

高等院校信息技术规划教材

数据结构与数据库 应用教程

于秀丽 编著



清华大学出版社

高等院校信息技术规划教材

数据结构与数据库 应用教程

于秀丽 编著

清华大学出版社
北京

内 容 简 介

本书是为“数据结构与数据库”课程编写的教材,也可作为学习数据结构与数据库技术的参考教材。

本书的前半部分为数据结构,包括线性表、栈、队列、串、数组、树和图等,以及查找和排序等操作;后半部分为数据库技术,包括数据库系统概述、关系模型与关系代数,关系数据库标准语言(SQL)、数据库设计及优化、数据库安全性与完整性、事务管理与恢复等,最后以一个综合实例介绍了数据库应用系统的开发过程。

本书概念清楚,重点突出,内容丰富,结构合理,思路清晰,示例翔实,每章后均附有习题。本书主要面向数据结构与数据库初学者,可作为信息管理与信息系统、计算机及相关专业的本科教学教材,也可供自学计算机基础知识的读者参考。

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

数据结构与数据库应用教程/于秀丽编著. —北京:清华大学出版社,2019

(高等院校信息技术规划教材)

ISBN 978-7-302-51422-0

I. ①数… II. ①于… III. ①数据结构—高等学校—教材 ②关系数据库系统—高等学校—教材 IV. ①TP311.1

中国版本图书馆 CIP 数据核字(2018)第 242139 号

责任编辑:张 玥 薛 阳

封面设计:常雪影

责任校对:焦丽丽

责任印制:丛怀宇

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

课件下载: <http://www.tup.com.cn>, 010-62795954

印装者:北京密云胶印厂

经 销:全国新华书店

开 本:185mm×260mm 印 张:18.75

版 次:2019 年 1 月第 1 版

定 价:45.00 元

字
数
印



次:2019 年 1 月第 1 次印刷

产品编号:078463-01

前 言

数据结构和数据库技术是信息技术的重要理论技术基础,不仅是高等学校计算机科学与技术类专业学生必修的两门专业基础课程,而且已成为非计算机专业的热门选修课。目前,有关数据结构和数据库技术的书籍有很多。随着课程建设的改革、课时的缩减,如何能使学生在有限的课时里更好地掌握这两门课程,并能在实际的软件开发过程中自觉地应用,一直是摆在广大教师面前的课题。本书结合目前教学的实际情况,梳理了对数据结构与数据库要求的知识点,并形成了便于学习和掌握的相应知识单元,通过大量案例来解释相关的原理及应用技术,注重学生实践能力的培养,内容通俗易懂。本书既可以作为信息管理与信息系统、计算机及相关专业的本科教材,也可供自学计算机基础知识的读者参考。

全书共 15 章,分为两大部分。前 8 章作为第一部分,系统地介绍了数据结构的相关理论及应用。第 1 章为数据结构绪论,主要介绍数据结构的相关概念和术语、算法的描述与分析方法;第 2 章为线性表,主要介绍了顺序表和链表的存储表示与实现;第 3 章为特殊线性表,主要介绍了栈、队列和串的存储表示与实现;第 4 章为数组,主要介绍了数组的存储表示与实现;第 5 章为树与二叉树,主要介绍了二叉树的基本知识、性质、存储及遍历应用;第 6 章为图,主要介绍了图的基本概念、存储及遍历应用;第 7 章为查找,主要介绍了静态查找、动态查找和哈希表;第 8 章为排序,主要介绍了直接插入排序、希尔排序、冒泡排序、快速排序、选择排序和归并排序等几种常用的排序算法及性能。第 9~15 章是第二部分,主要介绍的是数据库技术及其应用。第 9 章为数据库系统概述,主要介绍了数据、数据库、数据库管理系统和数据库系统等基本概念,数据库处理技术的发展,同时也介绍了数据模型、数据抽象、数据库模式等概念;第 10 章为关系模型与关系代数,主要介绍了关系数据库实现的基本理论、关系的定义和性质及专门的关系运算方法;第 11 章为关系数据库的标准语言——SQL,包括数据定义语言、数据控制语言和数据操纵语言;第 12 章为数据库设计及优化,主要介绍了数据库建模方法,包括概念模型设计过程、如何将概念模型转换为关系模型及关系模式规范化理论等;第 13 章为数据库安全性与完整性,主要包括数据库安全性、完整性的基本概念和措施,游标、存储过程和触发器的使用;第 14 章阐述了事务管理和恢复的相关技术,主要包括事务的概念、特性和并发控制,数据库的恢复与备份等;第 15 章以一个综合实例介绍了数据库应用系统的开发过程。内容讲解由浅入深,层次清晰,通俗易懂。

