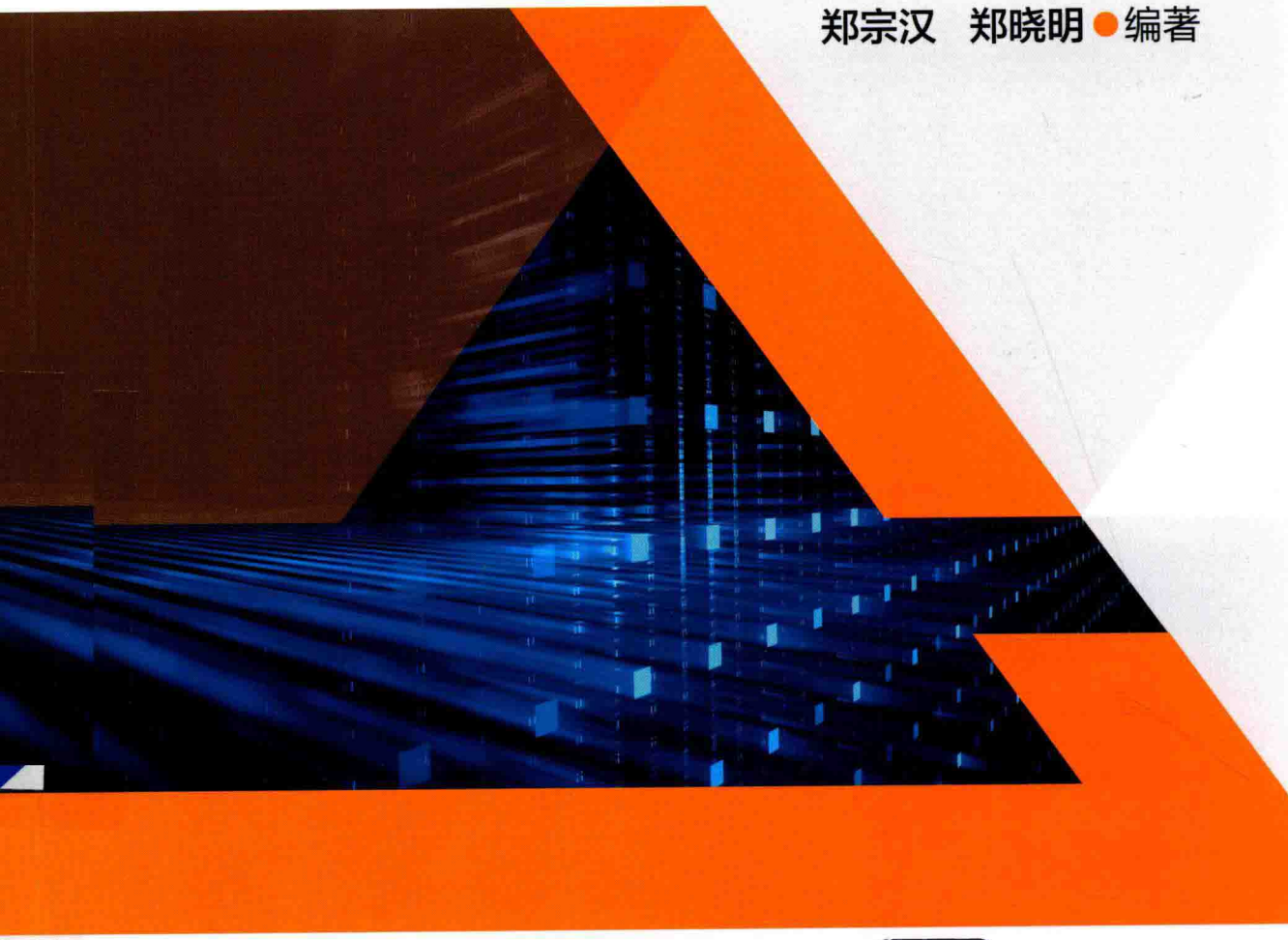
应用型
特色教材

算法设计与分析

（第3版）

郑宗汉　郑晓明●编著



清华大学出版社

算法设计与分析

（第3版）

郑宗汉　郑晓明　编著

清华大学出版社
北京

内 容 简 介

本书系统地介绍了算法设计与分析的概念和方法，共4篇内容。第1篇介绍算法设计与分析的基本概念，结合穷举法、排序问题及其他一些算法，对算法的时间复杂性的概念及复杂性的分析方法作了较为详细的叙述；第2篇以算法设计技术为纲，从合并排序、堆排序、离散集合的union和find操作开始，进而介绍递归技术、分治法、贪婪法、动态规划、回溯法、分支与限界法和随机算法等算法设计技术及其复杂性分析；第3篇介绍计算机应用领域里的一些算法，如图和网络流，以及计算几何中的一些问题；第4篇介绍算法设计与分析中的一些理论问题，如*NP*完全问题、计算复杂性问题、下界理论问题，最后介绍近似算法及其性能分析。

本书内容选材适当、编排合理、由浅入深、循序渐进、互相衔接、逐步展开，并附有大量实例，既注重算法的思想方法、推导过程和正确性的证明技术，也注重算法所涉及的数据结构、算法的具体实现和算法的工作过程。

本书可作为高等院校计算机专业本科生和研究生的教材，也可作为计算机科学与应用的科学技术人员的参考资料。

图书在版编目（CIP）数据

算法设计与分析 / 郑宗汉，郑晓明编著. —3版. —北京： 清华大学出版社，2017
ISBN 978-7-302-45720-6

Ⅰ. ①算… Ⅱ. ①郑… ②郑… Ⅲ. ①电子计算机－算法设计 ②电子计算机－算法分析
Ⅳ. ①TP301.6

中国版本图书馆CIP数据核字(2016)第288788号

责任编辑：苏明芳
封面设计：刘　超
版式设计：李会影
责任校对：王　颖
责任印制：刘海龙

