HZ BOOKS
华章教育

计 算 机 科 学 丛 书

《算法导论》第一作者托马斯 H. 科尔曼面向大众读者的算法著作

理解计算机科学中关键算法的简明读本，帮助您开启算法之门

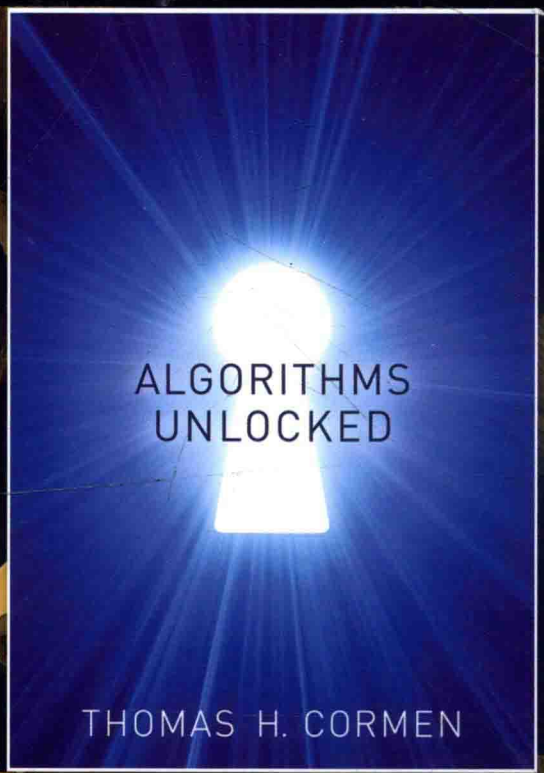
# 算法基础

## 打开算法之门

[美] 托马斯 H. 科尔曼（Thomas H. Cormen）著

王宏志 译

Algorithms Unlocked




机械工业出版社
China Machine Press

计 [redacted] 学 丛 书

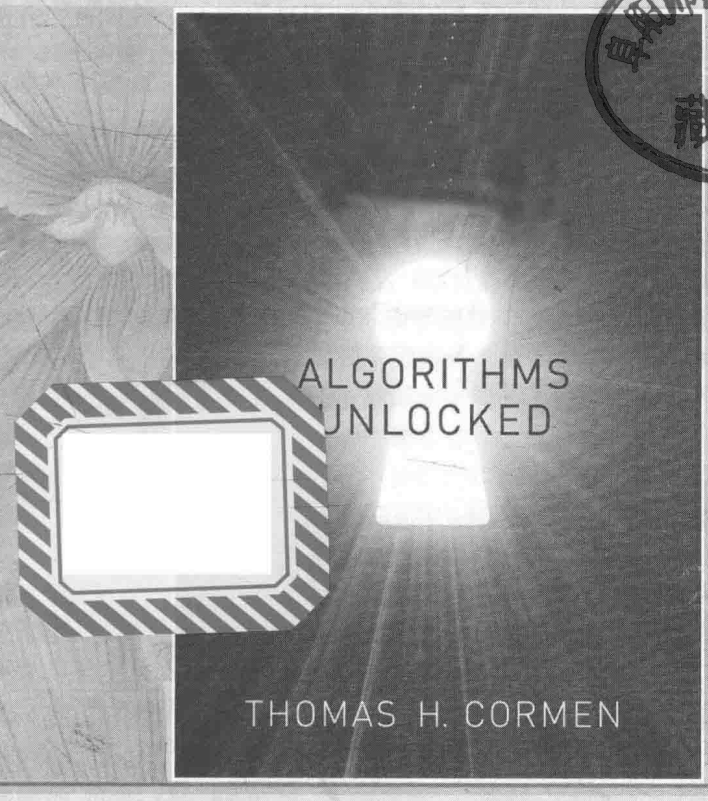
# 算法基础

## 打开算法之门

[美] 托马斯 H. 科尔曼（Thomas H. Cormen） 著
王宏志 译

Algorithms Unlocked






机械工业出版社
China Machine Press

**图书在版编目（CIP）数据**

算法基础：打开算法之门 /（美）科尔曼（Cormen, T. H）著；王宏志译 . —北京：机械工业出版社，2015.11
（计算机科学丛书）
书名原文：Algorithms Unlocked

