

“十三五”高等学校规划教材



工程软件开发 技术基础

GONGCHENG RUANJIAN KAIFA JISHU JICHU

臧铁钢 朱海华 主 编
梁睿君 冷 晟 马万太 副主编

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

“十三五”高等学校规划教材

工程软件开发技术基础

臧铁钢 朱海华 主 编
梁睿君 冷 晟 马万太 副主编

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和研究人员,以及高等学校工程专业类的学生对工程软件开发技术的需要而编写,主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的开发过程。

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,适合作为高等学校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

工程软件开发技术基础 / 臧铁钢, 朱海华主编. —
北京: 中国铁道出版社, 2017. 1
“十三五”高等学校规划教材
ISBN 978-7-113-19096-5

I. ①工… II. ①臧… ②朱… III. ①软件开发—高等学校—教材 IV. ①TP311.52

中国版本图书馆 CIP 数据核字 (2016) 第 325321 号

书 名: 工程软件开发技术基础
作 者: 臧铁钢 朱海华 主编

策 划: 翟玉峰

读者热线: (010) 63550836

责任编辑: 翟玉峰 徐盼欣

封面设计: 刘 颖

封面制作: 白 雪

责任校对: 张玉华

责任印制: 郭向伟

出版发行: 中国铁道出版社 (100054, 北京市西城区右安门西街 8 号)

网 址: [http:// www.51eds.com](http://www.51eds.com)

印 刷: 北京市燕鑫印刷有限公司

版 次: 2017 年 1 月第 1 版 2017 年 1 月第 1 次印刷

开 本: 787 mm×1 092 mm 1/16 印张: 19 字数: 413 千

印 数: 1~2 000 册

书 号: ISBN 978-7-113-19096-5

定 价: 45.00 元

版权所有 侵权必究

凡购买铁道版图书, 如有印制质量问题, 请与本社教材图书营销部联系调换。电话: (010) 63550836

打击盗版举报电话: (010) 51873659



Foreword

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。计算机技术是硬件技术和软件技术的结合，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都是相似的。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。本书主要论述了以下几个方面的内容：

(1) 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。

(2) 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章介绍了常用数据处理技术及其在工程中的应用。

(3) 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人员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。

(4) 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。

(5) 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、朱海华任主编，由梁睿君、冷晟、马万太任副主编。其中，臧铁钢编写第1、4章，朱海华编写第2、3、5章，梁睿君和冷晟编写第6章，马万太编写第7章，臧铁钢统稿。本书的编写与出版得到了中国铁道出版社的支持和帮助，在此表示诚挚的感谢。

由于编者水平有限，书中难免存在疏漏和不妥之处，恳切希望广大读者批评指正。

编 者

2016年10月

于南京航空航天大学



Contents

目 录

第1章 工程软件的基础元素	1	2.1.6 串	55
1.1 工程软件概述	1	2.2 非线性数据结构及其工程	
1.2 数据结构概述	3	应用	61
1.2.1 数据结构及其数据运算的		2.2.1 多维数组	61
概念	3	2.2.2 树与二叉树	68
1.2.2 数据结构的分类	6	2.2.3 图	78
1.2.3 数据结构的表示	8	小结	87
1.2.4 数据类型及数据抽象	10	习题	87
1.3 算法概述	12	第3章 常用数据处理技术及其在	
1.3.1 算法的概念	12	工程中的应用	90
1.3.2 算法的描述	13	3.1 查找技术	90
1.3.3 算法分析	14	3.1.1 顺序查找	91
小结	17	3.1.2 二分查找	91
习题	17	3.1.3 分块查找	93
第2章 常用数据结构及其在工程中的		3.1.4 二叉排序树查找	94
应用	18	3.1.5 Hash表技术	97
2.1 线性数据结构及其工程应用	18	3.2 排序技术	101
2.1.1 顺序表	19	3.2.1 互换排序	102
2.1.2 线性链表	23	3.2.2 插入排序	105
2.1.3 索引存储	35	3.2.3 选择排序	107
2.1.4 栈	39	3.2.4 归并排序	109
2.1.5 队列	49	3.2.5 基数排序	111

