



普通高等教育“十一五”国家级规划教材
教育部普通高等教育精品教材
中国大学出版社图书奖优秀教材一等奖

普通高校本科计算机专业特色教材精选·算法与程序设计

数据结构 (C++版) (第2版)

王红梅
胡 明 编著
王 涛



清华大学出版社



普通高等教育“十一五”国家级规划教材
教育部普通高等教育精品教材
中国大学出版社图书奖优秀教材一等奖

普通高校本科计算机专业特色教材精选·算法与程序设计

数据结构 (C++版) (第2版)

王红梅
胡明 编著
王涛



清华大学出版社
北京

内 容 简 介

数据结构是计算机专业教学计划中的核心课程,也是计算机及相关专业考研和水平等级考试的必考科目。要从事和计算机科学与技术相关的工作,尤其是计算机应用领域的开发和研制工作,必须具备坚实的数据结构基础。在本书第1版成功的基础上,作者进行了修订,作为第2版,本书内容更贴合《计算机学科专业硕士研究生入学考试基础综合考试大纲》,可读性和实用性更强。

本书介绍了数据结构、算法以及抽象数据类型的概念,介绍了线性表、栈、队列和串、数组、树和二叉树、图等常用数据结构,讨论了常用的查找、排序和索引技术,给出了较多的数据结构的应用实例。限于篇幅,把贯穿所有数据结构的综合案例放在了网站上,供读者下载。

本书内容丰富,层次清晰,讲解深入浅出,可作为计算机及相关专业本、专科数据结构课程的教材,也可供从事计算机软件开发和应用的工程技术人员阅读、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(C++版)/王红梅,胡明,王涛编著. —2版. —北京:清华大学出版社,2011.6
(普通高校本科计算机专业特色教材精选·算法与程序设计)
ISBN 978-7-302-24416-5

I. ①数… II. ①王… ②胡… ③王… 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

中国版本图书馆CIP数据核字(2010)第260409号

责任编辑:袁勤勇

责任校对:时翠兰

责任印制:何 芊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦A座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62795954,jsjic@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015,zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京市清华园胶印厂

装 订 者:三河市溧源装订厂

经 销:全国新华书店

开 本:185×260

印 张:19.25

字 数:438千字

版 次:2011年6月第2版

印 次:2011年6月第1次印刷

印 数:1~3000

定 价:29.00元

产品编号:038151-01

出版说明

INTRODUCTION

在我国高等教育逐步实现大众化后，越来越多的高等学校将会面向国民经济发展的第一线，为行业、企业培养各级各类高级应用型专门人才。为此，教育部已经启动了“高等学校教学质量和教学改革工程”，强调要以信息技术为手段，深化教学改革和人才培养模式改革。如何根据社会的实际需要，根据各行各业的具体人才需求，培养具有特色显著的人才，是我们共同面临的重大问题。具体地说，培养具有一定专业特色的和特定能力强的计算机专业应用型人才是计算机教育要解决的问题。

为了适应 21 世纪人才培养的需要，培养具有特色的计算机人才，急需一批适合各种人才培养特点的计算机专业教材。目前，一些高校在计算机专业教学和教材改革方面已经做了大量工作，许多教师在计算机专业教学和科研方面已经积累了许多宝贵经验。将他们的教研成果转化为教材的形式，向全国其他学校推广，对于深化我国高等学校的教学改革是一件十分有意义的事情。

清华大学出版社在经过大量调查研究的基础上，决定组织出版一套“普通高校本科计算机专业特色教材精选”。本套教材是针对当前高等教育改革的新形势，以社会对人才的需求为导向，主要以培养应用型计算机人才为目标，立足课程改革和教材创新，广泛吸纳全国各地的高等院校计算机优秀教师参与编写，从中精选出版确实反映计算机专业教学方向的特色教材，供普通高等院校计算机专业学生使用。

本套教材具有以下特点：

1. 编写目的明确

本套教材是在深入研究各地各学校办学特色的基础上，面向普通高等院校的计算机专业学生编写的。学生通过本套教材，主要学习计算机科学与技术专业的基本理论和基本知识，接受利用计算机解决实际问题的基本训练，培养研究和开发计算机系统，特别是应用系统的基本能力。

2. 理论知识与实践训练相结合

根据计算学科的三个学科形态及其关系,本套教材力求突出学科的理论与实践紧密结合的特征,结合实例讲解理论,使理论来源于实践,又进一步指导实践。学生通过实践深化对理论的理解,更重要的是使学生学会理论方法的实际运用。在编写教材时突出实用性,并做到通俗易懂,易教易学,使学生不仅知其然,知其所以然,还要会其如何然。

