者: 臧铁钢 梁睿君 冷 晟 马万太 樊树海

出版发行: 中国铁道出版社(100054,北京市宣武区右安门西街8号)

策划编辑: 严晓舟 商其坤

责任编辑: 苏 茜 秦绪好 王占清

封面制作: 白 雪

印 刷: 北京新魏印刷厂

开 本: 787×1092 1/16 印张: 19 字数: 462千

版 本: 2005年3月第1版 2005年3月第1次印刷

印 数: 1~5000册

书 号: ISBN 7-113-06389-6/TP·1428

定 价: 25.00元

版权所有 侵权必究

凡购买铁道版的图书,如有缺页、倒页、脱页者,请与本社计算机图书批销部调换。

前 言

随着科学技术的发展,在工程领域大规模应用计算机技术已是大势所趋。先进精密的生产过程和完备的生产管理体系都需要计算机技术的支持。计算机技术在工程领域的应用,标志着一场新的技术革命。

工程应用有其特殊的实用性强的一面,如何将计算机技术与工程技术相结合是广大工程技术人员所面临的难题。在工程中,计算机的应用领域有两大类:一是工业控制;二是工业管理。在工业控制领域,常用的计算机硬件形式有工控机、PLC 及一些嵌入式控制单元等;在工业管理领域,常用的是网络支持下的通用型计算机。计算机技术是硬件和软件的结合,计算机硬件提供了一个应用平台,具体的行为实现要靠软件系统,软件是计算机运作的核心。解决工程问题的方法有一定的通用性,无论计算机的应用领域和硬件形式如何不同,软件开发的基本原理是相似的。从了解工程软件应用基础技术入手,是全面掌握计算机工业应用技术的重要途径。

本书以工程软件应用技术为背景,全面介绍了工程软件在开发过程中涉及的基本概念、算法及开发技术等方面的内容,并以实例的形式介绍了工程软件的相关概念及开发过程。主要论述了以下几个方面的内容:

- 工程中常用的数据结构。数据结构是软件开发的基础,它直接影响到工程软件的质量。在本书的第1章和第2章对工程中常用的数据结构的概念和具体组织进行了论述;
- 工程常用算法。数据处理技术是工程软件的基础,算法是数据处理技术的核心。本书的第3章介绍了常用算法的概念、实现过程及其在工程中的应用;
- 操作系统。操作系统是计算机底层系统软件,工程应用软件是建立在操作系统平台上的。了解操作系统可以使应用软件的开发和操作者充分利用计算机资源。本书第4章对操作系统原理进行了阐述;
- 工程数据库技术。工程数据管理需要数据库技术的支持,工程数据库在计算机工程应用中的地位越来越重要。本书第5章对数据库原理和数据库工程应用进行论述;
- 工程软件开发技术。软件工程学研究软件开发过程、合理配置开发资源的学科。在本书第6章介绍了以软件工程为核心的工程软件开发过程、对工程软件的开发具有指导意义。

本书由臧铁钢编写第1、4章,马万太、樊树海编写第7章,梁睿君编写第2、5章,冷晟编写第3、6章,臧铁钢统编全稿。另外,樊树海、冯爱民、赵文鹏、魏孝斌、陈立辉、唐才峰、陈学峰及薛建彬老师对本书的编写提供了很多帮助,本书的编写与出版还得到了中国铁道出版社的支持和帮助,在此一并表示诚挚的感谢。

由于编者水平有限,书中难免还存在一些缺点和错误,恳切希望广大读者批评指正。

作 者

2005年1月

于南京航空航天大学

目 录

第 1 章 工程软件的基础元素.....	1
1-1 工程软件概述	1
1-2 数据结构概述	3
1-2-1 数据结构及其数据运算的概念	3
1-2-2 数据结构的分类	5
1-2-3 数据结构的表示	7
1-2-4 数据类型及数据抽象	9
1-3 算法概述	11
1-3-1 算法的概念	11
1-3-2 算法的描述	12
1-3-3 算法分析	13
1-4 小结	16
习题 1	16
第 2 章 常用数据结构及其在工程中的应用	17
2-1 线性数据结构及其工程应用	17
2-1-1 顺序表	17
2-1-2 线性链表	22
2-1-3 索引存储	35
2-1-4 栈	39
2-1-5 队列	50
2-1-6 串	56
2-2 非线性数据结构及其工程应用	62
2-2-1 多维数组	62
2-2-2 树与二叉树	69
2-2-3 图	79
2-3 小结	90
习题 2	90
第 3 章 常用数据处理技术及其在工程中的应用	92
3-1 查找技术	92
3-1-1 顺序查找	92
3-1-2 二分查找	93
3-1-3 分块查找	95
3-1-4 二叉排序树查找	96
3-1-5 Hash 表技术	99