本书具有以下特点。

(1) 本书面向应用型本科高校,根据相关专业的培养方案,服务于应用型和技能型的高级实用人才,结合该课程的先行课程和后续课程,组织相关知识点与内容。本书结构严谨,内容安排环环相扣,符合初学者的学习习惯。

(2) 吸取了同类教材的优点,注重理论和实践相结合。在知识点组织和案例设计等

内容安排上,既着眼于培养学生熟练掌握理论知识,又注意锻炼和培养学生程序设计过程中分析问题和解决问题的实践动手能力,启发学生的创新意识,使学生的理论知识水平和实践技能得到全面提升。

(3) 每个知识点都包括基础案例,知识内容层层推进,将知识点有机地串联在一起,使得学生易于接受和掌握相关知识内容。

(4) 提供配套的课件、例题和课后习题的参考答案。

本书由于秀丽编写。在本书的编写过程中,参阅了大量的参考书目和文献资料,在此向参考资料的作者表示由衷的感谢。在本书的出版过程中,得到了曹妍教授、陈佳教授和王旭坪教授的支持和帮助,还得到了清华大学出版社的大力支持,在此表示诚挚的感谢。此书的出版离不开我家人的鼓励和照顾,感谢他们的默默奉献。

由于作者水平有限,书中难免有不足和疏漏之处,敬请读者批评指正。

编 者

2018 年 2 月

目 录

第一部分 数据结构

第 1 章 绪论	3
1.1 数据结构的概念	3
1.1.1 数据结构的范畴	3
1.1.2 相关概念和术语	4
1.2 算法和算法分析	7
1.2.1 算法的基本概念	7
1.2.2 算法复杂度	11
小结	13
习题	14
第 2 章 线性表	15
2.1 线性表的逻辑结构	15
2.1.1 线性表的定义	15
2.1.2 线性表的基本操作	16
2.2 线性表的顺序存储及运算实现	17
2.2.1 顺序存储的特点	17
2.2.2 顺序表上的运算实现	17
2.3 线性表的链式存储及运算实现	21
2.3.1 链式存储的特点	22
2.3.2 链表上的运算实现	24
小结	26
习题	27
第 3 章 特殊线性表	28
3.1 栈	28
3.1.1 栈的定义	28
3.1.2 栈的存储及运算实现	29
3.2 队列	31
3.2.1 队列的定义	31
3.2.2 队列的存储及运算实现	33
3.3 串	35

3.3.1 串的定义	35
3.3.2 串的存储	37
小结	37
习题	38
第4章 数组	39
4.1 数组的定义	39
4.2 数组的存储及运算实现	40
小结	42
习题	42
第5章 树与二叉树	43
5.1 树	43
5.1.1 树的定义	43
5.1.2 相关术语	44
5.2 二叉树	45
5.2.1 二叉树的定义	45
5.2.2 二叉树的性质	46
5.2.3 二叉树的存储结构	47
5.3 二叉树的遍历	48
小结	50
习题	50
第6章 图	51
6.1 图的定义和术语	51
6.2 图的存储表示	53
6.3 图的遍历	55
小结	57
习题	58
第7章 查找	59
7.1 基本概念	59
7.2 静态查找表	60
7.2.1 顺序查找	60
7.2.2 折半查找	61
7.2.3 索引查找	62
7.3 动态查找表	63
7.3.1 二叉排序树	64

7.3.2 平衡二叉树	66
7.4 哈希表的查找	66
小结	69
习题	69
第8章 排序	70
8.1 基本概念	70
8.2 插入排序	71
8.2.1 直接插入排序	71
8.2.2 希尔排序	73
8.3 交换排序	74
8.3.1 冒泡排序	74
8.3.2 快速排序	76
8.4 选择排序	78
8.5 归并排序	79
小结	81
习题	82