3. 注意培养学生的动手能力

每种教材都增加了能力训练部分的内容,学生通过学习和练习,能比较熟练地应用计算机知识解决实际问题。既注重培养学生分析问题的能力,也注重培养学生解决问题的能力,以适应新经济时代对人才的需要,满足就业要求。

4. 注重教材的立体化配套

大多数教材都将陆续配套教师用课件、习题及其解答提示,学生上机实验指导等辅助教学资源,有些教材还提供能用于网上下载的文件,以方便教学。

由于各地区各学校的培养目标、教学要求和办学特色均有所不同,所以对特色教学的理解也不尽一致,我们恳切希望大家在使用教材的过程中,及时地给我们提出批评和改进意见,以便我们做好教材的修订改版工作,使其日趋完善。

我们相信经过大家的共同努力,这套教材一定能成为特色鲜明、质量上乘的优秀教材。同时,我们也希望通过本套教材的编写出版,为“高等学校教学质量和教学改革工程”做出贡献。

清华大学出版社

前言

PREFACE

本书第1版自2005年7月出版以来,国内有近100所院校将本书作为主讲教材,在此感谢读者的认可!本书之所以在全国范围内得到广泛应用和好评,我们认为在多年讲授数据结构课程的基础上,对教材进行了以下正确设计。

1. 定位准确,合理规划教学内容。能够抓牢核心概念,提炼基础性内容,侧重工程实践,减少形式化描述,注重算法设计与程序实现。

2. 遵循认知规律,理清教学主线。能够结合教学对象分析课程特点,根据学生的认知规律,按照从已知到未知的思维进程逐步推进教学内容,知识单元的拓扑结构安排合理,主线清晰。

3. 以知识为载体,注重能力培养。能够注意引导思维,通过讲思路讲过程讲方法,展现问题的求解过程。以算法为例,按照“提出问题→分析问题→解决问题”的过程,采用“图示理解→伪代码描述算法→C++描述算法”的三级模式,培养计算思维能力。精心安排的人物小传和思想火花,激发学习兴趣,对学生的思维方式产生有益的影响。

4. 分析难点,针对处理。针对数据结构内容抽象的特点,全书共设计了250多个插图,大量的插图将抽象的内容进行了具体化处理,降低了理解问题的复杂性。数据结构的实现需要较强的C++语言的应用能力,本书的抽象数据类型用“类+结构体”的形式实现,既解释了数据结构的本质内容,又简化了程序设计。

5. 立体化教材保证教学的有效实施。本套教材包括主教材、教师用书、学习辅导和习题解析、实验指导、考研辅导、电子课件、教学网站等,是真正意义上的立体化教材。

本书第2版主要进行了如下修订:第1版的体例和主要内容保持不变;根据《计算机学科专业硕士研究生入学考试基础综合考试大纲》对内容略有增删,增加了基数排序,删去了广义表;第2章~第6章每章增加了两个应用举例;将第0章预备知识移到了附录中;教材配套的教学网站运行良好,将随书光盘和综合实例的相关内容移到了教学网站上;将模

板的数据类型由 T 改为 DataType, 增强了可读性; 逻辑序号均从 1 开始, 数组下标均从 0 开始, 所有例外均是为了简化算法且做了说明, 例如, 图的编号从 0 开始, 排序算法 0 号单元用作暂存单元, 顺序查找算法 0 号单元用作监视哨。

由于作者的知识和写作水平有限, 虽然秉着认真负责的态度并尽了最大努力, 仍难免有缺点和错误, 欢迎专家和读者批评指正。

作者的邮箱是 wanghm@mail.cut.edu.cn。

教学网站的地址是 <http://jsj.ccut.edu.cn/sjgg>。

作 者

2011 年 5 月

目 录

CONTENTS

第 1 章 绪论	1
1.1 数据结构在程序设计中的作用	1
1.2 本书讨论的主要内容	3
1.3 数据结构的基本概念	5
1.3.1 数据结构	5
1.3.2 抽象数据类型	7
1.4 算法及算法分析	9
1.4.1 算法及其描述方法	9
1.4.2 算法分析	12
思想火花——好算法是反复努力和重新修正的结果	16
习题 1	17
思考题 1	19
第 2 章 线性表	21
2.1 线性表的逻辑结构	21
2.1.1 线性表的定义	21
2.1.2 线性表的抽象数据类型定义	22
2.2 线性表的顺序存储结构及实现	24
2.2.1 线性表的顺序存储结构——顺序表	24
2.2.2 顺序表的实现	25
2.3 线性表的链接存储结构及实现	30
2.3.1 单链表	30
2.3.2 循环链表	40
2.3.3 双链表	40
2.4 顺序表和链表的比较	42
2.4.1 时间性能比较	42
2.4.2 空间性能比较	43