3-2 排序技术	104
3-2-1 交换排序	105
3-2-2 插入排序	108
3-2-3 选择排序	110
3-2-4 归并排序	113
3-2-5 基数排序	115
3-2-6 排序方法的比较	115
3-2-7 排序技术在软件工程中的应用	115
3-3 小结	115
习题 3	116
第 4 章 操作系统原理	117
4-1 操作系统简介	117
4-1-1 操作系统的概念	117
4-1-2 操作系统的类型与功能	121
4-2 进程管理	125
4-2-1 进程的概念	125
4-2-2 进程控制	127
4-2-3 进程的同步、互斥和通信	131
4-2-4 进程调度	138
4-2-5 死锁	140
4-3 存储管理	142
4-3-1 存储管理概述	142
4-3-2 常用存储管理技术	145
4-4 文件管理	149
4-4-1 文件管理概述	149
4-4-2 文件系统管理方式	152
4-5 设备管理	157
4-5-1 设备管理的概念	157
4-5-2 设备管理结构	160
4-5-3 缓冲技术	161
4-5-4 设备分配	161
4-5-5 虚拟设备	163
4-6 小结	165
习题 4	165
第 5 章 数据库技术	166
5-1 数据库技术概述	166
5-1-1 数据管理与数据库技术	166
5-1-2 数据库系统的结构	174

5-2 数据模型	180
5-2-1 信息的三种世界及其描述	181
5-2-2 实体联系模型	182
5-2-3 层次模型	185
5-2-4 网状模型	187
5-2-5 面向对象模型	187
5-2-6 关系模型	188
5-3 关系化与规范化	192
5-3-1 关系代数及查询优化	192
5-3-2 关系模式的规范化	202
5-4 分布式数据库	218
5-4-1 分布式数据库概念	218
5-4-2 分布式数据库管理系统	219
5-4-3 分布式数据库设计	219
5-5 工程数据库设计	221
5-5-1 数据库设计的基本概念	221
5-5-2 工程数据库设计的基本步骤	222
5-5-3 工程数据字典	229
5-6 工程数据库的新技术与新应用	230
5-6-1 WWW 数据库	230
5-6-2 多媒体数据库	231
5-6-3 移动数据库	232
5-6-4 地理信息系统	233
5-6-5 数字图书馆	233
5-7 小结	233
习题 5	234
第 6 章 工程软件开发技术	235
6-1 软件工程	235
6-1-1 软件	235
6-1-2 软件工程的观念	238
6-1-3 软件支持环境	238
6-2 工程软件设计准则	239
6-2-1 抽象准则	240
6-2-2 结构化准则	240
6-2-3 模块化准则	241
6-2-4 信息隐藏和局部化准则	242
6-3 工程软件设计的基本技术	242
6-3-1 全生命周期法	243

6-3-2	原型法.....	244
6-3-3	面向数据的设计法.....	245
6-3-4	面向对象的设计方法.....	251
6-3-5	统一建模语言（UML）.....	257
6-4	工程软件的测试技术.....	263
6-4-1	工程软件测试的概念.....	263
6-4-2	工程软件的测试过程.....	264
6-4-3	工程软件的测试方法.....	266
6-5	工程软件质量评价与软件质量保证.....	272
6-5-1	工程软件质量标准.....	272
6-5-2	工程软件质量评价.....	273
6-5-3	工程软件质量保证——CMM.....	275
6-6	工程软件新技术.....	276
6-6-1	J2EE 技术.....	276
6-6-2	软件复用.....	280
6-7	小结.....	280
	习题 6.....	281
第 7 章	工程软件实例——人力资源管理系统 的开发.....	282
7-1	项目背景和需求.....	282
7-1-1	项目背景分析.....	282
7-1-2	用户需求.....	282
7-1-3	用户需求分析.....	283
7-2	系统设计.....	284
7-2-1	运行方案.....	284
7-2-2	系统总体结构.....	284
7-2-3	数据流图.....	285
7-2-4	实体—关系设计.....	286
7-2-5	状态—迁移模型.....	288
7-3	系统实现.....	290
7-3-1	数据库设计.....	290
7-3-2	程序设计.....	294