第二部分 数据库技术

第9章 数据库系统概述	85
9.1 数据库系统的作用	85
9.1.1 数据与数据管理	85
9.1.2 数据库应用	88
9.2 数据库处理技术的发展过程	91
9.2.1 人工管理阶段	91
9.2.2 文件系统阶段	92
9.2.3 数据库系统阶段	93
9.2.4 高级数据库阶段	95
9.3 数据模型	97
9.3.1 概念模型	97
9.3.2 数据模型	101
9.3.3 层次模型	103
9.3.4 网状模型	104
9.3.5 关系模型	106
9.3.6 面向对象模型	109
9.4 数据库系统的结构	111
9.4.1 数据库系统的三级模式结构	111
9.4.2 数据库系统的二级映像	113

9.4.3 数据库体系结构	114
9.5 数据库管理系统	117
9.5.1 DBMS 的工作模式	117
9.5.2 DBMS 的主要功能	118
9.5.3 DBMS 的组成	119
小结	120
习题	121
第 10 章 关系模型与关系代数	122
10.1 关系模型	122
10.2 关系代数	126
10.2.1 集合的三种基本运算——交、并、差	126
10.2.2 关系的基本运算	129
小结	133
习题	134
第 11 章 关系数据库标准语言——SQL	135
11.1 SQL 概述及特点	135
11.1.1 SQL 概述	135
11.1.2 SQL 的特点	136
11.1.3 SQL 的基本概念	137
11.2 SQL 的数据定义	138
11.2.1 数据库的定义	138
11.2.2 基本表的定义	141
11.2.3 索引的定义	147
11.3 SQL 的单表查询	149
11.3.1 SELECT 语句概述	149
11.3.2 投影运算	151
11.3.3 选择运算	153
11.3.4 排序运算	157
11.3.5 查询表	158
11.4 SQL 的连接查询	159
11.4.1 等值与非等值连接	159
11.4.2 自表连接	162
11.4.3 外连接	163
11.5 SQL 的聚合查询	166
11.5.1 聚合函数	166
11.5.2 分组聚合	167

11.6	SQL 的嵌套子查询	169
11.6.1	使用 IN 的子查询	169
11.6.2	使用比较运算符的子查询	170
11.6.3	使用存在量词 EXISTS 的子查询	172
11.7	集合运算	173
11.8	SQL 的数据操纵	174
11.8.1	插入数据	174
11.8.2	更新数据	176
11.8.3	删除数据	178
11.9	视图	179
11.9.1	创建视图	179
11.9.2	查询视图	181
11.9.3	视图更新	182
11.9.4	删除视图	183
小结		184
习题		184
第 12 章	数据库设计及优化	189
12.1	数据库设计方法	189
12.1.1	数据库和信息系统	189
12.1.2	数据库设计过程	190
12.2	需求分析	192
12.2.1	需求分析的任务	192
12.2.2	需求分析的步骤	193
12.2.3	需求分析的方法	193
12.3	概念结构设计	195
12.3.1	概念模型的基本概念	195
12.3.2	概念模型的表示方法	195
12.3.3	概念结构的特点	196
12.3.4	概念结构设计的方法	197
12.3.5	概念结构设计的步骤	198
12.4	规范化	200
12.4.1	关系模式规范化的必要性	200
12.4.2	函数依赖	201
12.4.3	范式与规范化	203
12.4.4	模式分解原则	207
12.4.5	规范化的本质分析与总结	207
12.5	逻辑结构设计	208

12.5.1	概念模型向关系模型的转换	208
12.5.2	数据模型的优化	211
12.5.3	数据库逻辑设计案例	212
12.6	数据库的物理设计	214
12.6.1	数据库物理设计的方法	214
12.6.2	确定数据库的物理结构	215
12.6.3	对物理结构进行评价	216
12.7	数据库的实施与维护	216
12.7.1	数据库的实施	216
12.7.2	数据库的维护	218
小结		219
习题		219
第 13 章	数据库安全性与完整性	221
13.1	数据库安全性	221
13.1.1	数据库安全的基本概念	221
13.1.2	用户管理	223
13.1.3	角色管理	225
13.2	数据库完整性	226
13.2.1	完整性约束的概念和类型	227
13.2.2	完整性约束的管理	228
13.3	Transact-SQL 基础	233
13.3.1	SQL 对象的命名规则和注释	233
13.3.2	数据类型	233
13.3.3	变量	237
13.3.4	函数	239
13.3.5	批处理和流程控制	242
13.4	游标	246
13.4.1	游标的使用	247
13.4.2	当前游标集的修改	250
13.5	存储过程	252
13.5.1	存储过程概述	252
13.5.2	创建和执行存储过程	252
13.5.3	修改和删除存储过程	254
13.6	触发器	255
13.6.1	触发器概述	255
13.6.2	创建触发器	256
13.6.3	删除和修改触发器	258