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦A座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：保定市中画美凯印刷有限公司
经　　销：全国新华书店
开　　本：185mm×260 mm　　**印　　张**：27.75　　**字　　数**：654千字
版　　次：2005年6月第1版　2017年10月第3版　　**印　　次**：2017年10月第1次印刷
印　　数：1～4000
定　　价：59.80元

产品编号：068461-01

前　言

计算机系统中的任何软件，都是按特定的算法来予以实现的。算法性能的好坏，直接决定了所实现软件性能的优劣。如何判定一个算法的性能？用什么方法来设计算法？所设计的算法需要多少运行时间、多少存储空间？在实现一个软件时，这些都是必须予以解决的问题。计算机中的操作系统、语言编译系统、数据库管理系统，以及各种各样的计算机应用系统中的软件，都离不开用具体的算法来实现。因此，算法设计与分析是计算机科学与技术的一个核心问题，也是大学计算机专业本科生及研究生必修的一门重要的专业基础课程。通过算法设计与分析这门课程的学习，读者能够掌握算法设计与分析的方法，并利用这些方法去解决在计算机科学与技术中所遇到的各种问题，设计计算机系统的各种软件中所可能遇到的算法，并对所设计的算法做出科学的评价。因此，算法设计与分析，不仅对计算机专业的科学技术人员，而且对使用计算机的其他专业技术人员，都是非常重要的。

本书内容选材适当、编排合理、由浅入深、循序渐进、互相衔接、逐步展开，在编写过程中，尽可能遵循下面几个原则。

（1）面对学生，尽可能用通俗的语言来表达深奥的问题。因此，公式推导尽可能详尽，因为这样才不会为难学生；所用到的专门术语或知识来龙去脉尽可能介绍清楚，因为这样，学生才不会感到错愕和不知所措；对问题的叙述，尽可能开门见山，使学生能较快地、容易地接触到问题的实质。

