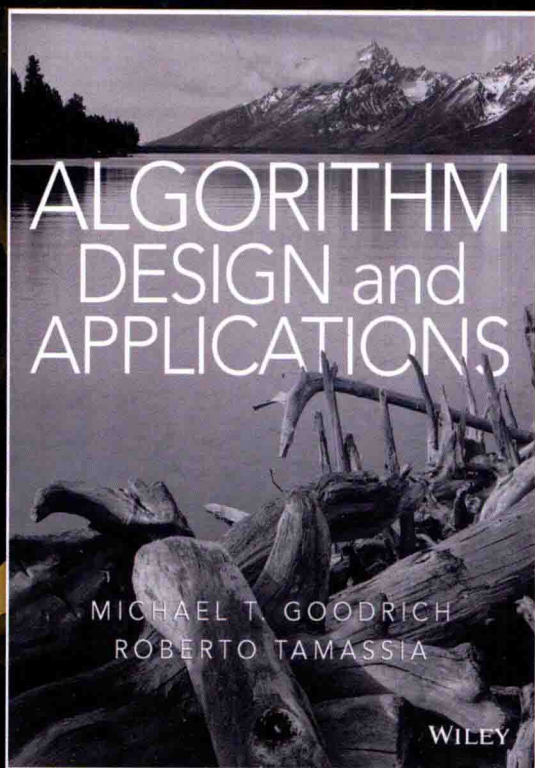


# 算法设计与应用

[美] 迈克尔 T. 古德里奇 (Michael T. Goodrich) 著  
    罗伯托·塔马西亚 (Roberto Tamassia)

乔海燕 李恣炜 王烁程 译

Algorithm Design and Applications



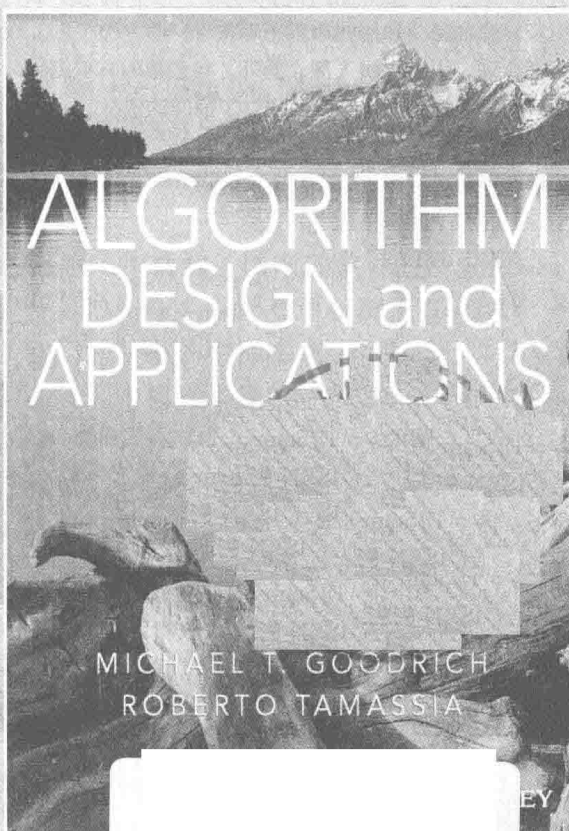
计 算 机 科 学 丛 书

# 算法设计与应用

[美] 迈克尔 T. 古德里奇 (Michael T. Goodrich) 著  
罗伯托·塔马西亚 (Roberto Tamassia)

乔海燕 李恣炜 王烁程 译

Algorithm Design and Applications



机械工业出版社  
China Machine Press

## 图书在版编目 ( CIP ) 数据

算法设计与应用 / (美) 迈克尔 T. 古德里奇 (Michael T. Goodrich), (美) 罗伯托·塔马西亚 (Roberto Tamassia) 著; 乔海燕, 李焱炜, 王烁程译. —北京: 机械工业出版社, 2017.10 (计算机科学丛书)

书名原文: Algorithm Design and Applications

ISBN 978-7-111-58277-9

I. 算… II. ①迈… ②罗… ③乔… ④李… ⑤王… III. 算法设计 IV. TP301.6

中国版本图书馆 CIP 数据核字 (2017) 第 259840 号

本书版权登记号: 图字 01-2016-6253

Copyright © 2015 John Wiley & Sons, Inc.

All rights reserved. This translation published under license. Authorized translation from the English language edition, entitled Algorithm Design and Applications, ISBN 978-1-118-33591-8, by Michael T. Goodrich, Roberto Tamassia, Published by John Wiley & Sons. No part of this book may be reproduced in any form without the written permission of the original copyrights holder.