ISBN 978-7-111-52076-4

I. 算… II. ①科… ②王… III. 电子计算机－算法理论 IV. TP301.6

中国版本图书馆 CIP 数据核字（2015）第 266872 号

**本书版权登记号：图字：01-2013-7586**



本书介绍了什么是计算机算法，如何描述它们，以及如何来评估它们。这些计算机算法将提供：利用计算机搜索信息的简单方式；解决各种排序问题的方法；利用有向无环图和最短路径法来解决基本问题的方法（可用于建模公路网络，任务间的依赖及金融关系）；解决字符串（例如 DNA 结构）问题的方法；密码学背后的基本原理；数据压缩的基础知识；以及甚至一些没有人能够理解如何在计算机上用相当长的时间来解决的问题。

本书适合作为计算机专业本科生“算法设计与分析”课程的教材，也适合相关专业人员阅读。

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：关　敏　　责任校对：董纪丽
印　　刷：中国电影出版社印刷厂　　版　　次：2016 年 1 月第 1 版第 1 次印刷
开　　本：185mm×260mm　1/16　　印　　张：15.5
书　　号：ISBN 978-7-111-52076-4　　定　　价：59.00 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

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，Brain W. Kernighan，Dennis Ritchie，Jim Gray，Afred 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，我们的联系方法如下：

华章网站：www. hzbook. com
电子邮件：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

# 译者序

Algorithms Unlocked

算法设计与分析是计算机科学的核心内容之一，算法设计与分析的能力也成为计算机科学从业者最重要的基本功之一，因而“算法设计与分析”是计算机专业学生的重要专业课程。尽管算法设计与分析很重要，但这门课程对许多读者来说稍显“高冷”，主要表现为其内容抽象、覆盖范围广、需要的数学基础多，因而学习算法设计与分析仿若攀登一座费时费力的高山。

针对这种情况，计算机领域的大牛 Thomas Cormen 出手了。他撰写了此书作为面向算法设计与分析初学者的入门书籍。本书有着如下几个鲜明的特点。

第一，本书仅仅使用了有限的数学知识。对于很多算法初学者来说，阻碍其学习的很重要的一个绊脚石就是算法设计与分析中涉及的大量数学知识，覆盖了概率论、代数、数学分析、图论等多个方面，而本书不需要读者具备这些方面的深入知识，为算法初学者提供了一条入门的捷径。

第二，本书语言通俗生动，并且把算法和现实中的问题紧密连接，避免出现大量算法分析细节。一方面，让算法真正成为生活中的一种思维方式，让读者深入了解算法思想的实际用途；另一方面，对于很多应用背后的算法知识，让读者在“知其然”的同时“知其所以然”。

第三，本书覆盖范围广，在 200 多页的篇幅中覆盖了图论算法、字符串算法、密码算法、数据压缩算法，甚至 NP-完全问题和不可判定问题，使读者可在最短的时间内掌握多种应用中不同的算法。

特别值得一提的是，本书的作者 Thomas Cormen 也是算法设计与分析方面的经典教材《算法导论》的作者之一，译者有幸参与了该书的翻译工作。《算法导论》是一本内容深入的算法设计与分析方面的大部头教材，而本书则可以看作是《算法导论》的一个薄薄的入门版本，通过阅读本书，读者可以用最短的时间轻松地窥见算法设计与分析的门径，奠定学习“算法设计与分析”课程的基础。

在本书英文版出版以后，译者应机械工业出版社的邀请开始了本书的翻译工作。由于水平有限且时间紧张，译文中一定存在许多不足，在此敬请各位同行、专家、学者和广大读者批评指正，欢迎大家将发现的错误或提出的意见与建议发送到邮箱 wangzh@hit.edu.cn，以改进本书的译本。