小结	259
习题	259
第 14 章 事务管理与恢复	260
14.1 事务	260
14.1.1 并发操作时产生的问题	260
14.1.2 事务的概念	262
14.1.3 事务的特性	263
14.2 并发控制	264
14.3 恢复与备份	266
14.3.1 数据库系统的故障	266
14.3.2 数据库备份	267
14.3.3 数据库恢复	268
小结	271
习题	271
第 15 章 数据库应用开发	272
15.1 ADO.NET 概述	272
15.2 系统分析	276
15.2.1 系统需求分析	276
15.2.2 系统用例分析	277
15.2.3 系统时序图	278
15.3 数据库分析和设计	279
15.3.1 数据库分析	279
15.3.2 数据库设计	279
15.4 数据库的连接和访问	281
15.4.1 数据库的连接	281
15.4.2 数据库的访问	282
15.5 系统界面设计及相关代码实现	284
15.5.1 酒店客房管理系统的首界面设计及其代码实现	284
15.5.2 客房信息管理界面的设计及其代码实现	286
小结	287
参考文献	288

第一部分 数据结构

第1章 绪 论

本章学习目标

- 熟练掌握数据结构中涉及的基本概念。
- 了解数据结构讨论的范畴。
- 理解算法的基本概念、描述方法以及评价标准。

本章首先介绍数据结构讨论的范畴,再介绍数据结构中涉及的基本概念和术语,最后介绍算法的概念、描述方法以及评价标准。

1.1 数据结构的概念

1.1.1 数据结构的范畴

数据结构是在整个计算机科学与技术领域上被广泛使用的术语。它用来反映一个数据的内部构成,即一个数据由哪些成分构成,以什么方式构成,呈什么结构。利用计算机进行数据处理是计算机应用的一个重要领域。在进行数据处理时,实际需要处理的数据元素一般有很多,而这些数据元素都需要存放在计算机中,因此,大量的数据元素在计算机中如何组织,以便提高数据处理的效率,并且节省计算机的存储空间,是进行数据处理的关键问题。显然,杂乱无章的数据是不便于处理的。而将大量的数据随意地存放在计算机中,实际上也是“自找苦吃”,对数据处理更是不利。

计算机已被广泛应用于数据处理。数据处理,是指对数据集中的各元素以各种方式进行运算,包括插入、删除、查找、更改等运算,也包括对数据元素进行分析。很多计算机工作者认为,程序设计的实质就是通过分析问题,确定数学模型和算法,然后再选择一个好的数据结构。即

程序 = 算法 + 数据结构

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,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。在数据处理领域中,建立数学模型有时并不十分重要,事实上,许多实际问题是无法表示成数学模型的。人们最感兴趣的是知道数据集中各数据元素之间存在什么关系,应如何组织它。例如,向量和矩阵就是数据结构,在这两种数据结构中,数据元素之间有着位置上的关系。又如,图书馆中的图书卡片目录,则是一个较为复杂的数据结构,列在各卡片上的各种书之间,可能在主题、作者等问题上相互关联,甚至一本书本身也有不同的相关成分。它们的数学模型无法用数学方程描述,而是用数据结构描述,解决此类问题的关键是设计出合适的数据结构。

下面请看三个例子。

【例 1.1】 学生入学登记表

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

表 1.1 学生入学登记表

学号	姓名	性别	出生日期	专业	贷款否	入学成绩	备注
2017001	张帆	男	1999-7-13	信息管理	F	580	
2017002	郭思达	男	1998-6-23	信息管理	T	563	
2017003	刘畅	女	1999-1-15	计算机	F	603	
2017004	李晓明	男	1998-12-15	信息管理	F	546	
2017005	白云	女	1999-5-25	计算机	F	539	