本书中文简体字版由约翰·威立父子公司授权机械工业出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

本书封底贴有 Wiley 防伪标签, 无标签者不得销售。

本书全面系统地介绍算法设计和算法应用的各个领域, 内容涵盖经典数据结构、经典算法、算法分析方法、算法设计方法以及算法在各个领域的应用, 还包含一些高级主题。本书采用应用驱动的方法引入各章内容, 内容编排清晰合理, 讲解由浅入深。此外, 各章都附有巩固练习、创新练习和应用练习三种类型的题目, 为读者理解和掌握算法设计和应用提供了很好的素材。

本书可作为高等院校计算机及相关专业“数据结构和算法”课程的本科生、研究生教材, 也可作为算法理论和实践工作者的参考手册。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 唐晓琳

责任校对: 殷虹

印刷: 中国电影出版社印刷厂

版次: 2018 年 1 月第 1 版第 1 次印刷

开本: 185mm × 260mm 1/16

印张: 32.75

书号: ISBN 978-7-111-58277-9

定价: 139.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的优势，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与 Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage 等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出 Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson 等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

华章网站：[www.hzbook.com](http://www.hzbook.com)

电子邮件：[hzjsj@hzbook.com](mailto:hzjsj@hzbook.com)

联系电话：(010)88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



华章科技图书出版中心

本书全面系统地介绍数据结构、算法设计、算法分析和算法应用。全书内容丰富，讲解由浅入深，既包含经典数据结构和算法、算法设计方法和分析方法以及算法应用，又包含众多高级主题，包括多维搜索、计算几何、密码算法、傅里叶变换和线性规划等。

本书采用应用驱动的方式组织各章内容。书中算法均给出伪代码描述，附带图文并茂的运行过程例子，以及严谨的复杂度分析。每章都附有大量的练习，并分为巩固练习、创新练习和应用练习三种类型，其中应用练习是将刚刚学习的知识应用到实际生活的具体场景中，能极好地加深读者对知识的理解。这种内容和练习的安排，为读者全面深刻地掌握有关内容提供了很好的素材和途径。

全书分六个部分。第一部分系统地讲解基础数据结构与算法，包括栈、队列、二叉搜索树、平衡二叉搜索树、优先队列和堆、哈希表和并查集结构等。通过学习这些章节，读者能由浅入深地掌握常见的数据结构，并为学习后续内容打下坚实的基础。

第二部分讲解排序和选择，包括归并排序和快速排序算法，以及如何进行快速的排序和选择。排序算法是极为重要的一类算法，本部分首先讲解了基于比较的归并排序和快速排序算法，然后讲解了桶排序和基数排序等算法，还讲解了选择和加权中位数算法。

第三部分讲解算法设计常用的技术，包括贪心法、分治法、动态规划、图的遍历。这些均为算法设计中常用的方法，本书用很多具体的算法和实例来讲解如何运用这些方法来设计算法，使得读者能够举一反三，根据实际问题利用这些方法来设计算法。

第四部分讲解图的算法，包括最短路径算法、最小生成树算法，以及网络流和匹配。通过学习这些章节，读者能掌握多种求解最短路径、最小生成树的算法，通过学习网络流和匹配的相关概念和算法，对图有更深入的理解。

第五部分讲解计算困难问题，包括 NP 完全问题，以及近似算法。本部分主要讨论 NP 难的问题，为解决这些计算困难问题，讨论了一些近似算法。

第六部分讲解高级主题，包括随机算法、B 树和外部存储器、多维搜索、计算几何、字符串算法、密码学、快速傅里叶变换和线性规划。这一部分覆盖面较广，涉及计算机科学的各个领域，读者可以借此了解更多有趣的主题。

本书作者 Michael T. Goodrich 和 Roberto Tamassia 均为知名大学教授，算法研究领域知名学者。他们不仅对算法理论和应用有深刻见解，而且对算法教学有实际经验。如作者在前言中所讲，本书涵盖内容丰富，既可作为数据结构和算法核心课程的教材，也可作为高年级本科生和研究生的高级算法课程的教材。

本书在翻译过程中得到了许多同行和编辑的大力支持。特别是，任利明对有关棒球的规则和术语做了说明，黄森中对图题的德文做了翻译。在此一并对给予我们帮助的同行和编辑表示感谢！

另外，为了便于读者的阅读，我们对算法伪代码也做了必要的翻译。限于译者水平，译文中难免出现疏漏和错误，欢迎大家批评指正！

译者

2017年4月于广州大学城

本书全面地介绍了计算机算法和数据结构的设计和分析。书中各章相对独立，所以教师和读者可以自由选择感兴趣的章节。此外，本书内容广泛，既包含了经典的算法，也包含了新兴的算法，具体如下：

- 渐近分析数学，包括平摊和随机化。
- 通用算法设计技术，包括贪心法、分治法和动态规划。
- 数据结构，包括列表、树、堆、搜索树、B 树、哈希表、跳跃表、并查集结构和多维树。
- 算法框架，包括 NP 完全性、近似算法和外存算法。
- 基本算法，包括排序、图算法、计算几何、数值算法、密码、快速傅里叶变换 (FFT) 和线性规划。