（2）实现算法的思想方法和推导过程尽可能详细和易于理解，因为这样，学生才能深刻地理解和掌握算法的工作原理。

（3）对算法的某些理论基础和定理的证明给予足够的重视，定义的叙述尽可能严谨，方法推导、定理证明的逻辑尽可能严密，因为这可以培养学生良好的逻辑思维能力和严谨规范的科学方法，而设计并实现一个算法，具有良好的逻辑思维能力和严谨规范的科学方法是很重要的。

（4）算法具体实现的描述、所涉及的数据结构和变量，都做出较为详尽的说明。因为有些学生尽管知道了思想方法，也了解了实现步骤，但具体实现起来，有时却觉得束手无策。对算法的具体实现和所用到的数据结构的较为详细的描述，可以培养学生的具体实践能力，使学生学会如何使用所学过的知识来设计和实现某些算法。

（5）算法的工作过程，尽可能详细说明。对工作过程中比较难于理解的某些算法，则通过实例，从头到尾地模拟算法的运行。因为这是学习、理解算法的关键，没有用实例来模拟算法的运行，学生学完了以后，对该算法也只能是懵懂的，不知其所以然。

（6）无论是算法的基本概念、算法复杂性的分析方法，还是算法的实现步骤，都尽可能提供大量实例加以解释说明。

本书先以最简单的穷举法为例，说明算法设计技术及算法分析的重要性；接着以排序问题中的一些基本算法和其他一些算法为例，说明算法复杂性的一般分析方法；然后以算法设计技术为纲，按照实现算法的思想方法、实现步骤、所涉及的数据结构、算法的具体描述及复杂性分析等几个方面，逐个介绍各种算法设计技术及其分析方法。

全书分为4篇。第1篇包括第1章和第2章，介绍算法设计与分析的基本概念。第1章介绍算法的定义及算法的时间复杂性的基本概念；第2章介绍算法时间复杂性的分析方法，并简单介绍与算法分析有关的最基本的数学工具。第2篇包括第3～9章，介绍算法设计的基本技术。第3章继续介绍排序问题和离散集合的操作，进一步对算法分析进行了阐述，并为下面各章中所涉及的问题作了技术上的准备；第4章介绍递归技术及分治方法，从理论上分析了分治算法的效率；第5章介绍贪婪法的设计方法及其正确性的证明；第6章介绍动态规划法的设计技术；第7章介绍回溯法的设计技术；第8章在回溯法的基础上介绍分支与限界方法的应用及其分析；第9章介绍3种类型的随机算法及其性能分析。第3篇包括第10～11章，涉及计算机应用领域里的一些算法。第10章介绍图和网络流的一些问题；第11章介绍计算几何中的一些问题。第4篇包括第12～15章，介绍算法设计与分析中的一些理论问题。第12章介绍NP完全问题；第13章介绍计算复杂性问题；第14章介绍下界理论问题；第15章介绍近似算法及其性能分析。

本书对第2版做了一些修订。感谢许存权、钟志芳、纪文远、苏明芳、邓艳诸位编辑为本书的出版付出的大量工作和努力，在此表示诚挚的谢意。

由于水平有限，书中难免存在不当之处，敬请读者指正。

编　者

目　录

第 1 篇　算法设计与分析的基本概念

第2篇 算法设计的基本技术

第3篇 计算机应用领域的一些算法

第1篇

算法设计与分析的基本概念

第 1 章　算法的基本概念

计算机系统中的任何软件，都是由大大小小的各种程序模块组成的，它们按照特定的算法来实现，算法的好坏直接决定了所实现软件性能的优劣。用什么方法来设计算法，所设计算法需要什么样的资源，需要多少运行时间、多少存储空间，如何判定一个算法的好坏……在实现一个软件时，这些都是必须予以解决的。计算机系统中的操作系统、语言编译系统、数据库管理系统以及各种各样的计算机应用系统中的软件，都必须用一个个的具体算法来实现。因此，算法设计与分析是计算机科学与技术的一个核心问题。