最后，我要感谢哈尔滨工业大学的孔欣欣同学在翻译过程中进行的辅助翻译工作。在完成译稿之后，我的爱人黎玲利博士阅读全文并提出了很多有益的意见，在此也表示感谢。同时感谢机械工业出版社的姚蕾编辑和朱劼编辑，由于她们的信任和支持，本书的翻译工作才得以顺利进行。

王宏志

# 前　言

Algorithms Unlocked

计算机是如何解决问题的呢？小小的 GPS 是如何只在几秒钟内就从无数条可能路径中找出到达目的地的最快捷路径的呢？在网上购物时，又如何防止他人窃取你的信用卡账号呢？解决这些问题，以及大量其他问题的答案均是**算法**。我写本书的目的就是为你打开算法之门，解开算法之谜。

我是《算法导论》的合著者之一。《算法导论》是一本特别好的书（当然，这是我个人的主观评价），但是它确实相当专业。

本书并不是《算法导论》，甚至不能被称为一本教材。它既没有对计算机算法领域进行广度或深度的研究，也没有遵照惯例来讲述设计计算机算法的方法，甚至连一道需要读者自己求解的难题或者练习题也没有。

那么，这是一本什么样的书呢？如果你符合如下条件，那么就可以开始阅读之旅了：

- 你对计算机如何解决问题感兴趣；
- 你想知道如何评估这些解决方案的质量；
- 你想了解计算方面的问题和这些问题的解决方案是如何与非计算机世界关联起来的；
- 你能处理一点数学运算；
- 你不需要编写过计算机程序（当然，编写过程序更好）。

一些计算机算法方面的书籍是讲述理论概念的，并涉及非常少的技术细节；一些书籍关注的全是技术细节；而另外一些书籍是介于这两者之间的。每类图书都有自己的定位，我将本书定位于介于两者之间。诚然，本书涉及了一些数学知识，并且部分地方阐述得非常仔细，但是我已经竭力避免深入阐述细节（或许除了本书的末尾部分，我无法克制住自己，阐述了一些细节知识）。

我认为本书有点像开胃菜。设想你去了一家意大利餐厅，点了一份开胃菜，直到吃完开胃菜，你才会决定是否点其余食物。开胃菜到

了，你就开始用餐了。或许你不喜欢吃开胃菜，并且决定不点其他菜了。可能你喜欢吃开胃菜，但是吃完它，你就感觉饱了，因此不需要点其他菜了。或者也有可能你喜欢吃开胃菜，但你并没有吃饱，此时你便开始期待其他菜了。将本书看作开胃菜，我希望能够产生后两种结果之一：或者读完了本书，你就很满足，感觉没有必要再深入探究算法世界了；或者你非常喜欢从本书中所学到的知识，以至于你想要学习更多算法方面的内容。每一章最后一节的标题为“拓展阅读”，其中会介绍更多关于该章主题的更为深入的书籍和文章。

## 你将从本书中学到什么

我无法断定你将从本书中学到什么。如下是我希望你能从本书中学到的：

- 什么是计算机算法，能够采用一种方式来描述计算机算法，以及如何评估算法。
- 在计算机中查找信息的简单方式。
- 在计算机中重排信息以使其以一种预定顺序排列的方法。（我们称这一任务为“排序”。）
- 如何解决那些能在计算机中以一种称为“图”的数学结构建模的基本难题。在许多应用中，利用图建模的领域包括：道路网（哪些十字路口到哪些十字路口有直接相连的道路，这些道路有多长），任务间的依赖关系（哪个任务必须在其他任务之前完成），金融关系（世界各国货币间的汇率是多少），或者是人与人之间的联系（谁认识谁？谁讨厌谁？哪个演员和哪个演员出现在同一个电影中等）。
- 如何解决关于文本字符串的问题。其中一些问题在某些领域有所应用，例如生物学领域，其中字符表示基本的分子，字符串表示 DNA 结构。
- 密码学背后的基本原理。即使你自己从来没有加密过一条信息，你的计算机很可能已经对它执行加密操作了（例如当你在网上购物时）。
- 数据压缩的基本概念，这远远超过了“f u cn rd ths u cn gt a gd jb n gd pay”背后的压缩原理。
- 计算机在任意合理的时间内都难以解决的一些问题，或者至少还没有人想出如何解决的问题。