## 应用驱动方法

计算机科学已经进入一个令人兴奋的时代。计算机已经从早期的计算引擎发展到现在的信息处理器，其应用遍布各个领域。此外，互联网的扩展推动了计算机在社会和商业中的新范式和新模式。例如，计算机可以用来存储和检索大规模数据，并且用于许多其他的应用领域，如运动、视频游戏、生物、制药、社交网络、工程和科学。因此，我们认为算法的讲授既要强调其数学分析，也要突出其实际应用。

为了达到这个目的，每章开篇都有该章主题应用的一个简短讨论。这些应用有的来自于主题的实际应用，有的是设想该章主题在实践中的可能应用。这些讨论的意图是为读者阅读各章时提供一定的背景和实际应用动机。除了这些应用的动机外，我们还给出算法的详细伪代码描述和完整的数学分析。事实上，我们认为数学的严谨性有其实际的意义。

## 写给教师

本书的结构便于教师自由地选择和讲授内容。各章节之间的依存度已经降至很低，以便教师可以按照自己喜欢的顺序授课。此外，依据内容的深度，每章的讲授时间在 1~3 节课。

## 课程样例

本书可作为多个课程的教材。例如，可用于算法核心课程的教材，即经典 CS7。另外，本书也可以用于高年级本科生或者研究生的数据结构课程、算法课程，或者两学期连续教授这两个课程的教材。为了突出这些选择，下面为这些可能的课程给出教学大纲样例。

### 算法核心课程 (CS7) 大纲样例

#### 第 1 章 算法分析

(跳读、略读或复习第 2~4 章的基本数据结构)<sup>⊖</sup>

- 第 5 章 优先队列和堆
- 第 6 章 散列表
- 第 7 章 并查集结构
- 第 8 章 归并排序和快速排序
- 第 9 章 快速排序和选择(如果时间允许)
- 第 10 章 贪心法
- 第 11 章 分治法
- 第 12 章 动态规划
- 第 13 章 图及遍历
- 第 14 章 最短路径
- 第 15 章 最小生成树
- 第 16 章 网络流和匹配(如果时间允许)
- 第 17 章 NP 完全性
- 第 18 章 近似算法

如果时间允许,可从第 19~26 章中选择内容,包括随机算法、计算几何、字符串算法、密码学、快速傅里叶变换(FFT)和线性规划。

#### 高年级本科生或者研究生的数据结构课程大纲样例

- 第 1 章 算法分析
- 第 2 章 基本数据结构
- 第 3 章 二叉搜索树
- 第 4 章 平衡二叉搜索树
- 第 5 章 优先队列和堆
- 第 6 章 散列表
- 第 7 章 并查集结构
- 第 8 章 归并排序和快速排序
- 第 13 章 图及遍历
- 第 14 章 最短路径
- 第 15 章 最小生成树
- 第 20 章 B 树和外部存储器
- 第 21 章 多维搜索

#### 高年级本科生或者研究生的算法课程大纲样例

(跳读、略读或者复习第 1~8 章)

- 第 9 章 快速排序和选择
- 第 10 章 贪心法
- 第 11 章 分治法
- 第 12 章 动态规划

⊖ 这些内容以及第 5 章和第 6 章是数据结构课程的基本内容,也是本课程的先行课。

- 第 16 章 网络流和匹配
- 第 17 章 NP 完全性
- 第 18 章 近似算法
- 第 19 章 随机算法
- 第 22 章 计算几何
- 第 23 章 字符串算法
- 第 24 章 密码学
- 第 25 章 快速傅里叶变换
- 第 26 章 线性规划

这门课程既可以作为一门独立的课程讲授，也可与上面的高级数据结构课程联合讲授。当然，还有其他的选择，在此不再赘述，将这些内容的创意安排留给教师。

### 三类练习

本书包含了 800 多个练习，分为三类：

- 巩固练习，测试对章节内容的理解。
- 创新练习，测试能否创造性地利用各章的技术方法。
- 应用练习，测试能否将各章内容应用于实际问题或者设想的问题。

这些练习的分布大致为巩固练习占 35%，创新练习占 40%，应用练习占 25%。

### 网络增值学习

本书有一个网站：

<http://www.wiley.com/college/goodrich/>

这个网站包含了章节内容的附加学习资源，特别为学生提供了以下内容：

- 本书大部分内容的 PDF 演示讲义。
- 某些练习的提示。

对于一些学生来说，有些创新练习和应用练习可能具有挑战性，因此他们会对这些提示感兴趣。

我们也为使用本书作为教材的教师提供了一个教学支持网站<sup>①</sup>，包括下列辅助资源：

- 本书一些练习的解。
- 本书大部分内容的可编辑 PowerPoint 演示文稿。

### 预备知识

本书假定读者具有一定的预备知识。特别是，假定读者有基本的数据结构知识，如数组和链表，并且对高级程序设计语言如 C、C++、Java 或者 Python 有一定的理解。因此，所有的算法都是用高级的伪代码描述，略去了一些细节（如错误条件测试），但对于具备一定知识的读者来讲，将算法描述转换为程序代码是容易的。