第 1 章 工程软件的基础元素

开发工程软件的目的是要解决各类实际工程问题。工程数据和算法是工程软件的基本组成元素。工程数据是工程软件系统的处理对象，数据结构是对工程数据的组织。组织结构良好的数据不仅可以提高工程数据处理的效率，而且数据的可靠性也能得到有效的保障。算法提供处理数据的手段和方法，是提取数据内涵的一系列行为的总称。本章将重点介绍数据结构、算法的概念及内容。

1-1 工程软件概述

科学计算、信息处理、过程控制是计算机的三大应用领域。计算机最早应用于科学计算，20 世纪 70 年代，计算机的应用范围扩大到数据信息处理，数据信息处理是当今计算机应用的重要领域。由于计算机具有逻辑运算的能力，从 20 世纪 60 年代起计算机被广泛应用于工业生产过程的实时监测和控制工作，这是工程软件大量出现的开始。伴随着网络技术的飞速发展，计算机应用的网络化已引起了生产方式的深刻变革。

目前，随着信息技术的普及，各类计算机软硬件系统已在工程应用领域得到了广泛的应用。没有工程软件的支持，计算机硬件就失去了作用。工程软件在实际工程应用中占有相当重要的地位。工程软件所面向的应用领域主要有如下几个方面：

1. 工程辅助系统

(1) 工程数值计算 EC (Engineering Computation)。人类发明计算机的最初目的就是为了解决复杂的计算问题。在现代科学技术、航空航天等需要高精度、高难度、高效率计算的领域，计算机发挥了很重要的作用。在工程设计领域，数值计算技术不仅仅将人们从复杂的计算中解放了出来，更重要的是，使用计算机计算得更精确、更高效，效益非常明显。

(2) 计算机辅助设计 CAD (Computer-Aided Design)。CAD 是利用计算机帮助各类工程设计人员进行设计的工程软件系统。CAD 系统可以取代传统的从图纸设计到工艺流程编制工作和烦琐的手工计算过程，使设计速度大大加快，精度、质量大大提高。在飞机设计、建筑设计、机械设计、船舶设计、大规模集成电路设计等工程领域应用也非常广泛。

(3) 计算机辅助制造 CAM (Computer-Aided Manufacturing)。CAM 是利用计算机进行生产设备的管理、控制和操作的技术，它主要是面对生产制造过程。采用 CAM 技术可以提高产品质量、降低成本、缩短生产周期、降低劳动强度。

(4) 计算机辅助工程教育 CAEE (Computer-Aided Engineering Education)。CAEE 是利用计算机和网络进行远程工程培训和工程教育的技术。CAEE 提供了丰富的教学内容和直观的教学手段，大大提高了工程教育和培训的效率和效果。通过 CAEE 系统，计算机可以在一定程度上代替教师的角色，对学生进行辅助教学。有了计算机的参与，教学内容和教学形式更加多样化、形象化。国际互联网的发展对 CAEE 在工程教育领域中的普及发挥了不容忽视

的作用,网上工程教育实现了工程教育资源的共享,为工程教育教学机构和广大学习者提供了教学信息服务,同时,网上教育还带来了最新的办学模式(远程教育)和最新的教学方式(多媒体教学)。

(5) 计算机集成制造系统 CIMS (Computer Integrated Manufacturing System)。CIMS 是集设计、制造和管理三大功能于一体的现代化工厂生产系统。CIMS 系统可以提供对各类生产技术支持,是实现自动化工厂的核心技术。

(6) 计算机辅助测试 CAT (Computer-Aided Test)。CAT 是利用计算机获取和处理大批量的工程数据,完成各种复杂的测试工作的系统。通过 CAT 技术可以实现方便快捷的工作环境监测和对产品的测试工作,是有效提高产品的质量和精度的关键技术。

(7) 工业控制 IC (Industry Control)。IC 是用计算机实现工业过程控制的技术。在厂矿企业的自动化生产线和交通运输的控制等领域,计算机被用于进行数据采集、自动检测和过程控制。工业控制系统中使用了大量的多种形式的计算机硬件,如 PLC、工控机等,实际的控制流程则是由工业控制软件来实现的,它是自动化技术的重要组成部分。

