



21世纪全国高等院校计算机教育“十一五”规划教材

数据结构 (C++版)

主 编	陈天凯	田俊华
副主编	杜兴勇	徐云娟
	李保红	辛建军



中国计划出版社

21世纪全国高等院校计算机教育“十一五”规划教材

数据结构（C++版）

本书编委会 编著



中国计划出版社

图书在版编目 (C I P) 数据

数据结构: C++版 / 《数据结构 (C++版)》编委会编著. —北京: 中国计划出版社, 2008. 7

21世纪全国高等院校计算机教育“十一五”规划教材
ISBN 978-7-80242-128-8

I. 数… II. 数… 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. TP311.12 TP312

中国版本图书馆CIP数据核字 (2008) 第087963号

内 容 简 介

本书是21世纪全国高等院校计算机教育“十一五”规划教材之一。本书采用面向对象的观点讨论数据结构技术, 从数据结构的基础知识讲起, 循序渐进地讲解, 且设置了大量的典型应用实例与算法设计。本书主要内容包括: 数据结构基础知识, 线性表, 栈和队列, 串, 数组和广义表, 树, 图, 查找, 排序, 递归, 文件, 以及实习题与实习指导。

本书注重内容的实用性, 既可作为工程型、应用型高等院校计算机专业数据结构课程的教材, 也可作为自学数据结构的教材和参考书。

21世纪全国高等院校计算机教育“十一五”规划教材 数据结构 (C++版)

本书编委会 编著

☆

中国计划出版社出版

(地址: 北京市西城区木樨地北里甲11号国宏大厦C座4层)

(邮政编码: 100038 电话: 63906433 63906381)

新华书店北京发行所发行

河北省高碑店市鑫宏源印刷包装有限责任公司印刷

787×1092 毫米 1/16 22.5 印张 547 千字

2008 年 7 月第 1 版 2008 年 7 月第 1 次印刷

印数 1—4000 册

☆

ISBN 978-7-80242-128-8

定价: 35.00元

本书编委会

主 编：陈天凯 田俊华

副主编：杜兴勇 徐云娟 李保红 辛建军

参 编：耿 彧 常国锋 郭遂峰 王 茹 李 娜 田 雪

前 言

数据结构课程是我国计算机教学中较早形成和完善的一门专业基础课程，也是计算机课程体系中的核心课程之一。在该课程中所介绍各类数据的逻辑结构、存储方式及相关的算法既是程序设计，特别是非数值性程序设计的基础，又是设计和实现系统软件及大型应用软件的重要基础。

本教材主要针对工程型、应用型院校的计算机类专业学生，以培养与提高学生的基本专业素质及综合应用能力为目标，注重体现教材的以下特色：

1. 先进性。采用面向对象的观点讨论数据结构技术，对每一种抽象的数据类型先定义成抽象类，然后再结合各种具体的存储结构加以实现，并以各实现类为线索对抽象类中各种操作的实现方法加以说明。

2. 实用性。在本教材中设置了大量的应用实例与算法设计的内容，尽量包括各章所介绍的类的功能演示及典型的应用问题。同时，本书中所有用 C++ 语言编写的代码均可在 VC++ 6.0 环境下调试通过。

3. 适应性。在本教材中对每一种数据类型都有相对比较规范的表述过程，对每一种算法都有相对比较规范统一的说明步骤，并通过图示、文字注释、实例的执行过程等多种方式来帮助学生理解算法，提高学生的计算机思维能力。