## 为了理解本书中的内容，你需要事先了解什么

正如我之前所说的，本书中涉及部分数学知识。如果你害怕数学，那么你可以尝试着跳过它，或者你也可以尝试着阅读涉及更少专业技术知识的书籍。但是我已经尽力做到令数学部分变得容易理解了。

我假定你从来没有写过，甚至从来没有读过一个计算机程序。如果你能看懂提纲的内容，就应该能够理解我是如何表达每一步算法，以及如何将这些步骤合并在一起组合成一个完整的算法的。如果你听到过如下笑话，那么你已经是在通往算法世界了：

> 你听说过被困在淋浴中的计算机专家吗？当时他在按照洗发瓶上的指示洗头发。指示说明是“打洗发露。冲洗。重复。”

本书使用了一种自由的写作风格，希望这种比较个性的方法能使本书的内容看起来更容易理解。有些章节依赖于前面章节的内容，但是这种依赖程度很轻。有些章节开始时不涉及专业技术知识，但是会逐步讲述专业技术知识。即使你感觉某一章太难了，这也不会影响下一章内容的学习。你也很可能会从下一章的开始部分受益。

## 报告错误

如果你在本书中发现了错误，请通过发送邮件至 algorithms-unlocked@mit. edu 来告知我。

## 致谢

本书中的许多内容都参考了《算法导论》的内容，因此多亏了《算法导论》的合著者——Charles Leiserson、Ron Rivest 以及 Cliff Stein 的帮助。你将发现我自始至终都在频繁地提到(插入)《算法导论》的内容，我们 4 个作者所写的《算法导论》早已众所周知了。在写本书时，我意识到我是多么想念和 Charles、Ron 及 Cliff 的合作。同时我仍然感谢在《算法导论》的前言部分所感谢的那些人。

同时，我也参考了在达特茅斯学院教书时所讲述的课程内容，尤其是计算机科学课程 1、5 和 25。感谢我的学生，通过他们精辟的见

解，我看出了当前这种教学方法很好；通过他们无情的沉默眼神，我看出了当前这种教学方式不理想。

本书是在 Ada Brunstein 的建议下撰写的。Ada Brunstein 是 MIT 出版社负责《算法导论》第 3 版的编辑。Ada 现在已经离开 MIT 出版社了，Jim DeWolf 接替了她的位置。刚开始时，本书被指定为 MIT 出版社的“基础知识”丛书的一部分，但是 MIT 出版社认为对于“基础知识”丛书而言，本书过于专业了。（想象一下——我写了一本对于 MIT 而言过于专业的书籍！）Jim 巧妙灵活地处理了这件事，允许我以自己的方式来写这本书，而不是按照 MIT 出版社初期的规定。同时，我还要感谢 MIT 出版社 Ellen Faran 和 Gita Devi Manaktala 的支持。

Julie Sussman，P. P. A.，是《算法导论》第 2 版和第 3 版的文字编辑，本书还是由她担任文字编辑，对此我感到非常兴奋。她是最好的、最专业的文字编辑。她让我放下所有顾虑。为了证明她的优秀，请看 Julie 关于我的第 5 章初稿所回复的一份电子邮件：

> 亲爱的 Cormen 先生，
>
> 当局逮捕了一个逃走的章节，这个章节被发现隐藏在你的书中。我们无法确定它是从哪本书中逃离的，但是我们无法想象这几个月中在你都不知晓的情况下，它是如何一直寄宿在你的书中的，因此我们别无选择，只能询问你。我们希望你能承担起修改这一章的任务，给它一个机会，让它成为书中的一个有用的公民。来自一个负责逮捕的军官的报告，
>
> Julie Sussman，附上。

你可能很好奇“P. P. A.”代表什么，事实上前两个字母代表“Professional Pain”，很可能你已经猜想到了“A”代表什么，但是我想要指出 Julie 的确以这个头衔自豪。因此非常非常感谢 Julie！