(8) 计算机仿真 CS (Computer Simulation)。CS 是利用计算机模拟进行工程、产品、决策的试验,能在生产计划实施前即对未来的生产过程进行预测。CS 技术在工艺流程优化、成本核算和新产品研制等领域应用前景广阔。

2. 工程事务处理

工程事务分为事务、管理和决策三个层次。

(1) 工程事务型系统又称工程电子数据处理系统或工程业务信息系统,主要提供工程业务中处理日常事务的功能,例如工程文档的编辑与打印、报表的填写与统计、文档检索、活动安排,以及其他日常数据的处理等。

(2) 工程管理型系统又称工程管理信息系统 (Engineering Management Information System)。它是一个以计算机为基础,对企业实施全面管理的信息系统。例如计划管理系统、财务管理系统、人事管理系统、统计管理系统等。

(3) 工程决策型系统。它是在上述事务处理和信息管理的基础上,增加了决策辅助功能而构成的。在企业活动中,管理和决策都是企业领导层的基本职能。该系统可帮助领导人员选择适当的决策。

3. 现代工程通信及信息交流

计算机网络技术的高速发展,使得计算机日渐成为一种现代通信和信息交流的工具。通过支持计算机网络的工程软件不仅可以实现远程的数据传输、状态监控和现场资源配置等工作,而且还能异地共享和交流各种生产信息资源。工程软件与网络技术的结合扩大了工程软件的应用领域,提高了工程软件的应用水平,这也是工程软件发展的趋势。

工程软件依照具体的应用领域有不同的应用,故种类和数量众多。工程软件是现代工业的核心之一,它的广泛应用体现了生产技术的进步。因此,掌握工程软件的开发技术对于工程技术人员尤为重要。

工程软件的三个基本要素是数据、数据结构和算法,本章将分别对这些基本元素进行论述。

1-2 数据结构概述

1-2-1 数据结构及其数据运算的概念

早期的计算机主要用于数值计算,20世纪70年代出现微处理机后,计算机被大量用于信息处理,计算机的应用领域被拓宽。20世纪90年代计算机网络日益普及,发展至今已成为集计算、数据处理、数据传输于一体的信息处理系统。现代计算机应用强调的是数据管理,数据结构则是数据管理的前提。

数据结构是计算机科学的重要基础,近年来已发展成为一个专门学科。根据各种实际需求,人们提出了许多组织数据的方法,从而构造了不同的数据结构,而且随着实际问题需要的增加,新的更加复杂的数据结构还在不断地被提出。

首先介绍与数据结构相关的概念和技术术语。

1. 数据

数据是对客观事物的符号表示,在计算机科学中是指所有能输入到计算机中并被计算机程序处理的符号的总称。对计算机而言,数据的含义极为广泛,如数字、文字、图形、图像、声音等都是数据。数据与信息不同,数据是以机器可读的形式存储的信息,当计算机把这些数据处理成人们可以理解的形式时,数据就成为信息,信息是特殊的数据集。

2. 数据元素

数据元素是数据的基本单位,是数据集合中的个体,是对事物属性中基本方面的测量。在不同的应用背景下也把数据元素称为结点、记录、表目等。一个数据元素可由一个或多个数据项组成,数据项是具有独立含义的数据的最小单位,在计算机程序中它通常被作为一个整体来进行考虑和处理。

3. 数据项

数据项是具有独立意义的最小数据单位,是对数据元素属性的描述。数据项也被称为域或字段。如在产品档案管理系统中,可以把一个产品的有关信息作为一个数据元素,它由产品号、名称、生产日期等数据项组成。

4. 数据类型

数据类型是指某种程序设计语言所允许使用的变量种类。各种高级语言提供的基本数据类型有所不同,如C++提供了整型、实型和字符型等基本数据类型。除基本类型外,C++还允许数组型、结构型、联合型等构造类型。一个数据类型不仅定义了相应变量可以设定的值的集合和存储方法,而且还规定了对变量允许进行的一组运算及其规则。所以,可以把数据类型看作是程序设计语言中已实现了的数据结构。

5. 数据对象

数据对象是性质相同的数据元素的集合,是数据的一个子集。如所有的字母构成的集合 $C=\{\text{'A'}, \text{'B'}, \dots, \text{'Z'}\}$ 是一个字母数据对象;在产品质量档案管理系统中所有的产品构成的集合是一个产品数据对象。