本书由陈天凯、田俊华主编，杜兴勇、徐云娟、李保红、辛建军担任副主编，耿彧、常国锋、郭遂峰、王茹、李娜、田雪参与编写。

由于作者水平有限，书中难免存在缺点与疏漏，敬请读者及同行们予以批评指正。

编者

2008年5月

目 录

第1章 绪论	1
1.1 数据结构概述	1
1.1.1 数据结构的重要性	1
1.1.2 有关概念和术语	4
1.1.3 数据结构课程的内容	5
1.2 面向对象技术	6
1.2.1 面向对象技术概述	6
1.2.2 C++基本概念	7
1.2.3 C++中的类和对象	10
1.2.4 C++模板与设计模式	17
1.3 数据和数据类型	23
1.3.1 抽象数据类型	24
1.3.2 数据结构的概念	24
1.3.3 数据的逻辑结构	25
1.3.4 数据的存储结构	25
1.3.5 数据的运算	26
1.3.6 数据的逻辑结构、存储结构及数据的运算的关系	27
1.4 算法和算法分析	28
1.4.1 算法	28
1.4.2 算法描述	28
1.4.3 算法性能分析与度量	29
1.5 小结与提高	30
1.6 思考与练习	30
第2章 线性表	31
2.1 线性表及其抽象数据类型说明	31
2.1.1 线性表及其逻辑结构	31
2.1.2 线性表的基本操作	32
2.1.3 线性表的抽象数据类型描述	33
2.2 线性表的顺序存储及运算实现	33
2.2.1 顺序表及顺序存储	33
2.2.2 顺序表类的定义	35
2.2.3 顺序表上基本运算的实现	36

2.2.4	顺序表应用举例	41
2.3	线性表的链式存储和运算实现	44
2.3.1	线性链表的存储结构	44
2.3.2	线性链表类的定义	46
2.3.3	线性链表的实现	47
2.4	循环链表	55
2.5	双向链表	61
2.6	可利用空间表	67
2.7	线性表的顺序存储和链式存储的比较	68
2.8	链式存储结构的应用	69
2.8.1	约瑟夫问题	69
2.8.2	一元多项式求和	73
2.9	小结与提高	77
2.10	思考与练习	77
第3章	栈和队列	79
3.1	栈	79
3.1.1	栈的定义	79
3.1.2	栈的顺序存储结构	81
3.1.3	栈的链式存储结构	85
3.1.4	顺序栈和链栈的比较	87
3.2	栈的应用	87
3.2.1	表达式求值	87
3.2.2	进制转换	93
3.2.3	火车调度	94
3.3	队列	95
3.3.1	队列的定义	95
3.3.2	队列的顺序存储结构	97
3.3.3	队列的链式存储结构	99
3.3.4	顺序队列与链式队列的比较	101
3.3.5	优先队列	103
3.4	队列的应用	104
3.4.1	基数排序法	104
3.4.2	火车车厢重排	106
3.5	双端队列	108
3.5.1	双端队列的定义	108
3.5.2	双端队列的操作	109
3.5.3	双端队列的应用	109
3.6	小结与提高	110
3.7	思考与练习	110

第4章 串	112
4.1 串类型的定义	112
4.2 串的存储结构	114
4.2.1 串的顺序存储结构	114
4.2.2 串的链式存储结构	115
4.3 串的操作	115
4.3.1 常用的C++字符串函数	115
4.3.2 串的抽象数据类型的描述	116
4.4 模式匹配	118
4.4.1 模式匹配的BF算法	118
4.4.2 模式匹配的KMP算法	119
4.5 串应用	123
4.5.1 在文本编辑中的应用	123
4.5.2 建立关键字索引	127
4.6 小结与提高	129
4.7 思考与练习	129
第5章 数组和广义表	130
5.1 数组的定义及抽象数据类型表示	130
5.1.1 C++中数组的定义	130
5.1.2 数组的抽象数据类型表示	131
5.1.3 数组的顺序存储结构	131
5.2 矩阵的压缩存储	133
5.2.1 特殊矩阵的压缩存储	133
5.2.2 稀疏矩阵的压缩存储	136
5.3 广义表	141
5.3.1 广义表的概念	141
5.3.2 广义表的存储结构	142
5.3.3 广义表的操作	144
5.4 小结与提高	147
5.5 思考与练习	147
第6章 树与二叉树	149
6.1 树的基本概念	149
6.1.1 树的定义和常用术语	149
6.1.2 树的抽象数据类型	152
6.1.3 树的遍历操作	153
6.2 二叉树	154
6.2.1 二叉树的定义	154
6.2.2 二叉树的性质	156

