

普通高校本科计算机专业

特色

教材精选

算法设计与分析

王红梅 编著

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

清华大学出版社



普通高校本科计算机专业**特色**教材精选

算法设计与分析

王红梅 编著

清华大学出版社
北京

内 容 简 介

算法设计与分析是计算机科学技术中处于核心地位的一门专业基础课,越来越受到重视。本书将计算机经典问题和算法设计技术很好地结合起来,系统地介绍了算法设计技术及其在经典问题中的应用。全书共12章,第1章介绍了算法的基本概念和算法分析方法,第2章从算法的观点介绍了NP完全理论,第3章~第11章分别介绍了蛮力法、分治法、减治法、动态规划法、贪心法、回溯法、分支限界法、概率算法和近似算法等算法设计技术,第12章基于图灵机计算模型介绍了计算复杂性理论。每章均附有一篇阅读材料,以通俗易懂的笔触介绍了算法领域的一些最新研究成果。书中所有算法均给出了伪代码,大部分算法还给出了C++描述,书中所有问题均给出了若干应用实例。

本书内容丰富,深入浅出,结合应用,图例丰富,可作为高等院校计算机专业本科和研究生学习算法设计与分析的教材,也可供工程技术人员和自学读者学习参考。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

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。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

算法设计与分析 / 王红梅编著. —北京:清华大学出版社,2006.7

(普通高校本科计算机专业特色教材精选)

ISBN 7-302-12942-8

I. 算… II. 王… III. ①电子计算机—算法设计—高等学校—教材 ②电子计算机—算法分析—高等学校—教材 IV. TP301.6

中国版本图书馆CIP数据核字(2006)第041915号

出版者:清华大学出版社

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

社总机:010-62770175

地址:北京清华大学学研大厦

邮编:100084

客户服务:010-62776969

责任编辑:袁勤勇

印刷者:北京市昌平环球印刷厂

装订者:三河市新茂装订有限公司

发行者:新华书店总店北京发行所

开本:185×260 印张:17 字数:401千字

版次:2006年7月第1版 2006年7月第1次印刷

书号:ISBN 7-302-12942-8/TP·8222

印数:1~3000

定价:23.00元

编审委员会

主任：蒋宗礼

副主任：李仲麟 何炎祥

委员：(排名不分先后)

王向东 宁 洪 朱庆生 吴功宜 吴 跃

张 虹 张 钢 张为群 余雪丽 陈志国

武 波 孟祥旭 孟小峰 胡金初 姚放吾

原福永 黄刘生 廖明宏 薛永生

秘书长：王听讲

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则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，决定组织编写一套《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”作出贡献。

清华大学出版社

前言

PREFACE

算法设计与分析是计算机科学技术中处于核心地位的一门专业基础课,越来越受到重视,CC2001和CCC2002都将“算法和复杂性”列为主领域,将算法设计策略、基本可计算性理论、P和NP问题类等算法设计技术和复杂性分析方法列为核心知识单元。

无论是计算科学还是计算实践,算法都在其中扮演着重要角色,算法被公认为是计算机科学的基石。翻开重要的计算机学术刊物,算法都占有一席之地,没有算法,计算机程序将不复存在。对于计算机专业的学生,学会读懂算法、设计算法,应该是一项最基本的要求,而发明算法则是计算机学者的最高境界。

提高学生的问题求解能力是高等教育的一个主要目标,在计算机科学的课程体系中,安排一门关于算法设计与分析的课程是非常必要的,因为这门课程能够引导学生的思维过程,告诉学生如何应用一些特定的算法设计策略来解决问题。学习算法还能够提高学生分析问题的能力。算法可以看作是解决问题的一类特殊方法——它不是问题的答案,而是经过精确定义的、用来获得答案的求解过程。因此,无论是否涉及计算机,特定的算法设计技术都可以看作是问题求解的有效策略。

本书将计算机经典问题和算法设计技术很好地结合起来,系统地介绍了算法设计技术及其在经典问题中的应用。通过同一算法设计技术在不同问题中的应用进行比较,牢固掌握算法设计技术的基本策略;通过不同的算法设计技术在同一问题中的应用进行比较,更容易体会到算法设计技术的思想方法,收到融会贯通的效果。所以,本书采用了模块化的设计思想,读者除了按本书组织的章节学习外,还可以将每种算法设计技术的问题提取出来,比较解决相同问题的不同解决方法。随着本书内容的不断展开,读者也将感受到综合应用多种算法设计技术有时可以更有效地解决问题。