2.5	线性表的其他存储方法	43
2.5.1	静态链表	43
2.5.2	间接寻址	45
2.6	应用举例	45
2.6.1	顺序表的应用举例——大整数求和	45
2.6.2	单链表的应用举例——一元多项式求和	47
	思想火花——好程序要能识别和处理各种输入	50
	习题2	50
	思考题2	53
第3章	栈和队列	55
3.1	栈	55
3.1.1	栈的逻辑结构	55
3.1.2	栈的顺序存储结构及实现	57
3.1.3	栈的链接存储结构及实现	61
3.1.4	顺序栈和链栈的比较	63
3.2	队列	64
3.2.1	队列的逻辑结构	64
3.2.2	队列的顺序存储结构及实现	65
3.2.3	队列的链接存储结构及实现	68
3.2.4	循环队列和链队列的比较	71
3.3	应用举例	71
3.3.1	栈的应用举例——表达式求值	71
3.3.2	队列的应用举例——火车车厢重排	73
	思想火花——直觉可能是错误的	75
	习题3	75
	思考题3	77
第4章	字符串和多维数组	79
4.1	字符串	79
4.1.1	字符串的定义	79
4.1.2	字符串的存储结构	80
4.1.3	模式匹配	81
4.2	多维数组	85
4.2.1	数组的定义	85
4.2.2	数组的存储结构与寻址	86
4.3	矩阵的压缩存储	87
4.3.1	对称矩阵的压缩存储	87

4.3.2	三角矩阵的压缩存储	88
4.3.3	对角矩阵的压缩存储	89
4.3.4	稀疏矩阵的压缩存储	90
4.4	应用举例	92
4.4.1	字符串的应用举例——凯撒密码	92
4.4.2	数组的应用举例——幻方	93
	思想火花——用常识性的思维去思考问题	94
	习题 4	94
	思考题 4	96
第 5 章	树和二叉树	99
5.1	树的逻辑结构	99
5.1.1	树的定义和基本术语	99
5.1.2	树的抽象数据类型定义	101
5.1.3	树的遍历操作	102
5.2	树的存储结构	103
5.2.1	双亲表示法	103
5.2.2	孩子表示法	104
5.2.3	双亲孩子表示法	106
5.2.4	孩子兄弟表示法	107
5.3	二叉树的逻辑结构	107
5.3.1	二叉树的定义	108
5.3.2	二叉树的基本性质	109
5.3.3	二叉树的抽象数据类型定义	112
5.3.4	二叉树的遍历操作	113
5.4	二叉树的存储结构及实现	114
5.4.1	顺序存储结构	114
5.4.2	二叉链表	115
5.4.3	三叉链表	120
5.4.4	线索链表	120
5.5	二叉树遍历的非递归算法	125
5.5.1	前序遍历非递归算法	125
5.5.2	中序遍历非递归算法	126
5.5.3	后序遍历非递归算法	127
5.6	树、森林与二叉树的转换	128
5.7	应用举例	131
5.7.1	二叉树的应用举例——哈夫曼树及哈夫曼编码	131
5.7.2	树的应用举例——八枚硬币问题	136

思想火花——调试程序与魔术表演	137
习题 5	138
思考题 5	140
第 6 章 图	143
6.1 图的逻辑结构	143
6.1.1 图的定义和基本术语	143
6.1.2 图的抽象数据类型定义	147
6.1.3 图的遍历操作	148
6.2 图的存储结构及实现	151
6.2.1 邻接矩阵	152
6.2.2 邻接表	155
6.2.3 十字链表	159
6.2.4 邻接多重表	159
6.2.5 邻接矩阵和邻接表的比较	160
6.3 最小生成树	161
6.3.1 MST 性质	161
6.3.2 Prim 算法	162
6.3.3 Kruskal 算法	165
6.4 最短路径	168
6.4.1 Dijkstra 算法	169
6.4.2 Floyd 算法	171
6.5 有向无环图及其应用	173
6.5.1 AOV 网与拓扑排序	173
6.5.2 AOE 网与关键路径	176
6.6 应用举例	179
6.6.1 图的应用举例 1——七桥问题	179
6.6.2 图的应用举例 2——七巧板涂色	180
思想火花——数据模型在问题求解中的作用	181
习题 6	181
思考题 6	185
第 7 章 查找技术	187
7.1 概述	187
7.1.1 查找的基本概念	187
7.1.2 查找算法的性能	188
7.2 线性表的查找技术	189
7.2.1 顺序查找	189