6. 数据结构

数据结构是带有结构的元素的集合，它是指数据之间的相互关系，即数据的组织形式。可以看作是相互间存在特定关系的数据元素的集合。这些特定关系包括：

(1) 数据元素之间的逻辑关系，即数据的逻辑结构；

(2) 数据元素及其关系在计算机存储器中的存储方式，即数据的存储结构，也就是数据元素的物理结构；

(3) 施加在数据上的操作，即数据的运算。

计算机加工处理的数据元素不是互相孤立的，它们彼此之间应当存在某些逻辑上的关联，不然，对这些数据元素的处理就是无意义的。这些联系需要在对数据进行存储和加工时反映出来。任何事物都存在结构意义上的限定，因此，数据作为对事物的描述自然也是存在结构的。数据结构是存在一种或多种特定关系的数据元素的集合，外在表现为数据的组织形式。数据结构一般包括三个方面的内容，被称为数据的三要素：数据间的逻辑关系、数据在计算机中的存储关系以及在这些数据上定义的运算。

数据的逻辑结构是数据间关系的描述，它只抽象地反映数据元素间的逻辑关系，而不涉及其在计算机中的存在方式。数据的逻辑结构是从逻辑关系上描述数据，它与数据的存储形式无关，是独立于计算机的。因此，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。数据的逻辑结构分为线性和非线性结构。当数据元素之间的逻辑关系可以用一个线性序列表示出来时，则称为线性数据结构，否则称为非线性数据结构。

数据的存储结构是逻辑结构在计算机存储器中的实际表示，它不仅存储数据元素的内容，还要把数据元素间的关联体现出来，它是逻辑结构用计算机所能理解的方式的具体实现。存储结构是依赖于计算机程序的，对具体的程序而言，存储结构是具体的，我们只在高级语言的层次上来讨论存储结构，这使得数据的逻辑结构与存储结构具有了某种程度的同一性。存储结构主要分为顺序结构和链式结构。顺序结构以元素在存储器中的相对位置来表示元素间的逻辑关系，链式结构借助元素存储地址的指针来表示元素间的逻辑关系，这是它们在具体程序中的重要区别。此外，存储结构还有索引存储和散列存储等。

数据的运算是在数据上所施加的一系列操作，称为抽象运算。它只考虑这些操作的功能，而不考虑具体的操作步骤。只有在确定了存储结构后，才会具体实施这些操作，即抽象运算是以逻辑结构为基础的，具体的实现要在存储结构上来完成的。数据的运算是数据结构的一个重要成分。研究任何一种数据结构都离不开对该结构上的数据运算及算法设计的讨论。

算法是为了解决特定问题而采取的有限操作步骤，是对解题方案的准确而完整的描述。数据结构与算法之间存在着密切的关系。一方面算法的效率受到数据结构的影响，另一方面针对不同数据结构也会采用不同的算法。数据结构和算法是计算机软件的两个不可分割的方面，其内容将在本章重点叙述。

总之，无论怎样定义数据结构，都应将该数据的逻辑结构、数据的存储结构及数据的运算这三个方面看成一个整体，要体现它们之间的联系。

在计算机发展的早期，计算机主要用于科学计算，处理的是数值数据。当时的数据特点是数据元素之间的关系简单、数据量小、形式较为一致。因此，程序设计人员更加注重的是

算法，而不注重对数据的组织、性质、关系的研究。

随着计算机产业的飞速发展和计算机应用的日益普及，计算机已越来越广泛地应用于各种非数值处理问题，其应用已逐步深入到事务管理、工业控制、公共通讯、教育和军事等领域。待处理的数据形式多样，数量越来越大，关系越来越复杂。要想对这些数据进行快速有效的处理，就必须要了解数据的性质、组织方式和相互之间的关系。这样才能采用适当的方式进行存储，选用高效的算法进行处理。这正是数据结构所要研究的内容。也就是说，数据结构是一门对计算机所处理的数据的表示、组织和操作进行系统研究的学科。

在计算机科学中，数据结构与算法是密不可分，缺一不可的。算法指的是对计算机问题的全过程及具体步骤的描述，计算机程序正是按照算法所描述的步骤对某种结构的数据进行加工处理的。如果没有数据结构，计算机就缺少了处理的基础；而没有算法，计算机就缺少了求解问题的方法。因此，著名的计算机科学家 N.Wirth 在论述算法与数据结构之间的关系时，就曾指出：