全书共12章,第1章介绍了算法的基本概念和算法分析方法,第2章从算法的观点非形式化地介绍了NP完全理论,第3章~第11章分别

介绍了蛮力法、分治法、减治法、动态规划法、贪心法、回溯法、分支限界法、概率算法和近似算法等算法设计技术，第 12 章基于图灵机计算模型介绍了计算复杂性理论。

书中所有问题均给出了若干应用实例，每章还设有一个实验项目，通过设计提高学生创造性思维的培养。每章均附有一篇阅读材料，以通俗易懂的笔触介绍了算法领域的一些最新研究成果，保证知识的先进性。书中所有算法均给出了伪代码，大部分算法还给出了 C++ 描述。在算法介绍上，注重对问题求解过程的理解，注重算法设计思路和分析过程的讲解，体现了“授之以渔”的教学理念。

王涛老师收集和整理了本书的阅读材料，参加本书编写的还有胡明、许建潮、孙卫佳、逢焕利、刘钢、陈志雨等老师，研究生张倩、魏卓调试了本书的全部算法。

由于作者的知识 and 写作水平有限，书稿虽几经修改，仍难免有缺点和错误。热忱欢迎同行专家和读者批评指正，使本书在使用中不断改进、日臻完善。

作者的电子邮箱是：wanghm@mail.ccut.edu.cn。

作者

2006 年 2 月

目 录

CONTENTS

第 1 章 绪论	1
1.1 算法的基本概念	1
1.1.1 为什么要学习算法	1
1.1.2 算法及其重要特性	3
1.1.3 算法的描述方法	4
1.1.4 算法设计的一般过程	5
1.1.5 重要的问题类型	8
1.2 算法分析	10
1.2.1 渐进符号	10
1.2.2 最好、最坏和平均情况	12
1.2.3 非递归算法的分析	13
1.2.4 递归算法的分析	14
1.2.5 算法的后验分析	16
1.3 实验项目——求最大公约数	18
阅读材料——人工神经网络与 BP 算法	19
习题 1	21
第 2 章 NP 完全理论	25
2.1 下界	25
2.1.1 平凡下界	26
2.1.2 判定树模型	26
2.1.3 最优算法	27
2.2 算法的极限	28
2.2.1 易解问题与难解问题	28
2.2.2 实际问题难以求解的原因	30
2.2.3 不可解问题	32
2.3 P 类问题和 NP 类问题	34

2.3.1	判定问题	34
2.3.2	确定性算法与 P 类问题	35
2.3.3	非确定性算法与 NP 类问题	35
2.4	NP 完全问题	36
2.4.1	问题变换与计算复杂性归约	37
2.4.2	NP 完全问题的定义	38
2.4.3	基本的 NP 完全问题	40
2.4.4	NP 完全问题的计算机处理	41
2.5	实验项目——SAT 问题	43
	阅读材料——遗传算法	43
	习题 2	47
第 3 章	蛮力法	49
3.1	蛮力法的设计思想	49
3.2	查找问题中的蛮力法	50
3.2.1	顺序查找	50
3.2.2	串匹配问题	52
3.3	排序问题中的蛮力法	56
3.3.1	选择排序	56
3.3.2	起泡排序	57
3.4	组合问题中的蛮力法	58
3.4.1	生成排列对象	58
3.4.2	生成子集	58
3.4.3	0/1 背包问题	59
3.4.4	任务分配问题	59
3.5	图问题中的蛮力法	61
3.5.1	哈密顿回路问题	61
3.5.2	TSP 问题	62
3.6	几何问题中的蛮力法	63
3.6.1	最近对问题	63
3.6.2	凸包问题	64
3.7	实验项目——串匹配问题	65
	阅读材料——蚁群算法	67
	习题 3	70
第 4 章	分治法	73
4.1	概述	73
4.1.1	分治法的设计思想	73