在数学背景方面，本书假定读者熟悉指数、对数、求和、极限和初等概率。尽管如此，本

① 关于本书教辅资源，只有使用本书作为教材的教师才可以申请，需要的教师可向约翰·威立出版公司北京代表处申请，电话：010-8418 7869，电子邮件：sliang@wiley.com。——编辑注



书第 1 章仍然复习了这些概念，并在附录 A 中给出了一些其他常用的数学知识，包括初等概率。

## 关于作者

Goodrich 教授和 Tamassia 教授是研究算法和数据结构的知名学者，在这个研究领域发表了许多论文，并应用于计算机安全、密码学、网络计算、信息可视化和几何计算。他们是美国国家科学基金会、陆军研究处和国防部高级研究计划局资助的许多项目的主要研究人员，在教育技术研究领域也相当活跃。

Michael Goodrich 于 1987 年在普渡大学计算机科学系获得博士学位。他是加州大学欧文分校计算机科学系首席教授。在这之前他是约翰霍普金斯大学的教授。他的研究兴趣包括算法的分析、设计和实现，以及数据安全、云计算、绘图和计算几何。他是富布赖特学者，美国科学促进会 (AAAS)、计算机协会 (ACM) 以及电气和电子工程师协会 (IEEE) 的会士。他是 IEEE 计算机协会技术成就奖、ACM 卓越服务奖和 Pond 本科教学优秀奖的获得者，还是《国际计算几何与应用期刊》(IJCGA) 和《图算法与应用期刊》(JGAA) 咨询委员会委员。

Roberto Tamassia 于 1988 年在美国伊利诺伊大学厄巴纳 - 香槟分校电子与计算机工程系获得博士学位。他是布朗大学计算机科学系 Plastech 教授，目前是布朗几何计算中心主任。他的研究兴趣包括：数据安全，应用密码学，云计算，算法的分析、设计和实现，绘图，计算几何。他是美国科学促进会 (AAAS)、计算机协会 (ACM) 以及电气和电子工程师协会 (IEEE) 的会士。他也是 IEEE 计算机协会技术成就奖的获得者。Tamassia 是《图算法与应用期刊》(JGAA) 和绘图学术会议的共同创始人，现在是 JGAA 的联合主编。

## 致谢

我们在算法研究和教育项目中有许多合作者，这些合作者帮助我们改进了本书的设计和內容。尤其要感谢 Jeff Achter、Vesselin Arnaudov、James Baker、Ryan Baker、Benjamin Boer、John Boreiko、Devin Borland、Lubomir Bourdev、Ulrik Brandes、Stina Bridgeman、Bryan Cantrill、Yi-Jen Chiang、Robert Cohen、David Ellis、David Emory、Jody Fanto、Ben Finkel、Peter Fröhlich、Ashim Garg、David Ginat、Natasha Gelfand、Esha Ghosh、Michael Goldwasser、Mark Handy、Michael Horn、Greg Howard、Benoît Hudson、Jovanna Ignatowicz、James Kelley、Evgenios Kornaropoulos、Giuseppe Liotta、David Mount、Jeremy Mullendore、Olga Ohrimenko、Seth Padowitz、Bernardo Palazzi、Charalampos Papamantou、James Piechota、Daniel Polivy、Seth Proctor、Susannah Raub、Haru Sakai、John Schultz、Andrew Schwerin、Michael Shapiro、Michael Shim、Michael Shin、Galina Shubina、Amy Simpson、Christian Straub、Ye Sun、Nikos Triandopoulos、Luca Vismara、Danfeng Yao、Jason Ye 和 Eric Zamore。

感谢编辑 Beth Golub 对本项目的热情支持，Wiley 的生产团队也给予我们很大帮助。许多人为本书做出了贡献，包括 Jayne Ziemba、Jennifer Welter、Debbie Martin、Chris Ruel、Julie Kennedy、Wanqian Ye、Joyce Poh 和 Janis Soo。

特别感谢 Michael Bannister、Jenny Lam 和 Joseph Simons 对第 26 章所做的贡献。感谢 Sidhartha Sen 和 Robert Tarjan 关于平衡搜索树富有启发的讨论。

真诚地感谢外部评审人员，特别是 Jack Snoeyink，他的详细评论和富有建设性的批评对本书非常有益。感谢其他外部评审人员，包括 John Donald、Hui Yang、Nicholas Tran、John Black、

My Thai、Dana Randall、Ming-Yang Kao、Qiang Cheng、Ravi Janarden、Fikret Ercal、Jack Snoeyink、S. Muthukrishnan、Elliot Anshelevich、Mukkai Krishnamoorthy、Roxanne Canosa、Michael Cutler、Roger Crawfis、Glencora Borradaile、Jennifer Welch。