6.2.3	二叉树的存储结构	158
6.2.4	二叉树的遍历	160
6.2.5	二叉树遍历的应用	165
6.3	线索二叉树	166
6.3.1	线索二叉树的类定义	166
6.3.2	中序线索二叉树	166
6.4	树、森林和二叉树的关系	169
6.4.1	树的存储结构	169
6.4.2	树、森林与二叉树的转换	174
6.4.3	森林的遍历	177
6.5	哈夫曼树及其应用	177
6.5.1	哈夫曼树的定义	177
6.5.2	哈夫曼树的构造	179
6.5.3	哈夫曼树在编码问题中的应用	181
6.6	小结与提高	182
6.7	思考与练习	182
第7章	图	184
7.1	图的基本概念	184
7.1.1	图的定义及基本概念	184
7.1.2	图的抽象数据类型	188
7.2	图的存储结构	189
7.2.1	邻接矩阵表示法	189
7.2.2	邻接表	190
7.3	图的遍历	193
7.3.1	深度优先搜索	193
7.3.2	广度优先搜索	195
7.4	图的连通性	196
7.4.1	无向图的连通分量	196
7.4.2	有向图的强连通分量	196
7.5	生成树	197
7.5.1	构造生成树	197
7.5.2	最小生成树	198
7.5.3	普里姆算法	198
7.5.4	克鲁斯卡尔算法	200
7.6	最短路径	202
7.6.1	单源最短路径	203
7.6.2	每一对顶点之间的最短路径	204
7.7	拓扑排序	205

7.8	关键路径	207
7.9	小结与提高	210
7.10	思考与练习	210
第8章	查找	213
8.1	基本概念	213
8.2	线性表的查找	214
8.2.1	顺序查找	214
8.2.2	折半查找	216
8.2.3	索引查找	219
8.2.4	分块查找	220
8.3	树表的查找	221
8.3.1	二叉查找树	222
8.3.2	平衡二叉树	228
8.3.3	B-树	236
8.4	哈希表的查找	242
8.4.1	哈希表	242
8.4.2	构造哈希函数的基本方法	243
8.4.3	解决冲突的方法	245
8.4.4	哈希表的查找方法	249
8.5	各种查找方法的比较	250
8.6	小结与提高	252
8.7	思考与练习	252
第9章	排序	254
9.1	基本概念	254
9.2	内部排序	256
9.2.1	插入排序	256
9.2.2	交换排序	259
9.2.3	选择排序	263
9.2.4	归并排序	264
9.2.5	基数排序	266
9.3	内部排序方法比较	268
9.4	外部排序	269
9.4.1	外部排序的简单方法	269
9.4.2	两路归并	270
9.4.3	多路归并	271
9.5	小结与提高	272
9.6	思考与练习	272

第10章 递归	273
10.1 递归的定义	273
10.2 常见递归问题	274
10.2.1 汉诺塔问题	274
10.2.2 八皇后问题	276
10.2.3 表达式树	278
10.3 递归的实现	282
10.4 消除递归	284
10.4.1 尾递归和单向递归的消除	284
10.4.2 用栈模拟系统运行时的栈	286
10.5 递归的评估	288
10.6 小结与提高	289
10.7 思考与练习	290
第11章 文件	291
11.1 外存储器的介绍	291
11.2 磁盘	292
11.3 有关文件的概念	293
11.3.1 文件及其类别	293
11.3.2 文件的操作	294
11.4 文件的组织	294
11.4.1 顺序文件	294
11.4.2 索引文件	297
11.4.3 散列文件	299
11.4.4 多关键字文件	300
11.5 文件的索引结构	303
11.6 小结与提高	305
11.7 思考与练习	306
第12章 实习题与实习指导	307
12.1 实习的目的和要求	307
12.2 实习步骤	308
12.3 实习报告和样例	309
12.3.1 实习报告	309
12.3.2 实习样例一	310
12.3.3 实习样例二	320
12.4 实习题	328
12.4.1 实习题一 C++类及模板的使用	328
12.4.2 实习题二 含并、交和差运算的集合类型	329
12.4.3 实习题三 用回溯法求解“稳定婚配”问题	330

12.4.4	实习题四	最佳任务分配方案求解.....	331
12.4.5	实习题五	以队列实现的仿真技术预测理发馆的经营状况.....	332
12.4.6	实习题六	利用树形结构的搜索算法模拟因特网域名的查询.....	333
12.4.7	实习题七	汽车牌照的快速查找.....	334
12.4.8	实习题八	旅游最佳路线的选择.....	335
12.4.9	实习题九	教务课程计划的辅助制定.....	336
附录A ASCII字符集			338
附录B 数制系统.....			341
主要参考文献.....			343