程序=数据结构+算法

这一公式强调了数据结构和算法之间的必然联系。

数据结构往往是与施加于其上的运算密切相关的。数据按一定的逻辑结构组织起来，再把数据结构用适当的存储方式存储到计算机中，其目的就是为了提高数据的运算效率，从而更有效地处理数据。各种逻辑结构有其相应的运算，每种逻辑结构有一个运算的集合。数据的运算是定义在数据的逻辑结构上的，运算的具体实现要在存储结构上进行。对于各类数据结构而言，它们在基本运算方面是相似的，最常用的运算有：

- 插入：其含义是在数据结构中增加新的结点；
- 删除：其含义是把指定的结点从数据结构中去除；
- 更新：其含义是改变指定结点的值或改变指定的某些结点之间的关系；
- 查找：查找的含义是在数据结构中查找满足一定条件的结点；
- 排序：排序的含义是对数据结构中各个结点按指定数据项的值，以升序或降序重新排列。

复杂的运算过程可以是以上各种基本操作的组合。

在工程软件系统中，往往要构建各类数据结构，以描述复杂的工程对象。数据结构的描述通常采用的是面向对象的数据类（Class）的描述方法。类不仅可以完整地说明数据的结构，通过类的封装机制还能有效地保护数据，降低数据结构的界面复杂度，并提高数据结构的利用率。更为重要的是，类还包含了对所描述的数据结构的一系列操作，使得数据结构及其操作运算完美地结合在了一起。

1-2-2 数据结构的分类

依据数据元素间关系的特点，数据结构可分为两大类：线性结构和非线性结构。

如果一个非空的数据结构满足下列两个条件：

- 有且有一个根结点；
- 每一个结点最多有一个前驱，最多有一个后继。

则称该数据结构为线性结构。在一个线性结构中，若插入或删除任何一个结点后不满足

以上的两个条件，则被操作后的数据结构就不是线性结构。线性结构可构成线性表，在该结构中所有的数据元素都按某种次序排列在一个序列中，如图 1-1 所示。

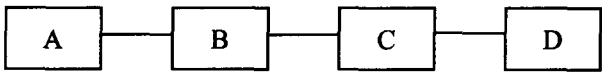


图 1-1 线性结构中数据元素间的线性关系

根据对线性结构中数据元素的存取方式的不同，又可将其分为直接存取结构、顺序存取结构和广义索引结构。对于直接存取结构，可以直接存取某一特定数据元素而不需先访问其前驱。数组、记录等都属于这一类。对于顺序存取结构，只能从数据序列中的第一个数据元素开始，按顺序依次查找，直到找到指定的元素。顺序表和链接表就属于该类型。广义索引与数组有类似之处，数组是依靠其下标进行索引，而广义索引是通过关键码进行索引的。通过设定数据记录中某一数据项或某一组数据项为关键码，通过关键码来索引记录。索引表和散列表都属于该类。

在非线性结构中各个数据元素不必保持在一个线性序列中，每个数据元素可能与多个其他数据元素发生关联。在非线性结构中，各数据元素之间的前驱与后继的关系要比线性结构复杂，因此，对非线性结构的存储与处理比线性结构要复杂得多。

线性结构与非线性结构都可以为空数据结构。一个空的数据结构究竟属于线性结构还是属于非线性结构要根据具体情况来定。如果对该数据结构的运算是按线性结构的规则来处理的，则属于线性结构；否则属于非线性结构。