4.1.2	分治法的求解过程	74
4.2	递归	75
4.2.1	递归的定义	75
4.2.2	递归函数的运行轨迹	77
4.2.3	递归函数的内部执行过程	77
4.3	排序问题中的分治法	78
4.3.1	归并排序	78
4.3.2	快速排序	80
4.4	组合问题中的分治法	83
4.4.1	最大子段和问题	83
4.4.2	棋盘覆盖问题	85
4.4.3	循环赛日程安排问题	87
4.5	几何问题中的分治法	88
4.5.1	最近对问题	89
4.5.2	凸包问题	90
4.6	实验项目——最近对问题	91
	阅读材料——鱼群算法	92
	习题 4	95
第 5 章	减治法	97
5.1	减治法的设计思想	97
5.2	查找问题中的减治法	98
5.2.1	折半查找	98
5.2.2	二叉查找树	100
5.3	排序问题中的减治法	101
5.3.1	堆排序	101
5.3.2	选择问题	105
5.4	组合问题中的减治法	106
5.4.1	淘汰赛冠军问题	106
5.4.2	假币问题	107
5.5	实验项目——8 枚硬币问题	109
	阅读材料——粒子群算法	109
	习题 5	112
第 6 章	动态规划法	115
6.1	概述	115
6.1.1	最优化问题	115
6.1.2	最优性原理	116

6.1.3	动态规划法的设计思想	117
6.2	图问题中的动态规划法	119
6.2.1	TSP 问题	119
6.2.2	多段图的最短路径问题	121
6.3	组合问题中的动态规划法	123
6.3.1	0/1 背包问题	123
6.3.2	最长公共子序列问题	126
6.4	查找问题中的动态规划法	128
6.4.1	最优二叉查找树	128
6.4.2	近似串匹配问题	132
6.5	实验项目——最大子段和问题	134
	阅读材料——文化算法	135
	习题 6	137
第 7 章	贪心法	139
7.1	概述	139
7.1.1	贪心法的设计思想	139
7.1.2	贪心法的求解过程	140
7.2	图问题中的贪心法	141
7.2.1	TSP 问题	141
7.2.2	图着色问题	144
7.2.3	最小生成树问题	145
7.3	组合问题中的贪心法	148
7.3.1	背包问题	148
7.3.2	活动安排问题	151
7.3.3	多机调度问题	153
7.4	实验项目——霍夫曼编码	155
	阅读材料——模拟退火算法	157
	习题 7	159
第 8 章	回溯法	161
8.1	概述	161
8.1.1	问题的解空间	161
8.1.2	解空间树的动态搜索(1)	163
8.1.3	回溯法的求解过程	165
8.1.4	回溯法的时间性能	166
8.2	图问题中的回溯法	168
8.2.1	图着色问题	168

8.2.2	哈密顿回路问题	170
8.3	组合问题中的回溯法	173
8.3.1	八皇后问题	173
8.3.2	批处理作业调度问题	175
8.4	实验项目——0/1 背包问题	177
	阅读材料——禁忌搜索算法	178
	习题 8	180
第 9 章	分支限界法	183
9.1	概述	183
9.1.1	解空间树的动态搜索(2)	183
9.1.2	分支限界法的设计思想	186
9.1.3	分支限界法的时间性能	188
9.2	图问题中的分支限界法	188
9.2.1	TSP 问题	188
9.2.2	多段图的最短路径问题	192
9.3	组合问题中的分支限界法	195
9.3.1	任务分配问题	195
9.3.2	批处理作业调度问题	198
9.4	实验项目——电路布线问题	200
	阅读材料——免疫算法	201
	习题 9	203
第 10 章	概率算法	205
10.1	概述	205
10.1.1	概率算法的设计思想	206
10.1.2	随机数发生器	207
10.2	舍伍德(Sherwood)型概率算法	207
10.2.1	快速排序	208
10.2.2	选择问题	209
10.3	拉斯维加斯(Las Vegas)型概率算法	210
10.3.1	八皇后问题	211
10.3.2	整数因子分解问题	212
10.4	蒙特卡罗(Monte Carlo)型概率算法	214
10.4.1	主元素问题	215
10.4.2	素数测试问题	216
10.5	实验项目——随机数发生器	218
	阅读材料——DNA 计算与 DNA 计算机	219