本书手稿主要用 Latex 完成文字部分，用微软 PowerPoint、Visio 和 Adobe FrameMaker 完成图形部分。

最后，真诚地感谢 Isabel Cruz、Karen Goodrich、Giuseppe Di Battista、Franco Preparata、Ioannis Tollis 和我们的父母在本书编写的各个阶段给予的建议、鼓励和支持。感谢他们在我们写书之余为我们提供了多彩的生活。

Michael T. Goodrich

Roberto Tamassia

# 目 录

Algorithm Design and Applications

出版者的话	
译者序	
前言	
第1章 算法分析	1
1.1 分析算法	2
1.1.1 伪代码	2
1.1.2 随机存取机模型	4
1.1.3 基本操作数目的计算	4
1.1.4 递归算法的分析	6
1.1.5 渐近表示法	6
1.1.6 渐近表示法的重要性	10
1.2 相关数学知识复习	11
1.2.1 求和	11
1.2.2 对数和幂	12
1.2.3 简单的证明技术	13
1.2.4 概率基础	16
1.3 算法分析案例	18
1.3.1 最大子数组问题的第一个解	18
1.3.2 一种改进的求最大子数组 算法	19
1.3.3 线性时间的最大子数组算法	20
1.4 平摊分析	21
1.4.1 平摊技术	22
1.4.2 对一个可扩展数组实现的 分析	24
1.5 练习	26
本章注记	31
<b>第一部分 数据结构</b>	
第2章 基本数据结构	34
2.1 栈和队列	35
2.1.1 栈	35
2.1.2 队列	37
2.2 列表	38
2.2.1 基于索引的列表	39
2.2.2 链表	40
2.3 树	44
2.3.1 树的定义	45
2.3.2 树的遍历	46
2.3.3 二叉树	49
2.3.4 表示树的数据结构	53
2.4 练习	55
本章注记	58
第3章 二叉搜索树	59
3.1 搜索和更新	60
3.1.1 二叉搜索树的定义	62
3.1.2 二叉搜索树中的搜索	62
3.1.3 二叉搜索树中的插入	64
3.1.4 二叉搜索树中的删除	64
3.1.5 二叉搜索树的性能	65
3.2 范围查询	66
3.3 基于索引的搜索	68
3.4 随机构造二叉搜索树	70
3.5 练习	72
本章注记	75
第4章 平衡二叉搜索树	76
4.1 秩和旋转	77
4.2 AVL树	79
4.3 红黑树	82
4.4 弱AVL树	85
4.5 伸展树	91
4.6 练习	97
本章注记	101
第5章 优先队列和堆	102
5.1 优先队列	103
5.2 PQ排序、选择排序和插入排序	103
5.2.1 选择排序	104
5.2.2 插入排序	106
5.3 堆	107
5.3.1 基于数组结构的二叉树	109
5.3.2 堆中的插入	110
5.3.3 堆中的删除	112

5.4 堆排序 .....	114	8.2.1 随机快速排序 .....	165
5.5 扩展优先队列 .....	118	8.2.2 原地快速排序 .....	166
5.6 练习 .....	119	8.3 基于比较的排序的下界 .....	168
本章笔记 .....	122	8.4 练习 .....	169
第6章 散列表 .....	123	本章笔记 .....	172
6.1 映射 .....	124	第9章 快速排序和选择 .....	173
6.1.1 映射的定义 .....	124	9.1 桶排序和基数排序 .....	174
6.1.2 查找表 .....	125	9.1.1 桶排序 .....	174
6.2 散列函数 .....	126	9.1.2 基数排序 .....	175
6.2.1 分量求和 .....	127	9.2 选择 .....	176
6.2.2 多项式求值函数 .....	127	9.2.1 随机快速选择 .....	176
6.2.3 基于表格的散列 .....	128	9.2.2 确定性选择 .....	178
6.2.4 取模 .....	128	9.3 加权中位数 .....	179
6.2.5 随机线性和多项式函数 .....	129	9.4 练习 .....	181
6.3 碰撞处理与再散列 .....	129	本章笔记 .....	183
6.3.1 拉链法 .....	129		
6.3.2 开放寻址法 .....	130	<b>第三部分 基本技术</b>	
6.3.3 线性探测 .....	131	第10章 贪心法 .....	186
6.3.4 平方探测 .....	133	10.1 分份背包问题 .....	187
6.3.5 双重散列 .....	133	10.2 任务调度 .....	189
6.3.6 再散列 .....	134	10.3 文本压缩和哈夫曼编码 .....	191
6.4 布谷鸟散列 .....	134	10.4 练习 .....	195
6.5 通用散列 .....	138	本章笔记 .....	197
6.6 练习 .....	140	第11章 分治法 .....	198
本章笔记 .....	142	11.1 递推与主定理 .....	199
第7章 并查集结构 .....	143	11.2 整数乘法 .....	203
7.1 并查集及其应用 .....	144	11.3 矩阵乘法 .....	204
7.1.1 连通分支 .....	144	11.4 极大点集问题 .....	205
7.1.2 迷宫建筑和渗透理论 .....	145	11.5 练习 .....	206
7.2 基于列表的实现 .....	146	本章笔记 .....	209
7.3 基于树的实现 .....	148	第12章 动态规划 .....	210
7.4 练习 .....	153	12.1 矩阵连乘 .....	211
本章笔记 .....	155	12.2 通用技术 .....	213
		12.3 望远镜调度 .....	213
		12.4 博弈策略 .....	216
		12.4.1 硬币行 .....	216
		12.4.2 概率博弈策略与逆向 归纳法 .....	217
		12.5 最长公共子序列问题 .....	219
		12.5.1 问题定义 .....	219
<b>第二部分 排序和选择</b>			
第8章 归并排序和快速排序 .....	158		
8.1 归并排序 .....	159		
8.1.1 分而治之 .....	159		
8.1.2 归并排序和递推方程 .....	162		
8.2 快速排序 .....	163		