1.1　引　　言

“算法”这一术语是从英文 Algorithm 一词翻译而来的，但直到 1957 年，西方著名的《韦伯斯特新世界词典》也未将这一单词收录其中。据西方数学史家的考证，古代阿拉伯的一位学者写了一部名著——*Kitāb al-jabr Wa'lmuqābala*（《复原和化简的规则》），作者的署名是 Abū 'Abd Allāh Muhammad ibn Mūsā al-Khwārizmī。从字面上看，其含义是“穆罕默德（Muhammad）的父亲，摩西（Moses）的儿子，Khwārizm 地方的人”。后来，这部著作流传到了西方，结果从作品名称中的 al-jabr 派生出 Algebra（代数）一词；从作者署名中的 al-Khwārizmī 派生出 Algorism（算术）一词，最后，又从 Algorism 衍生出 Algorithm。随着时间的推移，Algorithm 这个词的含义已变得面目全非了，成了本书要讨论的内容——算法。

1.1.1　算法的定义和特征

欧几里得曾在他的著作中描述过求两个数的最大公因子的过程。20 世纪 50 年代，欧几里得所描述的这个过程被称为 Euclides Algorithm for gcd，国内将其翻译为“求最大公因子的欧几里得算法”，Algorithm（算法）这一术语在学术上具有了现在的含义。下面通过一个例子来认识一下该算法。

算法 1.1　欧几里得算法
输入：正整数 m,n
输出：m,n 的最大公因子

```
1. unsigned euclid_gcd(unsigned m,unsigned n)
2. {
3.     unsigned     r;
```

```
4.    while(n!=0) {
5.        r = m % n;
6.        m = n;
7.        n = r;
8.    }
9.    return m;
10. }
```

在此用一种类 C 语言来叙述最大公因子的求解过程。今后，在描述其他算法时，还可能结合一些自然语言的描述，以代替某些烦琐的具体细节，从而更好地说明算法的整体框架。同时，为了简明、直观地访问二维数组元素，假定在函数调用时，二维数组可以直接作为参数传递，在函数中可以动态地分配数组，等等。读者可以容易地把这种类 C 语言程序转换为 C 语言程序。

这个算法由一个循环组成，第 4 行判断循环体的执行条件，若 n 为 0，结束循环的执行，转到第 9 行，把 m 作为结果还回，算法结束；若 n 非 0，就执行第 5~7 行的循环体：第 5 行把 m 除以 n 的余数赋予 r，第 6 行把 n 的值赋予 m，第 7 行把 r 的值赋予 n，然后转到第 4 行继续判断是否执行循环体。按照上面这组规则，给定任意两个正整数，总能返回它们的最大公因子。读者可以自行证明这个算法的正确性。

根据上面这个例子，可以给算法如下的定义：

定义 1.1　算法是解某一特定问题的一组有穷规则的集合。

算法设计的先驱者唐纳德 •E.克努特（Donald E.Knuth）对算法的特征做了如下的描述：

（1）有限性。算法在执行有限步之后必须终止。算法 1.1 中，对输入的任意正整数 m、n，在 m 除以 n 的余数赋予 r 之后，再通过 r 赋予 n，从而使 n 值变小。如此往复进行，最终或者使 r 为 0，或者使 n 递减为 1。这两种情况最终都使 r=0，而使算法在有限步后终止。

（2）确定性。算法的每一个步骤都有精确的定义，要执行的每一个动作都是清晰的、无歧义的。例如，在算法 1.1 的第 5 行中，如果 m、n 是无理数，那么 m 除以 n 的余数是什么，就没有一个明确的界定。确定性的准则意味着必须确保该算法在执行第 5 行时，m 和 n 的值都是正整数。算法 1.1 中规定了 m、n 都是正整数，从而保证了后续各个步骤中都能确定地执行。