习题 10	221
第 11 章 近似算法	223
11.1 概述	223
11.1.1 近似算法的设计思想	223
11.1.2 近似算法的性能	224
11.2 图问题中的近似算法	225
11.2.1 顶点覆盖问题	225
11.2.2 TSP 问题	226
11.3 组合问题中的近似算法	228
11.3.1 装箱问题	228
11.3.2 子集和问题	231
11.4 实验项目——TSP 问题的近似算法	235
阅读材料——量子密码技术	235
习题 11	237
第 12 章 计算复杂性理论	239
12.1 计算模型	239
12.1.1 图灵机的基本模型	240
12.1.2 k 带图灵机和时间复杂性	241
12.1.3 离线图灵机和空间复杂性	244
12.2 P 类问题和 NP 类问题	245
12.2.1 非确定性图灵机	245
12.2.2 P 类语言和 NP 类语言	246
12.3 NP 完全问题	247
12.3.1 多项式时间变换	247
12.3.2 Cook 定理	248
12.4 实验项目——NP 完全问题树	251
阅读材料——算法优化策略	251
习题 12	254
参考文献	255

算法理论研究的是算法的设计技术和分析技术,前者是指面对一个问题,如何设计一个有效的算法;后者则是对已设计的算法,如何评价或判断其优劣。二者是相互依存的,设计出的算法需要检验和评价,对算法的分析反过来又将改进算法的设计。

1.1 算法的基本概念

算法的概念在计算机科学领域几乎无处不在,在各种计算机软件系统的实现中,算法设计往往处于核心地位。例如,操作系统是现代计算机系统中不可缺少的系统软件,操作系统的各个任务都是一个单独的问题,每个问题由操作系统中的一个子程序根据特定的算法来实现。用什么方法来设计算法,如何判定一个算法的优劣,所设计的算法需要占用多少时间资源和空间资源,在实现一个软件系统时,都是必须予以解决的重要问题。

1.1.1 为什么要学习算法

用计算机求解任何问题都离不开程序设计,而程序设计的核心是算法设计。一般来说,对程序设计的研究可以分为4个层次:算法、方法学、语言和工具,其中算法研究位于最高层次。算法对程序设计的指导可以延续几年甚至几十年,它不依赖于方法学、语言 and 工具的发展与变化。例如,用于数据存储和检索的 Hash 算法产生于 20 世纪 50 年代,用于排序的快速排序算法发明于 20 世纪 60 年代,但它们至今仍被人们广为使用,可是程序设计方法已经从结构化发展到面向对象,程序设计语言也变化了几代,至于编程工具很难维持 3 年不变。所以,对于从事计算机专业的人士来说,学习算法是非常必要的。

学习算法还能够提高人们分析问题的能力。算法可以看作是解决问

题的一类特殊方法——它不是问题的答案,而是经过精确定义的^①、用来获得答案的求解过程。因此,无论是否涉及计算机,特定的算法设计技术都可以看作是问题求解的有效策略。著名的计算机科学家科努思(Donald Knuth)是这样论述这个问题的:“受过良好训练的计算机科学家知道如何处理算法,如何构造算法、操作算法、理解算法以及分析算法,这些知识远不只是为了编写良好的计算机程序而准备的。算法是一种一般性的智能工具,一定有助于我们对其他学科的理解,不管是化学、语言学、音乐还是另外的学科。为什么算法会有这种作用呢?我们可以这样理解:人们常说,一个人只有把知识教给别人,才能真正掌握它。实际上,一个人只有把知识教给计算机,才能真正掌握它,也就是说,将知识表述为一种算法……比起简单地按照常规去理解事物,用算法将其形式化会使我们获得更加深刻的理解。”

算法研究的核心问题是时间(速度)问题。人们可能有这样的疑问:既然计算机硬件技术的发展使得计算机的性能不断提高,算法的研究还有必要吗?