根据前驱与后继关联的不同，可进一步分为层次结构和群结构。

层次结构是按层次划分的数据元素的集合，其特点是：每一确定层次上的元素可以有多个的直接下层子元素。典型的层次结构是树型结构，如图 1-2 所示。

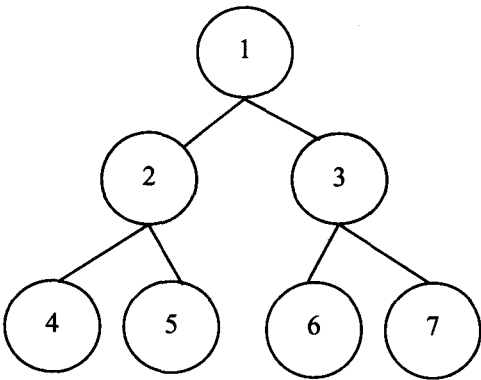


图 1-2 典型的树型结构

群结构中所有元素之间无顺序关系，数据元素通过特定的关系联系在一起。这类结构中 包含集合、图结构等，典型的图结构如图 1-3 所示。

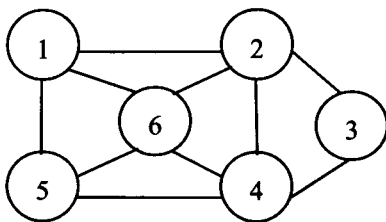


图 1-3 典型的图结构

此外，数据结构还可分为逻辑结构和存储结构（即物理结构）。数据的逻辑结构是指从解决问题的角度出发，为实现必要的功能所建立的数据结构。它属于用户的逻辑视图，是面向具体问题的。数据的存储结构是指数据在计算机中存放的方式，它是数据逻辑结构的物理存储方式，是属于具体实现的物理视图，是面向具体计算机的。数据的逻辑结构根据问题所要实现的功能来建立，数据的存储结构是根据解决问题所要求的响应速度、处理时间、修改时间、存储空间和单位时间的处理量等建立起来的，是逻辑数据的存储映像。

1-2-3 数据结构的表示

1. 数据的逻辑结构

对数据间关系的描述即数据的逻辑结构，形式上可以用一个二元组来表示，即：

$$B=(K, R)$$

其中 K 为结点的有穷集合，即 K 是由有限个结点所构成的集合； R 是 K 上的有穷集合，即 R 是由有限关系所构成的集合，其中的每个关系都是从 K 到 K 的关系。

结点的类型是指结点的值的组成方式，即数据的类型。设 K 是结点的有穷集合且每个结点 k ($k \in K$) 的值可由 n 个字段组成，第 i 个字段记为 W_{ik} ，所以 k 可以表示为：

$$k=(W_{1k}, W_{2k}, W_{3k}, \dots, W_{nk}), n \geq 1$$

其中每个 W_{ik} 是一个组合项，它可由一个或多个初等项或组合项组成，也可为一个初等项。

结点可以分为两大类型，即初等类型和组合类型。只包含一个初等项的结点属于初等类型。初等类型有 5 种基本的数据类型，即整数类型、实数类型、布尔类型、字符类型及指针类型。按指针的地址操作类型可分为绝对地址指针或相对地址指针。组合类型是由初等类型以某种方式组合而成的。

数据的逻辑结构可分为线性数据结构和非线性数据结构。数据结构里有且仅有一个终端结点和一个开始结点，所有结点都最多只有一个前驱和一个后继，这种数据结构称为线性数据结构，典型的线性数据结构为线性表。其他数据结构为非线性数据结构，典型的非线性数据结构有树型结构和图结构。

数据结构和结点之间具有整体与局部概念上的关联。数据结构由若干结点及其相互关系所组成。把结点看作一个整体，重点讨论结点之间的关系。当研究结点的内部结构时，有时可以把一个结点看作一个数据结构，而把组成该结点的数据项看成数据结构中的结点，但这

时各结点可以是相同类型的，也可以是不同类型的。

2. 数据的存储结构

数据的存储结构与数据的物理存储方式是密切相关的。实际上，计算机的存储器具有有限个存储单元，每个单元用惟一的地址作为操作标识，各存储单元的地址是连续编码的。数据的逻辑结构向存储结构的转换是通过建立一个从结点到地址单元的映像来实现的，存储结构要在物理层次上映像出数据结点间的关系。

有4种基本的物理存储映像方法：顺序方法、链接方法、索引方法和散列方法。现分别介绍如下：

(1) 顺序方法

所有数据元素存放在一片连续的存储单元中，逻辑上相邻的元素存放在计算机内仍然相邻。顺序存储方法主要用于线性数据结构。把逻辑上相邻的结点存储在物理上相邻的存储单元内，结点之间的关系由存储单元的连接关系来体现。每个结点所占的存储单元可以相同，也可以不同。但无论所占单元相同或不同，总体上都是一个一个地填满整个存储区域。

(2) 链接方法

所有元素存放在可以不连续的存储单元中，但元素之间的关系可以通过地址确定，逻辑上相邻的元素存放到计算机内存后不一定是相邻的。链接方法最大的特点就是每个数据结点都附有指针字段，即结点所占的存储单元分为两部分：存放结点本身的信息和存入后继结点的存储单元的一个或多个地址。通常链接的最后一个结点的指针可以通过置0而终结。

(3) 索引方法

使用该方法在存放元素的同时，还需建立一个附加的索引表。索引表中的每一项称为索引项，索引项的一般形式为：

(关键字，地址)