7.2.2	折半查找	191
7.3	树表的查找技术	195
7.3.1	二叉排序树	195
7.3.2	平衡二叉树	201
7.4	散列表的查找技术	206
7.4.1	概述	206
7.4.2	散列函数的设计	207
7.4.3	处理冲突的方法	210
7.4.4	散列查找的性能分析	213
7.4.5	开散列表与闭散列表的比较	214
	思想火花——把注意力集中于主要因素,不要纠缠于噪声	214
	习题 7	215
	思考题 7	218
第 8 章	排序技术	219
8.1	概述	219
8.1.1	排序的基本概念	219
8.1.2	排序算法的性能	220
8.2	插入排序	221
8.2.1	直接插入排序	221
8.2.2	希尔排序	223
8.3	交换排序	225
8.3.1	起泡排序	225
8.3.2	快速排序	228
8.4	选择排序	231
8.4.1	简单选择排序	231
8.4.2	堆排序	233
8.5	归并排序	238
8.5.1	二路归并排序的非递归实现	238
8.5.2	二路归并排序的递归实现	241
8.6	分配排序	242
8.6.1	桶式排序	242
8.6.2	基数排序	244
8.7	各种排序方法的比较	247
	思想火花——学会“盒子以外的思考”	249
	习题 8	250
	思考题 8	253

第9章 索引技术	255
9.1 索引的基本概念	255
9.2 线性索引技术	256
9.2.1 稠密索引.....	256
9.2.2 分块索引.....	257
9.2.3 多重表.....	258
9.2.4 倒排表.....	258
9.3 树形索引	259
9.3.1 2-3 树	259
9.3.2 B_树	262
9.3.3 B ⁺ 树	266
思想火花——随处可见的索引.....	267
习题9	268
附录A 预备知识	271
A.1 数学术语	271
A.2 级数求和	271
A.3 集合	272
A.4 关系	273
附录B C++ 语言基本语法	275
B.1 程序结构	275
B.2 数据类型	276
B.3 控制语句	278
B.4 输入与输出	279
B.5 动态存储分配	280
B.6 函数	281
B.7 类与对象	284
B.8 模板	287
B.9 异常处理	288
附录C 词汇索引	289
参考文献	294

第 1 章

CHAPTER

绪 论

用计算机求解任何问题都离不开程序设计,程序设计的实质是数据表示和数据处理。数据要被计算机处理,首先必须能够存储在计算机的内存中,这项任务称为**数据表示**,其核心是数据结构;一个实际问题的求解必须满足各项处理要求,这项任务称为**数据处理**,其核心是算法。数据结构课程主要讨论数据表示和数据处理的基本方法。本章将概括地介绍数据结构和算法的基本概念。

【学习重点】

- ① 数据结构及相关概念;
- ② 数据的逻辑结构和存储结构,二者之间的关系;
- ③ 算法及特性;
- ④ 大O记号。

【学习难点】

- ① 抽象数据类型的理解和使用;
- ② 伪代码的理解和使用;
- ③ 算法的时间复杂度分析。

1.1 数据结构在程序设计中的作用

用计算机求解任何问题都离不开程序设计,但是计算机不能分析问题并产生问题的解决方案,必须由人(即程序设计者)分析问题,确定问题的解决方案,采用计算机能够理解的指令描述这个问题的求解步骤(即编写程序),然后让计算机执行程序最终获得问题的解。程序设计的一般过程^①如图 1-1 所示。

^① “数据结构”课程是计算机软件基础课程群的核心课程,在作者编著的计算机软件基础课程群系列教材《数据结构(C++版)》(王红梅,胡明,王涛编著,清华大学出版社)、《程序设计基础——从问题到程序》(胡明,王红梅编著,清华大学出版社)、《算法设计与分析》(王红梅编著,清华大学出版社)中,都是通过这个程序设计过程培养学生的计算思维能力和程序设计能力。