12.5.2 应用动态规划解 LCS 问题	219
12.6 0-1 背包问题	221
12.7 练习	223
本章笔记	227
第 13 章 图及遍历	228
13.1 图的术语和表示方法	228
13.1.1 图的一些术语	229
13.1.2 图的操作	231
13.1.3 表示图的数据结构	232
13.2 深度优先搜索	235
13.3 广度优先搜索	237
13.4 有向图	239
13.4.1 遍历有向图	240
13.4.2 传递闭包	242
13.4.3 有向 DFS 和垃圾回收	244
13.4.4 有向无环图	246
13.5 双连通分量	249
13.6 练习	252
本章笔记	255

#### 第四部分 图算法

第 14 章 最短路径	258
14.1 单源最短路径	258
14.2 Dijkstra 算法	259
14.3 Bellman-Ford 算法	264
14.4 有向无环图中的最短路径	266
14.5 所有顶点对之间的最短路径	268
14.5.1 动态规划最短路径算法	268
14.5.2 通过矩阵乘法计算最短路径	269
14.6 练习	272
本章笔记	275
第 15 章 最小生成树	276
15.1 最小生成树的性质	276
15.2 Kruskal 算法	279
15.3 Prim-Jarník 算法	282
15.4 Borůvka 算法	285
15.5 练习	286
本章笔记	288
第 16 章 网络流和匹配	290
16.1 流与割	290

16.1.1 割	292
16.1.2 剩余容量和增流路径	293
16.2 最大流算法	295
16.2.1 Ford-Fulkerson 算法	295
16.2.2 Edmonds-Karp 算法	297
16.3 最大二分图匹配	300
16.4 棒球赛的淘汰	301
16.5 最低成本流	302
16.6 练习	307
本章笔记	309

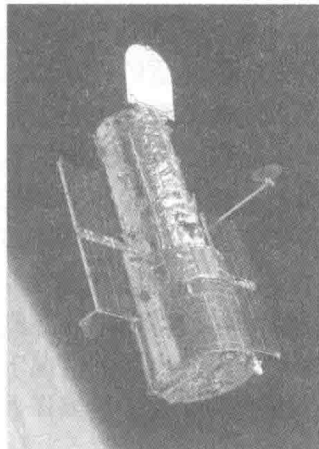
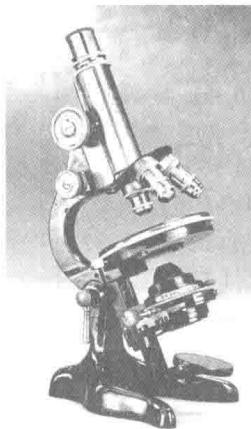
#### 第五部分 计算困难问题

第 17 章 NP 完全性	312
17.1 P 和 NP	313
17.1.1 定义复杂类 P 和 NP	314
17.1.2 一些有趣的 NP 问题	316
17.2 NP 完全性	318
17.2.1 多项式时间归约和 NP 难度	318
17.2.2 Cook-Levin 定理	318
17.2.3 如何证明一个问题是 NP 完全问题	319
17.3 合取范式可满足问题和 3 可满足问题	321
17.4 顶点覆盖、团和集合覆盖	324
17.5 子集和与背包问题	326
17.6 哈密顿回路和 TSP	328
17.7 练习	330
本章笔记	333
第 18 章 近似算法	334
18.1 几何旅行商问题	336
18.1.1 Metric-TSP 的一个 2 近似算法	336
18.1.2 Christofides 近似算法	337
18.2 覆盖问题的近似	338
18.2.1 顶点覆盖的 2 近似算法	338
18.2.2 集合覆盖的对数近似	339
18.3 多项式时间近似方法	340
18.4 回溯和分支定界	342
18.4.1 回溯法	342