（3）输入。一个算法有 0 个或多个输入，它是由外部提供的，作为算法开始执行前的初始值或初始状态。算法的输入是从特定的对象集合中抽取的。算法 1.1 中的两个输入 m、n，就是从正整数集合中抽取的。

（4）输出。一个算法有一个或多个输出，这些输出与输入有着特定的关系，实际上是输入的某种函数。不同取值的输入，产生不同结果的输出。算法 1.1 中的输出是输入 m、n 的最大公约数。

（5）能行性。算法的能行性指的是算法中有待实现的运算都是基本的运算，原则上可以由人们用纸和笔在有限的时间里精确地完成。算法 1.1 用一个正整数来除另一个正整数、判断一个整数是否为 0 以及整数赋值等，这些运算都是能行的。因为整数可以用有限的方

式表示，而且至少存在一种方法来完成一个整数除以另一个整数的运算。如果所涉及的数值必须由展开成无穷小数的实数来精确地完成，则这些运算就不是能行的了。

必须注意到，在实际应用中，有限性的限制是不够的。一个实用的算法，不仅要求运行的步骤有限，同时也要求运行这些步骤所花费的时间是人们可以接受的。如果一个算法需要执行数以百亿亿计的运算步骤，从理论上说，它是有限的，最终可以结束。但是，这样的算法，除非拿到超级计算机去运行外，否则，以当代一般的计算机每秒数亿次的运算速度，也必须运行数百年以上时间。对于期望以一般的计算机作为平台来运行这样的算法，这是人们所无法接受的，因而是不实用的算法。同时也应注意到上述的确定性，指的是算法要执行的每一个动作都是确定的，并非指算法的执行结果是确定的。大多数算法不管在什么时候运行同一个实例，所得结果都一样，这种算法称为确定性算法；有些算法在不同的时间运行同一个实例，可能会得出不同的结果，这种算法称为不确定的算法或随机算法。

算法设计的整个过程，可以包含对问题需求的说明、数学模型的拟制、算法的详细设计、算法的正确性验证、算法的实现、算法分析、程序测试和文档资料的编制。本书所关心的是串行算法的设计与分析，其他相关的内容以及并行算法可参考专门的书籍。这里只在涉及有关内容时，才对相应的内容进行论述。

1.1.2 算法设计的例子——穷举法

例 1.1 百鸡问题。

公元5世纪末，我国古代数学家张丘建在他所撰写的《算经》中提出了这样一个问题："鸡翁一，值钱五；鸡母一，值钱三；鸡雏三，值钱一。百钱买百鸡，问鸡翁、母、雏各几何？"意思是公鸡每只5元，母鸡每只3元，小鸡3只1元，用100元钱买100只鸡，求公鸡、母鸡、小鸡的只数。

令 a 为公鸡只数，b 为母鸡只数，c 为小鸡只数。根据题意，可列出下面的约束方程：

$$a+b+c=100 \tag{1.1.1}$$

$$5a+3b+c/3=100 \tag{1.1.2}$$

$$c\%3=0 \tag{1.1.3}$$

其中，运算符"/"为整除运算，"%"为求模运算。式（1.1.3）表示 c 被3除，余数为0。

这类问题用解析法求解有困难，但可用穷举法来求解。所谓穷举法，就是从有限集合中，逐一列举集合的所有元素，对每个元素逐一判断和处理，从而找出问题的解。

上述百鸡问题中，a、b、c 的可能取值范围为0~100，对在此范围内的 a、b、c 的所有组合进行测试，凡是满足上述3个约束方程的组合，都是问题的解。如果把问题转化为用 n 元钱买 n 只鸡，n 为任意正整数，则式（1.1.1）、式（1.1.2）变成：

$$a+b+c=n \tag{1.1.4}$$

$$5a+3b+c/3=n \tag{1.1.5}$$

于是，可用下面的算法来实现。