3.2.6 排序方法的比较	111	习题	165
3.2.7 排序技术在软件工程 中的应用	112	第5章 数据库技术	166
小结	112	5.1 数据库技术概述	166
习题	112	5.1.1 数据管理与数据库 技术	166
第4章 操作系统原理	114	5.1.2 数据库系统的结构	174
4.1 操作系统简介	114	5.2 数据模型	181
4.1.1 操作系统的概念	114	5.2.1 信息的三种世界及其 描述	181
4.1.2 操作系统的类型与 功能	118	5.2.2 实体-联系模型	183
4.2 进程管理	123	5.2.3 层次模型	185
4.2.1 进程的概念	123	5.2.4 网状模型	186
4.2.2 进程控制	125	5.2.5 面向对象模型	187
4.2.3 进程的同步、互斥和 通信	128	5.2.6 关系模型	188
4.2.4 进程调度	135	5.3 关系化与规范化	191
4.2.5 死锁	138	5.3.1 关系代数及查询优化	192
4.3 存储管理	140	5.3.2 关系模式的规范化	201
4.3.1 存储管理概述	140	5.4 分布式数据库	217
4.3.2 常用存储管理技术	143	5.4.1 分布式数据库概述	217
4.4 文件管理	148	5.4.2 分布式数据库管理 系统	218
4.4.1 文件管理概述	148	5.4.3 分布式数据库设计	218
4.4.2 文件系统管理方式	151	5.5 工程数据库设计	220
4.5 设备管理	156	5.5.1 数据库设计的基本 概念	220
4.5.1 设备管理的概念	156	5.5.2 工程数据库设计的基本 步骤	221
4.5.2 设备管理结构	159	5.5.3 工程数据字典	227
4.5.3 缓冲技术	160	5.6 工程数据库的新技术与 新应用	229
4.5.4 设备分配	161		
4.5.5 虚拟设备	163		
小结	164		

5.6.1 WWW数据库	229	6.4.3 工程软件的测试方法	266
5.6.2 多媒体数据库	231	6.5 工程软件质量评价与软件	
5.6.3 移动数据库	231	质量保证	272
5.6.4 地理信息系统	232	6.5.1 工程软件质量标准	272
5.6.5 数字图书馆	232	6.5.2 工程软件质量评价	273
小结	233	6.5.3 工程软件质量保证	275
习题	233	6.6 工程软件新技术	276
第6章 工程软件开发技术	235	6.6.1 Java EE技术	276
6.1 软件工程	235	6.6.2 软件复用	280
6.1.1 软件	235	小结	281
6.1.2 软件工程的概念	238	习题	281
6.1.3 软件支持环境	239	第7章 工程软件实例——人力资源	
6.2 工程软件设计准则	240	管理系统的开发	282
6.2.1 抽象准则	240	7.1 项目背景和需求	282
6.2.2 结构化准则	241	7.1.1 项目背景分析	282
6.2.3 模块化准则	241	7.1.2 用户需求	283
6.2.4 信息隐藏和局部化		7.1.3 用户需求分析	283
准则	242	7.2 系统设计	284
6.3 工程软件设计的基本技术	243	7.2.1 运行方案	284
6.3.1 全生命周期法	243	7.2.2 系统总体结构	285
6.3.2 原型法	244	7.2.3 数据流图	285
6.3.3 面向数据的设计法	245	7.2.4 实体—关系设计	286
6.3.4 面向对象的设计方法	251	7.2.5 状态—迁移模型	288
6.3.5 统一建模语言 (UML)	257	7.3 系统实现	290
6.4 工程软件的测试技术	263	7.3.1 数据库设计	290
6.4.1 工程软件测试的概念	263	7.3.2 程序设计	294
6.4.2 工程软件的测试过程	264	小结	296

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，在工程界，科学计算也是计算机最早应用的领域；计算机的应用范围扩大到工程数据信息处理后，就一直都是计算机工程应用的重要领域，现在已成为计算机工程应用不可缺少的部分；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过程控制之中。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。在工业4.0和中国制造2025技术体系中，计算机和网络是基本的硬件基础，在其上是大量的工程软件，用以支持高效率的大规模定制需求。