第 1 章

绪论

计算机科学是一门研究数据表示和数据处理的科学。数据是计算机化的信息，它是计算机可以直接处理的最基本和最重要的对象。无论是进行科学计算、数据处理、过程控制以及文件的存储和检索，还是在数据库技术等计算机应用领域中，都需要对数据进行加工处理。因此，要设计出一个结构好、效率高的程序，必须研究数据的特性、数据间的相互关系及其对应的存储表示，并利用这些特性和关系设计出相应的算法和程序。



本章主要内容

- ❏ 数据结构概述
- ❏ 面向对象技术
- ❏ 数据和数据类型
- ❏ 算法和算法分析

1.1 数据结构概述

1.1.1 数据结构的重要性

数据结构是计算机科学与技术专业的专业基础课，是十分重要的核心课程。所有的计算机系统软件和应用软件都要用到各种类型的数据结构。因此，要想更好地运用计算机来解决实际问题，仅掌握几种计算机程序设计语言是难以应付众多复杂的课题的。要想有效地使用计算机、充分发挥计算机的性能，还必须学习和掌握好数据结构的有关知识。打下“数据结构”这门课程的扎实基础，对于学习计算机专业的其他课程，如操作系统、编译原理、数据库管理系统、软件工程、人工智能等都是十分有益的。

在计算机发展的初期，人们使用计算机的目的主要是处理数值计算问题。当使用计算机解决一个具体问题时，一般需要经过下列几个步骤：首先，要从该问题中抽象出一个适当的数学模型；然后，设计或选择一个解此数学模型的算法；最后，编写出程序进行调试、测试，直至得到最终的解答。例如，求解梁架结构中应力的数学模型可以是线性方程组，该方程组可以使用迭代算法来求解。

由于当时所涉及的运算对象是简单的整型、实型或布尔类型数据,所以程序设计者的主要精力是集中于程序设计的技巧上,而无须重视数据结构。随着计算机应用领域的扩大和软、硬件的发展,非数值计算问题显得越来越重要。据统计,当今处理非数值计算性问题占用了 90%以上的机器时间。这类问题涉及到的数据结构更为复杂,数据元素之间的相互关系一般无法用数学方程式加以描述。因此,解决这类问题的关键不再是数学分析和计算方法,而是要设计出合适的数据结构。下面所列举的就是属于这一类的具体问题。

【例 1-1】 学生信息检索系统。当需要查找某个学生的有关情况的时候,或者想查询某个专业或年级的学生的有关情况的时候,只要建立了相关的数据结构,按照某种算法编写了相关程序,就可以实现计算机自动检索。由此,可以在学生信息检索系统中建立一张按学号顺序排列的学生信息表(表 1-1)和分别按姓名、专业、年级顺序排列的索引表(表 1-2、表 1-3、表 1-4)。由这 4 张表构成的文件便是学生信息检索的数学模型,计算机的主要操作是按照某个特定要求(如给定姓名)对学生信息文件进行查询。

诸如此类的还有电话自动查号系统、考试查分系统、仓库库存管理系统等。在这类文档管理的数学模型中,计算机处理的对象之间通常存在着一种简单的线性关系,这类数学模型可称为线性的数据结构。

表 1-1 学生信息表

学 号	姓 名	性 别	专 业	年 级
980001	吴承志	男	计算机科学与技术	1998 级
980002	李淑芳	女	信息与计算科学	1998 级
990301	刘 丽	女	数学与应用数学	1999 级
990302	张会友	男	信息与计算科学	1999 级
990303	石宝国	男	计算机科学与技术	1999 级
000801	何文颖	女	计算机科学与技术	2000 级
000802	赵胜利	男	数学与应用数学	2000 级
000803	崔文靖	男	信息与计算科学	2000 级
010601	刘 丽	女	计算机科学与技术	2001 级
010602	魏永鸣	男	数学与应用数学	2001 级

表 1-2 姓名索引表

姓 名	学 生 编 号	姓 名	学 生 编 号
崔文靖	8	魏永鸣	10
何文颖	6	吴承志	1
李淑芳	2	赵胜利	7
刘 丽	3, 9	张会有	4
石宝国	5		