我并不是一个密码破译者，关于密码学原理的那一章极大地归功于 Ron Rivest、Sean Smith、Rachel Miller 以及 Huijia Rachel Lin 的帮助。那一章中有一个关于棒球手势的脚注说明，这要感谢达特茅斯学院的棒球教练 Bob Whalen，是他耐心地向我解释了棒球手势系统中的一些手势。Ilana Arbisser 核实了计算生物学家对齐 DNA 序列的方式与第 7 章所介绍的方式一致。Jim DeWolf 和我仔细思考了本书的书

名，但是“Algorithms Unlocked”这一书名最终是由达特茅斯学院的一个本科生 Chander Ramesh 提出的。

达特茅斯学院计算机科学系是一个很好的工作去向。我的同事个个才华横溢，我们的专职人员也都是首屈一指的。如果你希望编写一个本科生或者研究生级别的计算机科学程序，或者如果你在寻找一个计算机科学专业的教授职位，建议你申请达特茅斯学院。

最后，感谢我的妻子 Nicole Cormen、我的父母 Renee 和 Perry Cormen、我的姊妹 Jane Maslin 以及 Nicole 的父母 Colette 和 Paul Sage，感谢他们对我的爱和支持。我的父亲确信在 1.1 节中的图形是 5，而不是 S。

Tom Cormen

于新罕布什尔州汉诺威

# 目 录

Algorithms Unlocked

第1章
Algorithms Unlocked

# 什么是算法以及为什么应该关注算法

让我们从我经常被问到的一个问题开始："什么是算法?"[⊖]

一个常见的回答是，"完成一个任务所需的一系列步骤。"在日常生活中经常会碰到算法。刷牙的时候会执行一个算法：打开牙膏盖，拿出牙刷，持续执行挤牙膏操作直到足量的牙膏涂在你的牙刷上，盖上牙膏盖，将牙刷放到嘴的1/4处，上下移动牙刷N秒，等等。如果你必须乘通勤车去工作，乘通勤车也是一个算法。诸如此类。

但是本书是关于运行在计算机上的算法的，或者更概括地来讲，是关于运行在计算设备上的算法的。正如你日常所运行的算法会影响你每天的生活一样，在计算机上运行的算法也会影响你的生活。你使用过GPS来寻找旅行路线吗？它运行一种称为"最短路径"的算法以寻求路线。你在网上购买商品吗？那么你会使用（应该正在使用）一个运行加密算法的安全网站。当你在网上购买商品时，它们是由一个私营快递公司发货的吗？它使用算法将包裹分配给不同的卡车，然后确定每个司机发件的顺序。算法运行在各种设备上——在你的笔记本上，服务器上，智能手机上，嵌入式系统上（例如你的车中，你的微波炉中，或者气候控制系统中）——无处不在！

运行在计算机上的算法和你在日常生活中执行的算法有什么区别呢？当粗略地描述一个算法时，你可能能够容忍它的非精确性，但是计算机不能。例如，如果你开车上班，你的drive-to-work算法可能会说"如果交通不畅，可以选择其他路线"。虽然你可能知道"交通不畅"是什么意思，但是计算机不知道。

因此，一个计算机算法是完成一个任务所需的一系列步骤，且这些步骤需要足够精确地描述，以使得计算机能够运行它。如果你已经用Java、C、C++ 、Python、Fortran、Matlab或者类似的编程语言编写过哪怕一丁点的计算机程序，那么你会对精确度标准的含义有一些



⊖ 或者，正如一个我曾经一起打曲棍球的同伴问我的，"什么是nalgorithm?"

概念。如果你从来没有编写过计算机程序，那么当你看了本书中如何描述算法后，可能你会对精确度有一点概念。

我们思考下一个问题："我们想从一个计算机算法中获取什么?"