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。常见的工程软件有如下几种系统：

1. 工程辅助系统

(1) 工程数值计算 (Engineering Computation, EC)。人类发明计算机的最初目的就是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 (Computer-Aided Design, CAD) 和计算机辅助工艺过程设计 (Computer Aided Process Planning, CAPP)。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非常广泛。CAPP是指借助于计算机进行数值计算、逻辑判断和推理等功能来制定零件机械加工工艺过程。借助于CAPP系统,可以解决手工工艺设计效率低、一致性差、质量不稳定、不易达到优化等问题。

(3) 计算机辅助制造(Computer-Aided Manufacturing, CAM)。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。在机械加工过程中用到的数控加工程序就是一种低层CAM系统的一部分。

(4) 计算机辅助工程教育(Computer-Aided Engineering Education, CAEE)。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。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(Computer Integrated Manufacturing System, CIMS)。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可以提供对各类生产技术的支

持,是实现自动化工厂的核心技术。

(6) 计算机辅助测试(Computer-Aided Test, CAT)。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(Industry Control, IC)。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航运的控制等领域,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(Computer Simulation, CS)。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,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。

企业资源规划(Enterprise Resource Planning, ERP)系统是一类典型的工程事务软件。ERP的功能包括制造、供销、财务管理，以及多工厂(车间)管理、质量管理、实验室管理、设备维修管理、仓库管理、运输管理、过程控制接口、数据采集接口、企业电子邮件管理、法规与标准索引、项目管理、金融投资管理、市场信息管理等。

随着智能工厂技术的逐步深入研究与应用，工程事务处理过程的智能化已成为核心技术之一，相应的工程软件也正得到重视。

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，而且可以异地共享和交流各种生产信息资源。目前，物联网技术正在工业领域得到推广应用，这是网络技术与计算机深度结合的结果。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。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，并且广泛应用于工业领域。现代计算机软件强调的是数据管理，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。

3. 数据项

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，是对数据元素属性的描述。数据项也被称为域或字段。例如，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。各种高级语言所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规则。所以，可以把数据类型看作程序设计语言中已实现了结构化的数据类别。

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。例如，所有的字母构成的集合 $C= \{ 'A', 'B', \dots, 'Z' \}$ 是一个字母数据对象；在产品质量档案管理系统中由某类产品构成的集合是一个产品数据对象，如传动结构对象中就包含了主动轴和从动轴等零件。

6. 数据结构

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载体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。数据结构是带有经过组织，并形成特定结构的元素的集合，它体现了数据之间的相互关系，即体现了数据的组织形式。例如，在描述轴系零件的装配关系时，可能会用到堆栈结构，可以描述机器的装拆顺序。总之，数据结构一般包括三方面的内容，也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。

(1) 数据元素之间的逻辑关系，即数据的逻辑结构。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。逻辑结构总是和所要解决的问题相对应，并不涉及数据元素在计算机存储器中的具体存

放状态，即数据的逻辑结构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从具体问题抽象出来的数据模型。数据的逻辑结构分为线性与非线性结构。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。

(2) 数据元素在计算机存储器中的物理存储结构，体现的是数据元素在物理存储空间中的实际分布情况。数据的物理存储结构不仅要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。

(3) 施加在数据上的操作，即数据的运算。一种特定的数据结构其实已经限定了施加在其上的运算操作方式。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。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的数据运算及算法的讨论。

算法是为了解决特定问题而采取的有限运算步骤，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，其应用已逐步深入事务管理、工业控制、公共通信、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：

$$\text{程序} = \text{数据结构} + \text{算法}$$

这一公式揭示了数据结构和算法之间的天然联系。

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，其目的就是实现运算过程和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，即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。对于数据本身而言，有各种具体的数学运算；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：

- ① 插入：其含义是在数据结构中增加新的结点。
- ② 删除：其含义是把指定的结点从数据结构中去除。
- ③ 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。
- ④ 查找：其含义是在数据结构中查找满足一定条件的结点。
- ⑤ 排序：其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。

针对数据结构的复杂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（Data Class）的描述方法。类不仅可以完整地说明数据的结构，而且通过类的封装机制能有效地保护数据，降低数据结构的界面复杂度，并使得数据结构能以开发者易用的形式被使用。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一起。