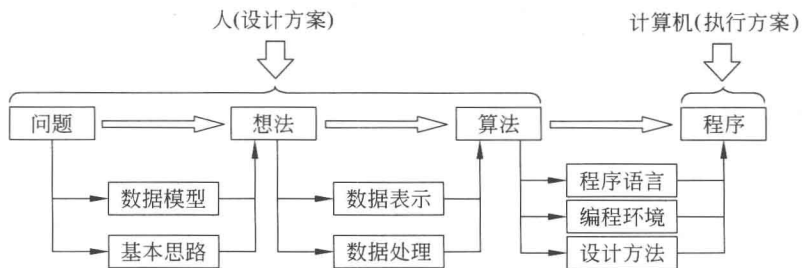


图 1-1 程序设计的一般过程

由问题到想法需要分析问题,抽象出具体的数据模型(待处理的数据以及数据之间的关系,即数据结构),形成问题求解的基本思路;由想法到算法需要完成数据表示(将数据以及数据之间的关系存储到计算机的内存中)和数据处理(具体的操作步骤,将问题求解的基本思路形成算法);由算法到程序需要将算法的操作步骤转换为某种程序设计语言对应的语句,换言之,就是用某种程序设计语言描述数据处理的过程。

图灵^①奖获得者沃思^②给出了一个著名的公式:数据结构+算法=程序。从这个公式可以看到,数据结构和算法是构成程序的两个重要的组成部分,一个“好”程序首先是将问题抽象出一个适当的数据结构,然后基于该数据结构设计一个“好”算法。对于许多实际的问题,写出一个可以正确运行的程序还不够,如果这个程序在规模较大的数据集上运行,那么运行效率就成为一个重要的问题。下面以手机电话号码查询问题为例,说明数据结构在程序设计中的作用。

例 1-1 手机电话号码查询问题

假设某手机中存储了如表 1-1 所示的若干电话号码,如果将这个电话号码集合组织为线性结构(即线性排列),则查找某人的电话号码只能进行顺序查找。

表 1-1 某手机中的电话号码

姓名	王靓靓	赵刚	韩春颖	李琦勇	...	张强
电话	13833278900	13944178123	15594434552	13833212779	...	13331688900

如果将这个电话号码集合进行分组,即组织为树结构,如图 1-2 所示,则查找某人的电话号码可以只在某个分组中进行。显然,后者的查找效率更高,当数据量较大时差别就更大。

^① 图灵(Alan Mathison Turing 1912—1954)出生于伦敦,24岁提出图灵机理论(这一理论奠定了整个现代计算机的理论基础),31岁参与 COLOSSUS 的研制,33岁设想仿真系统,35岁提出自动程序设计概念,38岁发表论文《计算机能思考吗?》,论证了人工智能的可能性,被人们推崇为人工智能之父。美国计算机协会(ACM)为了纪念图灵对计算机科学的贡献,从1966年起设立图灵奖。图灵奖被誉为计算机界的诺贝尔奖。

^② 沃思(Niklaus Wirth)1934年生于瑞士。1968年设计并实现了 PASCAL 语言(被誉为 PASCAL 之父),1971年提出了结构化程序设计,1976年设计并实现了 Modula 2 语言。除了程序设计语言之外,沃思在其他方面也有许多创造,如扩充了著名的巴科斯范式,发明了语法图等。1984年获图灵奖。

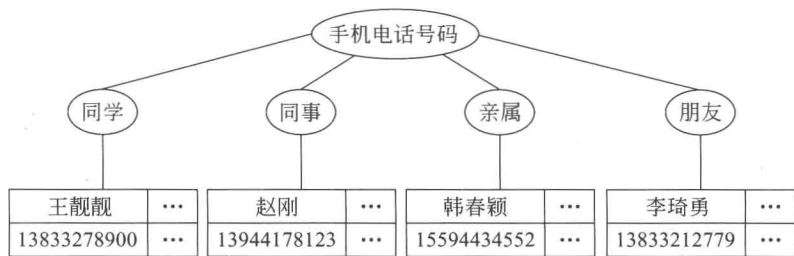


图 1-2 分组——将电话号码集合组织为树结构

1.2 本书讨论的主要内容

计算机能够求解的问题一般可以分为数值问题和非数值问题。数值问题抽象出的数据模型通常是数学方程,例如求解梁架结构中应力的模型是线性方程组;预报人口增长情况的模型是微分方程;非数值问题抽象出的数据模型通常是线性表、树、图等数据结构。下面请看三个例子。