其中的关键字是能惟一标识一个结点的数据项。

在线性结构里，结点可以排成序列 $k_1, k_2, \dots, k_n$ ，每个 k_i 在序列里都有对应的位置 i ，这个位置为结点的索引。用结点的索引号 i 来确定结点的存储地址有两种方法：

- 建立附加的索引表，索引表中第 i 项的值就是第 i 个结点的存储地址；
- 每个结点所占单元数相等时，可用结点位置值指出结点的存储地址，即第 i 个结点 k_i 的地址为：

$$LOCATION(k_i) = (i - 1) \cdot p + q$$

其中 p 为结点所占单元数， q 为 k_1 的存储地址。

(4) 散列方法

通过构造散列函数，用函数的值来确定元素存放的地址。根据结点的值来确定它的存储地址。散列方法又称为关键码—地址转移法。在结点中取一个或几个字段的值 W_{ik} 做关键码，结点 k 的存储地址为：

$$LOCATION(k) = f(W_{ik})$$

其中 f 为散列函数。

在散列方法中关键的问题是要恰当地选择散列函数以解决存储地址碰撞的问题。

一个数据结构的存储映像都是以上 4 种基本映像之一或者是它们的组合。一个逻辑结构可以有不同的映像方法。逻辑结构相同但存储结构不同视为不同的数据结构,例如,线性表为一逻辑结构,顺序存储时为顺序表,链接存储时称为线性链表。到底选择什么存储方式,取决于对数据运算的方便性和可利用的存储资源等诸多方面的因素。

数据结构的描述有两类方法。一类是逻辑描述法,即采用数学或图形描述的方法,这类方法主要用于数据逻辑结构的描述;另一类方法是采用具有面向对象特性的算法语言来描述,采用诸如 Effel、C++ 等算法语言既能描述数据的逻辑结构,同时还能描述数据结构的具体实现过程。

1-2-4 数据类型及数据抽象

在数据结构中往往涉及数据类型 (Data type) 的概念,虽然数据类型与数据结构本质上相同,如数据结构中的数组与数据类型中的数组,数据结构中的串与数据类型中的串都有同一性,但二者的意义却不相同。简单地讲,数据结构定义了一组按某种关系结合在一起的数据元素,重点是研究数据元素之间的关系。数据类型则最早出现在高级算法语言中,用以刻画数据操作对象的特性。每一种高级算法语言都定义有自己的数据类型,在用高级算法语言编写的程序中,每个变量、常量或表达式都有一个它所属的确定的数据类型。数据类型明显或隐含地规定了在程序执行期间变量或表达式所有可能取值的范围。因此,数据类型是一个值的集合和定义在这个值的集合上的一组操作的总称。例如:在 C++ 语言中,整数类型 (int) 的值集范围是 $[-32768, 32767]$,主要运算有 $+$ 、 $-$ 、 $*$ 、 $/$ 、 $\%$ 等。从硬件角度看,它们的实现涉及到“字”、“字节”、“位”、“位的运算”等。但从用户角度看,并不需要了解整数在计算机内是如何表示的,运算细节是如何实现的。用户只需了解整数的加法或取模运算的抽象特性。而不必了解“位运算”细节,就可运用高级算法语言进行程序设计。可以说,数据类型是数据结构在算法语言中的实现。

数据类型可分为简单类型和结构类型两种。简单类型中的每个数据都是无法再分割的整体,数据值内部就无结构可言,如一个整数、实数、字符、指针、枚举量等都是无法再分割的整体,所以它们所属的类型均为简单类型。结构类型由简单类型按照一定的规则构造而成,其数据结构就是相应元素的集合(它实际上是该结构类型中的一个值)和元素之间所含关系的集合,并且结构类型中可以包含结构类型。所以结构类型中的数据可以分解为若干个简单数据或结构数据,每个结构数据仍可再分。例如,在 C++ 语言中的数组是一种结构类型,它由固定个数的同种数据类型顺序排列而成。

数据类型也可被定义为一种数据结构和能够对该数据结构进行的操作的集合。对于简单类型,其数据结构就是相应取值范围内的所有数据,每一个数据值是不可分的独立整体,因而数据值内部就无结构可言。对于结构类型,其数据结构就是相应元素的集合和元素之间所含关系的集合。

总之,数据结构是指计算机处理的数据元素的组织形式和相互关系,而数据类型是某种程序设计语言中已实现的数据结构。在程序设计语言提供的数据类型的支持下,就可以