计算机算法解决计算问题。我们希望从一个计算机算法中获取两个结果：给定一个问题输入，它应该总能够产生该问题的正确输出结果，并且在运行该算法时，应该能够有效地利用计算资源。让我们依次看看这两个必要条件。

## 1.1 正确性

产生问题的一个正确解决方案意味着什么呢? 我们通常会精确地定义一个正确的解决方案涉及的内容。例如，为了寻找出最佳旅行路线，你的GPS会产生一个正确的解决方案。该方案可能是从你所在位置到目的地的所有可能路线中最快的路线，也可能是具有最短距离的路线，或者是能使你最快到达目的地同时也能免交过路费的路线。当然，你的GPS确定路线时所使用的信息可能不完全匹配实际情况。除非你的GPS能够获取实时路况信息，否则它可能假定穿过一条道路的时间等于道路的长度除以道路的限定时速。然而，如果道路拥挤，当你在寻找一条最快路线时，GPS可能不能给你提供好的建议。然而，即使算法的输入是不正确的，我们仍然可以说GPS所提供的路线选择算法是正确的，即对于给定的输入，该路线选择算法输出最快的路线。

然而，对于某些问题，可能难以判定甚至不可能判定一个算法是否产生了正确的输出。以光学字符识别为例。这个11×6像素的图像表示5还是S呢?

一些人可能会说它是5，而其他人可能说它是S，因此我们也不能判定计算机的输出是否正确。在本书中，我们将只关注有确定解的计算机算法。



然而，有些时候，我们可以接受可能会产生错误解的算法，只要产生错误解的频率可以被控制。加密算法就是一个范例。最常用的

RSA 加密系统依赖于确定大数是否为素数，这里的大数指相当大的数，如数百位那么长。如果你曾经写过一个计算机程序，你可能能够写出一个判定数 $n$ 是否是素数的程序。它将测试从 2 到 $n-1$ 的所有候选除数，如果这些候选除数中有一个除数确实能被 $n$ 整除，那么 $n$ 是合数。如果 2 和 $n-1$ 之间的任何数均不能被 $n$ 整除，那么 $n$ 是素数。但是如果 $n$ 是数百位长的数，那就会产生大量的候选除数，即使是一个运行相当快的计算机进行相应的检查操作也会超过合理的运行时间。当然，可以进行一些优化操作，例如当检测出 2 不是 $n$ 的除数后，在候选除数中可以去除所有的偶数，或者循环到候选除数等于 $\sqrt{n}$ 时终止（由于若 $d>\sqrt{n}$，且 $n \bmod d=0$，那么 $\frac{n}{d}<\sqrt{n}$，$n \bmod (n/d)=0$；这说明若 $n$ 能整除一个大于 $\sqrt{n}$ 的数，则 $n$ 也必定能够整除一个小于 $\sqrt{n}$ 的数）。如果 $n$ 是一个数百位的数，则尽管 $\sqrt{n}$ 的位数是数百位的一半，但是它仍然是一个非常大的数。好消息是，我们知道一个可以高效测试一个数是否是素数的算法。坏消息是，该算法可能会得出错误的结论。特别是，当该算法得出 $n$ 是合数时，则 $n$ 一定是一个合数，但是若该算法得出 $n$ 是一个素数，$n$ 实际上也可能是一个合数。但是坏消息也不全不好，我们可以对其加以控制，使得错误率降到足够低，例如每执行 $2^{50}$ 次才会出现一次错误。那是相当罕见的了——大约每千万亿次才出现一次错误——在 RSA 中应用这个方法来判定一个数是否是素数对于大多数人而言是安全的。

对于另一类算法——近似算法，正确性也是一个需要着重考量的问题。近似算法适合于优化问题，即根据一些量化测度来寻找最优解的问题。例如 GPS 中寻找最快路径问题就是一个优化问题，它的量化测度是旅程中花费的时间。对某些问题，我们找不到任何可以在合理的时间内求解出最优解的算法，但是我们能够找到一个近似算法，它可以在合理的时间内求解得出一个近似的最优解。“近似最优”就是近似算法输出的解的量化测度值介于最优解的量化测度值的一个已知因子之内。只要指定了目标因子，我们就可以说一个近似算法的正确解是任意一个量化测度值在最优解目标因子之内的解决方案。



## 1.2 资源利用

什么样的算法才能称为高效使用计算资源的算法呢？我们在讨论