计算机的功能越强大,人们就越想去尝试更复杂的问题,而更复杂的问题需要更大的计算量。现代计算技术在计算能力和存储容量上的革命仅仅提供了计算更复杂问题的有效工具,无论硬件性能如何提高,算法研究始终是推动计算机技术发展的关键。下面看几个例子。

1. 检索技术

20世纪50年代~60年代,检索的对象是规模比较小的数据集合。例如,编译系统中的标识符表,表中的记录个数一般在几十至数百这样的数量级。

20世纪70年代~80年代,数据管理采用数据库技术,数据库的规模在K级或M级,检索算法的研究在这个时期取得了巨大的进展。

20世纪90年代以来,Internet引起计算机应用的急速发展,海量数据的处理技术成为研究的热点,而且数据驻留的存储介质、数据的存储方法以及数据的传输技术也发生了许多变化,这些变化使得检索算法的研究更为复杂,也更为重要。

近年来,智能检索技术成为基于Web信息检索的研究热点。使用搜索引擎进行Web信息检索时,经常看到一些搜索引擎前50个搜索结果中几乎有一半来自同一个站点的不同页面,这是检索系统缺乏智能化的一种表现。另外,在传统的Web信息检索服务中,信息的传输是按pull的模式进行的,即用户找信息。而采用push的方式,是信息找用户,用户不必进行任何信息检索,就能方便地获得自己感兴趣的信息,这就是智能信息推送技术。这些新技术的每一项重要进步都与算法研究的突破有关。

2. 压缩与解压缩

随着多媒体技术的发展,计算机的处理对象由原来的字符发展到图像、图形、音频、视频等多媒体数字化信息,这些信息数字化后,其特点就是数据量非常庞大,同时,处理多媒

^① 算法固有的精确性限制了它能够解决的问题种类,比如说,我们无法找到一个使人生活快乐的算法,也不能找到一个使人富有和出名的算法。

体所需的高速传输速度也是计算机总线所不能承受的。因此,对多媒体数据的存储和传输都要求对数据进行压缩。声音文件的MP3压缩技术说明了压缩与解压缩算法研究的巨大成功,一个播放3~4分钟歌曲的MP3文件通常只需3MB左右的磁盘空间。

3. 信息安全与数据加密

在计算机应用迅猛发展的同时,也面临着各种各样的威胁。一位酒店经理曾经描述了这样一种可能性:“如果我能破坏网络的安全性,想想你在网络上预订酒店房间所提供的信息吧!我可以得到你的名字、地址、电话号码和信用卡号码,我知道你现在的位置,将要去哪儿,何时去,我也知道你支付了多少钱,我已经得到足够的信息来盗用你的信用卡!”这的确是一个可怕的情景。所以,在电子商务中,信息安全是最关键的问题,保证信息安全的一个方法就是对需要保密的数据进行加密。在这个领域,数据加密算法的研究是绝对必需的,其必要性与计算机性能的提高无关。

1.1.2 算法及其重要特性

算法(algorithm)被公认为是计算机科学的基石。通俗地讲,算法是解决问题的方法。严格地说,算法是对特定问题求解步骤的一种描述,是指令的有限序列,此外,算法还必须满足下列5个重要特性(如图1.1所示):

(1) 输入:一个算法有零个或多个输入。算法的输入来源于两种方式:一种是从外界获得数据,另一种是由算法自己产生被处理的数据。

(2) 输出:一个算法有一个或多个输出。既然算法是为解决问题而设计的,那么算法实现的最终目的就是要获得问题的解。没有输出的算法是无意义的。

(3) 有穷性:一个算法必须总是(对任何合法的输入)在执行有穷步之后结束,且每一步都在有穷时间内完成。

(4) 确定性:算法中的每一条指令必须有确切的含义,不存在二义性。并且,在任何条件下,对于相同的输入只能得到相同的输出。

(5) 可行性:算法描述的操作可以通过已经实现的基本操作执行有限次来实现。

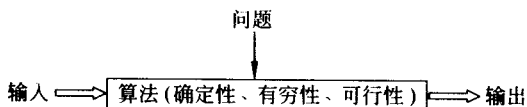


图 1.1 算法的概念

概念回顾

算法和程序不同。程序(program)是对一个算法使用某种程序设计语言的具体实现,原则上,算法可以用任何一种程序设计语言来实现。算法的有穷性意味着不是所有的计算机程序都是算法。