表 1-3 专业索引表

专 业	学 生 编 号	专 业	学 生 编 号
计算机科学与技术	1, 5, 6, 9	数学与应用数学	3, 7, 10
信息与计算科学	2, 4, 8		

表 1-4 年级索引表

年 级	学 生 编 号	年 级	学 生 编 号
2000 级	6, 7, 8	1998 级	1, 2, 3
2001 级	9, 10	1999 级	4, 5

【例 1-2】 八皇后问题。在八皇后问题中，处理过程不是根据某种确定的计算法则，而是利用试探和回溯的探索技术求解。为了求得合理布局，在计算机中要存储布局的当前状态。从最初的布局状态开始，一步步地进行试探，每试探一步形成一个新的状态，整个试探过程形成了一棵隐含的状态树。为了描述方便，将八皇后问题简化为四皇后问题，如图 1-1 所示。回溯法求解过程实质上就是一个遍历状态树的过程。在这个问题中所出现的树也是一种数据结构，它可以应用在许多非数值计算的问题中。

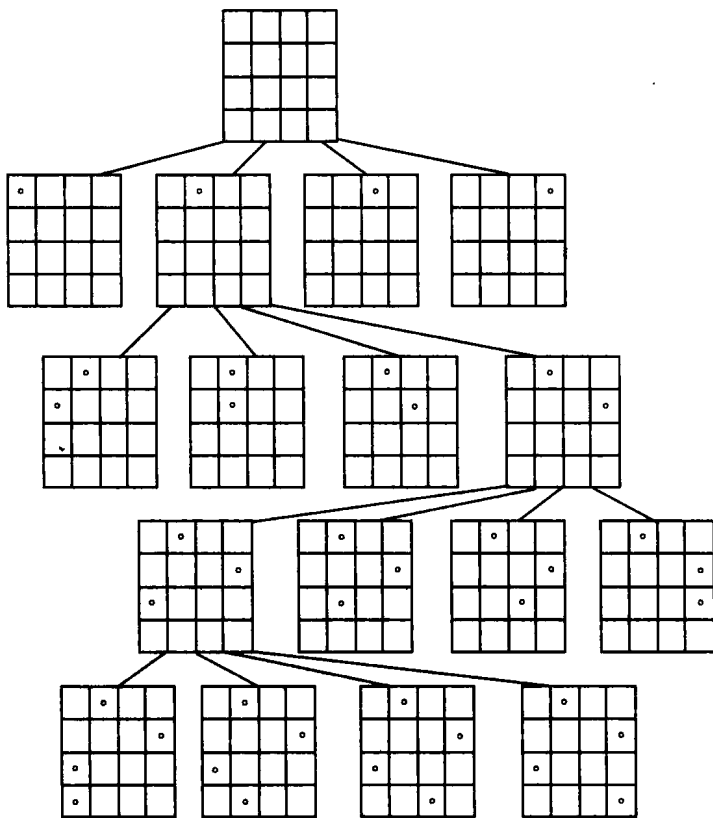


图 1-1 四皇后问题中隐含的状态树

【例 1-3】 教学计划编排问题。一个教学计划包含许多课程，在教学计划包含的许多课

程之间,有些必须按规定的先后次序进行,有些则没有次序要求。即有些课程之间有先修和后续的关系,有些课程可以任意安排次序。计算机专业的课程设置如表 1-5 所示,各个课程之间的次序关系可用一个称作“图”的数据结构来表示,如图 1-2 所示。有向图中的每个顶点表示一门课程,如果从顶点 v_i 到 v_j 之间存在有向边 $\langle v_i, v_j \rangle$,则表示课程 i 必须先于课程 j 进行。

表 1-5 计算机专业的课程设置

课 程 编 号	课 程 名 称	先 修 课 程
C_1	计算机导论	无
C_2	数据结构	C_1, C_4
C_3	汇编语言	C_1
C_4	C 程序设计语言	C_1
C_5	计算机图形学	C_2, C_3, C_4
C_6	接口技术	C_3
C_7	数据库原理	C_2, C_9
C_8	编译原理	C_4
C_9	操作系统	C_2