例 1-2 学籍管理问题

用计算机来完成学籍管理,就是由计算机程序处理学生学籍登记表,实现增、删、改、查、统计等功能。表 1-2 所示是一张简单的学生学籍登记表。在学籍管理问题中,计算机的操作对象是每个学生的学籍信息——称为档案表项,各档案表项之间的关系可以使用称为线性表的数据结构来描述。

表 1-2 学生学籍登记表

学号	姓名	性别	出生日期	政治面貌
0001	陆宇	男	1986/09/02	团员
0002	李明	男	1985/12/25	党员
0003	汤晓影	女	1986/03/26	团员
...

例 1-3 人机对弈问题

计算机之所以能和对弈,是因为对弈的策略实现已存入计算机。在对弈问题中,计算机的操作对象是对弈过程中可能出现的棋盘状态——称为格局,而格局之间的关系是由对弈规则决定的。因为从一个格局可以派生出多个格局,所以,这种关系通常不是线性的。如图 1-3(a)所示为井字棋^①的一个格局,从该格局出发可以派生出 5 个新的格局,从新的格局出发,还可以再派生出新的格局,如图 1-3(b)所示。格局之间的关系可以用称

^① 井字棋又称三子连珠,由两个人对弈,棋盘为 3×3 的方格,当一方的三个棋子占同一行、同一列或同一对角线时便为胜方。

为树的数据结构来描述。

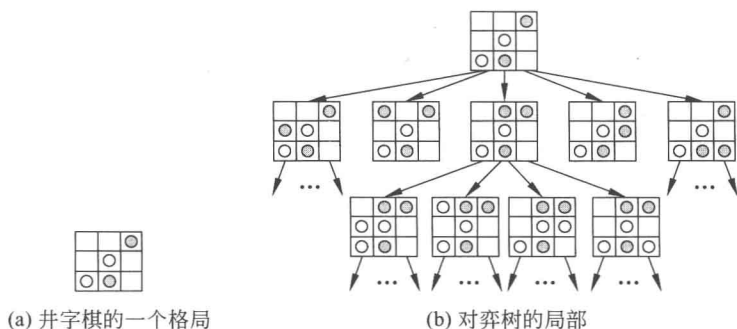


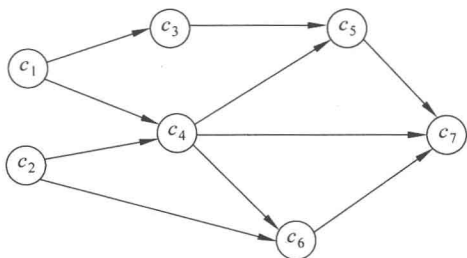
图 1-3 对弈问题中格局之间的关系

例 1-4 教学计划编排问题

一个教学计划包含许多课程,有些课程之间必须按规定的先后次序进行,图 1-4(a)列出了计算机软件方向的一些课程以及课程之间的次序关系。在教学计划编排问题中,计算机的操作对象是课程,课程之间的次序关系可以用称为图的数据结构来描述。如图 1-4(b)所示,其中每个顶点表示一门课程,如果从顶点 c_i 到 c_j 之间存在边 $\langle c_i, c_j \rangle$,则表示课程 c_i 是课程 c_j 的先修课。

课程编号	课程名称	先修课程
c_1	高等数学	无
c_2	计算机导论	无
c_3	离散数学	c_1
c_4	程序设计语言	c_1 、 c_2
c_5	数据结构	c_3 、 c_4
c_6	计算机原理	c_2 、 c_4
c_7	数据库原理	c_4 、 c_5 、 c_6

(a) 计算机软件方向的一些课程



(b) 课程之间的次序关系

图 1-4 课程以及课程之间的次序关系

本书讨论非数值问题的数据组织和处理,主要内容有如下 4 点。

(1) 数据的逻辑结构: 线性表、树、图等数据结构,其核心是如何组织待处理的数据以及数据之间的关系。

(2) 数据的存储结构: 如何将线性表、树、图等数据结构存储到计算机的存储器中,其核心是如何有效地存储数据以及数据之间的逻辑关系。

(3) 算法: 如何基于数据的某种存储结构实现插入、删除、查找等基本操作,其核心是如何有效地处理数据。

(4) 常用数据处理技术: 包括查找技术、排序技术、索引技术等。