1.2.2 数据结构的分类

数据结构常被用于描述被研究对象的空间或逻辑结构，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：

- (1) 有且只有一个根结点。
- (2) 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。直观上看，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。在实际软件设计中，线性结构可用于描述有先后顺序的事物。

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

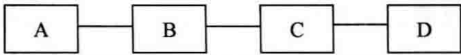


图1-1 线性结构中数据元素间的线性关系

数据元素而不需先访问其前驱。数组、记录等都属于这一类，只要知道数据元素的标号就能直接访问该元素。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，只有知道表的首地址才能依次访问各元素。广义索引与数组有类似之处，数组是依靠其下标进行索引的，而广义索引是通过关键码进行索引的。广义索引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。在描述工程对象时，数据的线性结构被广泛采用。例如，在描述生产流程时，各工序就可以被描述为一个线性表。

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的复杂，因此对非线性结构的存储与处理比线性结构要复杂得多。工程软件中对工程对象结构的描述往往可以采用一种称为树的非线性数据结构进行描述。例如，对机器的机械结构和控制系统结构的描述就是典型的树形结构。

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。

非线性数据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于一个的直接下层子元素。典型的层次结构是树形结构，如图1-2所示。

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。例如，在分析机器各功能部分的相互作用关系时，就会采用群结构。

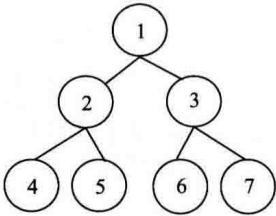


图1-2 典型的树形结构

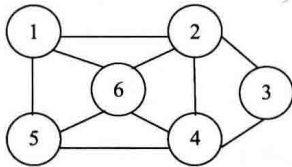


图1-3 典型的图结构

此外，从数据的存在方式来看，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，为实现必要的运算所建立的基于逻辑的数据结构，它属于用户的逻辑视图，是面向具体工程问题的。数据的存储结构是指数据在计算机中存放的方式，它是数据逻辑结构向物理存储结构的映射，是属于具体计算机存储空间

视图。数据的逻辑结构根据问题所要实现的运算来建立，数据的存储结构的选择是根据解决问题所要求的响应速度、处理效率、存储空间利用率等因素权衡的结果，是逻辑数据的实际存储映像。同一种类型的数据结构可以根据需要采用不同的存储结构。

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对数据逻辑结构的描述，形式上可以用一个二元组来表示，即

$$B = (D, R)$$

其中， D 为结点的有穷集合，即 D 是由有限个结点所构成的集合； R 是体现结点间关系的有穷集合，即 R 是由有限关系所构成的集合，其中的每个关系都是从 D 集合中一个结点到另一个结点的关系。

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

$$d = (W_{1d}, W_{2d}, W_{3d}, \dots, W_{nd}), n \geq 1$$

其中，每个 W_{id} 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，即整型、实型、布尔型、字符型及指针型。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，如枚举类型和结构类型等。

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。一般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，每个单元用唯一的地址作为操作标识，各存储单元的地址是人为连续编码的。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射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。一种具有特定逻辑结构的数据结构可以根据需要采用多种物理存储映像方法来表达。

1) 顺序方法

使用顺序方法存储数据，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

使用链接方法存储数据, 所有元素可以存放在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

使用索引方法存储数据, 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, 关键字是能唯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 $d_1, d_2, \dots, d_n$, 每个 d_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:

(1) 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。

(2) 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 d_i 的地址为

$$\text{LOCATION}(d_i) = (i-1) \cdot p + q$$

其中, p 为结点所占单元数, q 为 d_1 的存储地址。

4) 散列方法

使用散列方法存储数据, 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。散列方法又称关键码-地址转移法。在结点中取一个或几个字段的值 W_{id} 做关键码, 结点 d 的存储地址为

$$\text{LOCATION}(d) = f(W_{id})$$

其中, f 为散列函数。

散列方法关键的问题是要恰当地选择散列函数以解决存储地址冲突的问题。

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它们的组合。一个逻辑结构可以有不同的存储映像方法。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。具体选择什么存储方式, 取决于算法和计算机资源等约束条件。

目前, 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; 另一类是采用具有面向对象特性的算法语言来描述, 采用诸