【例 1.2】 学院组织机构

学院组织机构之间的数据关系如图 1.1 所示,呈现出一种很自然的层次关系,数据与数据成一对多的关系,这就是我们所说的树状结构。

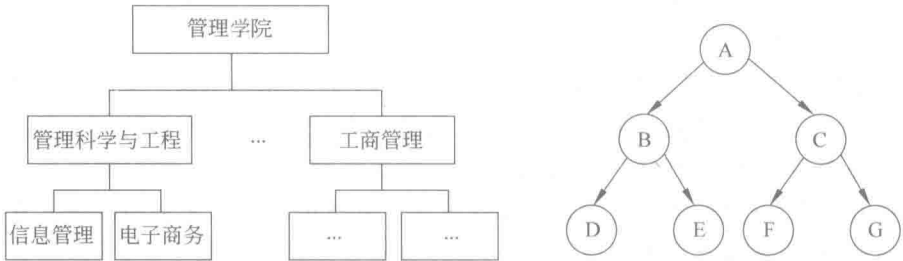


图 1.1 学院组织结构图

【例 1.3】 交通路网

从一个地方到另外一个地方可以有多条路径。交通路网中的一个结点和另外几个结点都有联系,本问题是一种典型的网状结构问题,网状结构是一种可以灵活地描述事物及其之间关系的数据模型,如图 1.2 所示。

由以上几个例子可见,描述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,概括地说: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。

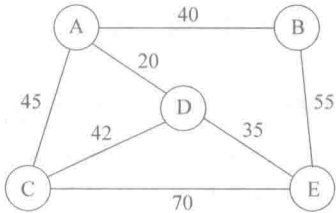


图 1.2 交通路网图

1.1.2 相关概念和术语

简单地说,数据结构是指相互有关联的数据元素的集合,是数据存在的形式。数据结

构有逻辑上的数据结构和物理上的数据结构之分。逻辑上的数据结构反映数据之间的逻辑关系,而物理上的数据结构反映数据在计算机内部的存储安排。在系统地学习数据结构知识之前,下面首先对一些基本概念和术语赋予确切的含义。

数据(Data):是对客观事物的符号表示,是指所有能被输入到计算机中,且能被计算机处理的符号的集合。在计算机科学中,数据就是计算机操作的对象总称,是计算机处理信息的某种特定符号的表示形式,如图像、声音等都可以通过编码归属于数据的范畴。

数据元素(Data Element):是数据的基本单位,是数据(集合)中的一个“个体”。数据元素具有广泛的含义。一般来说,现实世界中客观存在的一切个体都可以是数据元素。例如,表示数值的各个数 18、11、35、23、16、…可以作为数值的数据元素;表示家庭成员的各成员名父亲、儿子、女儿可以作为家庭成员的数据元素。

数据项(Data Item):是数据结构中讨论的最小单位,数据元素可以是数据项的集合。例如,描述一个运动员的信息为一个数据元素,而运动员信息中的每一项(如运动员编号、运动员姓名、俱乐部名称等)为一个数据项。

关键字(Keyword):是指能识别一个或多个数据元素的数据项。若能起到唯一识别作用,则称之为主关键字(如运动员编号),否则称之为次关键字(如运动员姓名)。

数据对象(Data Object):是具有相同特性的数据元素的集合,如整数、实数等,它是数据的一个子集。

数据结构(Data Structure):是带结构的数据元素的集合,或者说,数据结构是相互之间存在着某种逻辑关系的数据元素的集合。

一般情况下,在具有相同特征的数据元素集合中,各个数据元素之间存在有某种关系(即联系),这种关系反映了该集合中的数据元素所固有的一种结构。数据的**逻辑结构**是对数据之间关系的描述,指反映数据元素之间逻辑关系的数据结构。数据的逻辑结构有两个要素:一是数据元素的集合,通常记为 D ;二是反映了 D 中各数据元素之间的关系,通常记为 R 。即一个数据结构一般可以用二元组表示成:

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

其中, D 是数据元素的有限集; R 是 D 上关系的有限集。

数据的逻辑结构分为两大类型:线性结构与非线性结构。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,也最多有一个后继。

则称该数据结构为线性结构,又称线性表。由此可以看出,在线性结构中,各数据元素之间的前驱和后继关系是很简单的,在一个线性结构中插入或删除任何一个结点后还应是线性结构。如果一个数据结构不是线性结构,则称之为非线性结构。显然,在非线性结构中,各数据元素之间的前驱和后继关系要比线性结构复杂。

数据的逻辑结构可归结为集合、线性、树状和图状 4 类结构,如图 1.3 所示。

(1) 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(Linear Structure)中的数据元素之间存在“一对一”的关系。数组、队