18.4.2 分支定界法 .....	344	21.2.3 优先范围树 .....	405
18.5 练习 .....	344	21.3 四叉树和 $k$ -d 树 .....	406
本章注记 .....	346	21.3.1 四叉树 .....	406
<b>第六部分 高级主题</b>		21.3.2 $k$ -d 树 .....	407
第 19 章 随机算法 .....	350	21.4 练习 .....	409
19.1 随机排列的生成 .....	351	本章注记 .....	411
19.2 稳定婚姻和优惠券收集 .....	352	第 22 章 计算几何 .....	412
19.2.1 优惠券收集问题分析 .....	353	22.1 几何对象上的操作 .....	413
19.2.2 稳定婚姻问题 .....	354	22.2 凸壳 .....	416
19.3 最小割 .....	356	22.2.1 礼品包装算法 .....	417
19.3.1 收缩边 .....	357	22.2.2 Graham 扫描算法 .....	419
19.3.2 计算最小割 .....	357	22.3 线段相交 .....	421
19.3.3 更快的算法 .....	359	22.4 求最近点对 .....	423
19.4 寻找素数 .....	360	22.5 练习 .....	425
19.5 切尔诺夫界 .....	364	本章注记 .....	428
19.5.1 马尔可夫不等式 .....	364	第 23 章 字符串算法 .....	429
19.5.2 示性随机变量之和 .....	364	23.1 字符串操作 .....	430
19.5.3 几何型随机变量之和 .....	366	23.2 Boyer-Moore 算法 .....	431
19.6 跳跃表 .....	367	23.3 Knuth-Morris-Pratt 算法 .....	435
19.6.1 搜索 .....	368	23.4 基于散列的词典匹配 .....	437
19.6.2 更新操作 .....	368	23.5 字典树 .....	441
19.6.3 跳跃表的概率分析 .....	370	23.5.1 标准字典树 .....	441
19.7 练习 .....	371	23.5.2 压缩字典树 .....	443
本章注记 .....	374	23.5.3 后缀字典树 .....	445
第 20 章 B 树和外部存储器 .....	376	23.5.4 搜索引擎 .....	447
20.1 外部存储器 .....	377	23.6 练习 .....	448
20.2 (2,4) 树和 B 树 .....	379	本章注记 .....	450
20.2.1 多叉搜索树 .....	379	第 24 章 密码学 .....	451
20.2.2 (2,4) 树 .....	381	24.1 最大公约数 .....	452
20.2.3 $(a,b)$ 树和 B 树 .....	386	24.1.1 一些初等数论知识 .....	452
20.3 外部存储器排序 .....	389	24.1.2 欧几里得 GCD 算法 .....	453
20.4 在线缓存算法 .....	391	24.2 模运算 .....	454
20.5 练习 .....	396	24.2.1 模幂运算 .....	456
本章注记 .....	398	24.2.2 模乘法逆 .....	458
第 21 章 多维搜索 .....	399	24.3 加密操作 .....	459
21.1 范围树 .....	400	24.4 RSA 密码系统 .....	461
21.2 优先搜索树 .....	403	24.5 El Gamal 密码系统 .....	463
21.2.1 构造优先搜索树 .....	403	24.6 练习 .....	464
21.2.2 在优先搜索树中搜索 .....	404	本章注记 .....	465

第 25 章 快速傅里叶变换 .....	466	26.2.1 松弛型 .....	484
25.1 卷积 .....	467	26.2.2 扩展的例子 .....	485
25.2 原始单位根 .....	468	26.2.3 单纯形算法 .....	487
25.3 离散傅里叶变换 .....	469	26.3 对偶 .....	488
25.4 快速傅里叶变换算法 .....	471	26.4 线性规划的应用 .....	491
25.5 练习 .....	475	26.5 练习 .....	492
本章注记 .....	478	本章注记 .....	497
第 26 章 线性规划 .....	479	附录 A 一些有用的数学知识 .....	498
26.1 定义问题 .....	480	参考文献 .....	501
26.2 单纯形法 .....	483		

## 算法分析



显微镜：美国政府图像，来自 DeWitt Stetten, Jr. 医学研究博物馆 N. I. H. 医学仪器陈列馆。哈勃望远镜：美国政府图像，NASA STS-125 宇航员 - 2009 年 5 月 25 日拍摄

科学家经常需要区别对待不同规模的研究对象。对于从微观到天体的不同规模物体，科学家研发了不同的观察工具。同样，计算机科学家也必须处理研究对象的规模问题，但是他们考虑的主要是数据量大小问题，而不是对象的物理规模。在信息技术的世界中，可扩展性 (scalability) 是指系统能够很好地适应不断增大的输入量或工作量的能力。计算机系统是否具有可扩展性，意味着技术解决方案是否可以成功地应用于市场或者科学问题，还是当数据量增加时该方案变得不可行。本书考虑的是可扩展的算法和数据结构。

简单地说，算法 (algorithm) 是在有限的时间内一步步完成某个任务的过程，数据结构 (data structure) 是系统地组织和访问数据的方法。这些概念是计算的核心，本书探讨设计和实现正确且有效的数据结构和算法的范式 and 原则。为了度量算法和数据结构的可扩展程度，必须有精确地分析可扩展性的方法。