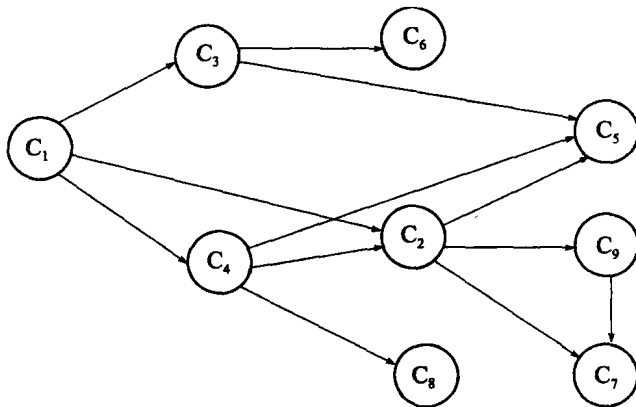


图 1-2 表示课程之间优先关系的有向图

由以上 3 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,可以说数据结构课程主要研究的是,非数值计算的程序设计中所出现的计算机操作对象、对象之间的关系,以及具体的操作。

学习数据结构的目的是了解计算机处理对象的特性,将实际问题中所涉及的处理对象在计算机中表示出来,并对它们进行处理。与此同时,通过算法训练来提高学生的思维能力,通过程序设计的技能训练来促进学生的综合应用能力和专业素质的提高。

1.1.2 有关概念和术语

在系统地学习数据结构知识之前,先对一些基本概念和术语赋予确切的含义。

- **数据 (Data)**：信息的载体。它能够被计算机识别、存储和加工处理，是计算机程序加工的原料，应用程序处理各种各样的数据。计算机科学中，所谓数据就是计算机加工处理的对象，它可以是数值数据，也可以是非数值数据。数值数据是一些整数、实数或复数，主要用于工程计算、科学计算和商务处理等；非数值数据包括字符、文字、图形、图像、语音等。
- **数据元素 (Data Element)**：数据的基本单位。在不同的条件下，数据元素又可称为元素、结点、顶点、记录等。例如，学生信息检索系统中学生信息表中的一个记录、八皇后问题中状态树的一个状态、教学计划编排问题中的一个顶点等，都被称为一个数据元素。

有时，一个数据元素可由若干个数据项 (Data Item) 组成。例如，学籍管理系统中学生信息表的每一个数据元素就是一个学生记录，它包括学生的学号、姓名、性别、籍贯、出生年月、成绩等数据项。这些数据项可以分为两种：一种叫做初等项，如学生的性别、籍贯等，这些数据项是在数据处理时不能再分割的最小单位；另一种叫做组合项，如学生的成绩，它可以再划分为数学、物理、化学等更小的项。通常，在解决实际问题时，把每个学生记录当作一个基本单位进行访问和处理。

- **数据对象 (Data Object) 或数据元素类 (Data Element Class)**：具有相同性质的数据元素的集合。在某个具体问题中，数据元素都具有相同的性质（元素值不一定相等），属于同一数据对象（数据元素类），数据元素是数据元素类的一个实例。例如，在交通咨询系统的交通网中，所有的顶点是一个数据元素类，顶点 A 和顶点 B 各自代表一个城市，是该数据元素类中的两个实例，其数据元素的值分别为 A 和 B。
- **数据结构 (Data Structure)**：互相之间存在着一种或多种关系的数据元素的集合。在任何问题中，数据元素都不会是孤立的，在它们之间都存在着这样或那样的关系，这种数据元素之间的关系称为结构。

从上面所介绍的数据结构的概念中可以知道，一个数据结构有两个要素：一个是数据元素的集合；另一个是关系的集合。在形式上，数据结构通常可以采用一个二元组来表示：

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

其中， D 是数据元素的有限集， R 是 D 上关系的有限集。

数据结构包括数据的逻辑结构和数据的物理结构。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，它与数据的存储无关。研究数据结构的目的是在计算机中实现对其的操作，为此还需要研究如何在计算机中表示一个数据结构。数据结构在计算机中的标识（又称映像）称为数据的物理结构或存储结构。它所研究的是数据结构在计算机中的实现方法，包括数据结构中元素的表示及元素间关系的表示。

1.1.3 数据结构课程的内容

数据结构与数学、计算机硬件和软件有十分密切的关系。数据结构是介于数学、计算机硬件和计算机软件之间的一门计算机科学与技术专业的核心课程，是高级程序设计语言、