本书使用的主要分析工具是描述算法或数据结构操作的运行时间 (running time) 和空间的使用 (space usage)。因为时间是宝贵的资源，因此运行时间是可扩展性的一种自然度量。从经济和科学的应用方面来讲，运行时间是一个重要的考虑因素，因为每个人都希望计算机应用程序运行得尽可能快。

本章首先描述算法分析的基本框架，包括描述算法的语言、该语言的计算模型以及计算运行时间所考虑的因素。关于递归算法的分析，本章也做了简短介绍。运行时间的大  $O$  表示法在 1.1.5 节引入。这些工具是设计和分析算法的主要理论工具。

在继续介绍算法分析框架之前，1.2 节将重温算法分析需要的一些重要的数学知识，包括求和、对数、证明技术和基础概率。在有了算法分析的背景知识后，1.3 节讨论一个算法分析案例，这是在面试中经常出现的一个测试问题。1.4 节讨论一个有趣的分析技术，即所谓的平摊技术，用于描述多个操作的总体性能。最后是本章的练习，其中有些问题是根据几家主流软件公司和互联网公司面试过程中常见的问题而设计的。



## 1.1 分析算法

算法或数据结构操作的运行时间通常取决于一系列因素，那么，度量的正确方法是什么呢？如果一个算法已经实现，可以通过记录程序在不同的测试输入下的实际运行时间来度量。通过程序语言中的系统调用或运行程序的操作系统的相应命令，这种方法能够准确地度量运行时间。一般情况下，我们关心的是运行时间如何依赖于输入的规模。为此，可以对许多不同规模的测试输入记录相应的运行时间，并将结果可视化，即用  $x$  轴表示输入的规模  $n$ 、 $y$  轴表示相应的运行时间  $t$  作图（见图 1-1）。这种方法需要选择合适的样本输入，并做大量的测试，才能从中对算法做出合理的统计推断。

一般情况下，一个算法或数据结构方法的运行时间随着输入规模的增加而增加，当然对于相同规模的输入运行时间也可能不一样。同时，算法的实现、编译和运行的硬件环境（处理器、时钟速率、内存、磁盘等）和软件环境（操作系统、程序语言、编译器、解释器等）等都对算法的运行时间有影响。在所有其他因素相同的情况下，如果一个计算机的处理器更快，或者一个程序被编译成本地机器代码运行而不是在虚拟机上解释执行，那么相同的算法在相同的输入数据上的运行时间会更短。

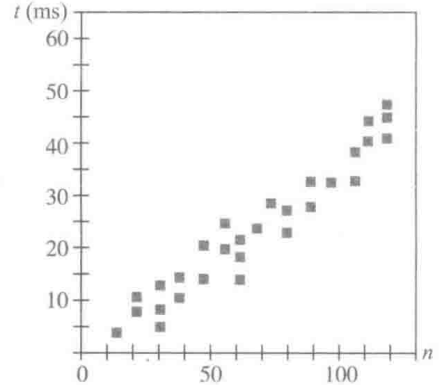


图 1-1 一个算法运行时间的实验研究结果。坐标  $(n, t)$  表示当输入规模为  $n$  时算法的运行时间是  $t$  毫秒(ms)

### 对通用分析方法的需求

通过实验取得算法的运行时间是有价值的，但有一定的局限性：

- 实验只能在一组有限的测试输入上进行，并且必须确保这些输入都是有代表性的。
- 比较两种算法的效率很困难，除非实验是在相同的硬件和软件环境中进行的。
- 必须首先实现算法，然后才能用实验方法度量算法的运行时间。

因此，虽然实验对算法分析有重要作用，但是，单靠这种方法是不够的。因此，除了实验方法，需要一种分析的方法，该方法能够：

- 考虑所有可能的输入。
- 允许用一种独立于硬件和软件环境的方式评估任何两种算法的相对效率。
- 通过算法的高阶描述完成运行时间评估，无需实现算法或者进行实验。

这种方法的目标是用一个输入规模  $n$  的函数  $f(n)$  刻画算法的运行时间。典型的函数包括  $n$  和  $n^2$ 。例如，用“算法的运行时间正比于  $n$ ”这样的语句表达，如果进行实验度量，那么对任何规模为  $n$  的输入，其运行时间不超过  $cn$ ，其中  $c$  是依赖于实验的硬件和软件环境的常数。给定  $A$  和  $B$  两个算法，如果  $A$  的运行时间正比于  $n$ ， $B$  的运行时间正比于  $n^2$ ，那么  $A$  的效率高于  $B$ ，因为函数  $n$  的增长率小于函数  $n^2$  的增长率。

现在可以“撸起袖子”给出算法分析的方法了。这种方法包括以下几个部分：

- 一种描述算法的语言。
- 一个执行算法的计算模型。
- 一个度量算法运行时间的尺度。
- 一个评价算法运行时间的方法，包括评价递归算法的方法。

本节将对以上部分做详细描述。

### 1.1.1 伪代码

程序员通常需要用便于人类阅读的方法描述算法。这样的描述不能是电脑程序，但是比普