HZ BOOKS
华章教育

计 算 机 科 学 丛 书

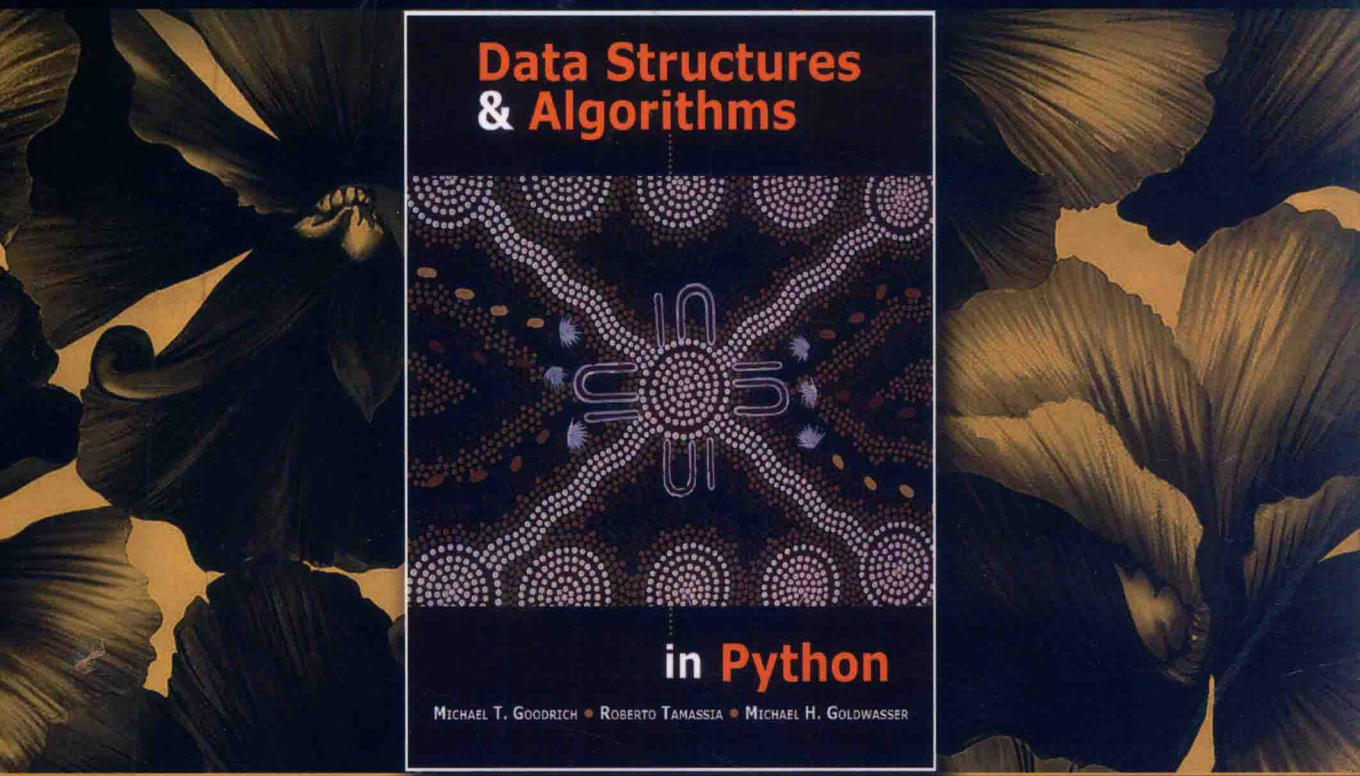
WILEY

数据结构与算法

Python语言实现

迈克尔·T. 古德里奇（Michael T. Goodrich）
[美] 罗伯托·塔马西亚（Roberto Tamassia） 著
迈克尔·H. 戈德瓦瑟（Michael H. Goldwasser）
张晓 赵晓南 等译

Data Structures and Algorithms in Python



机械工业出版社
China Machine Press

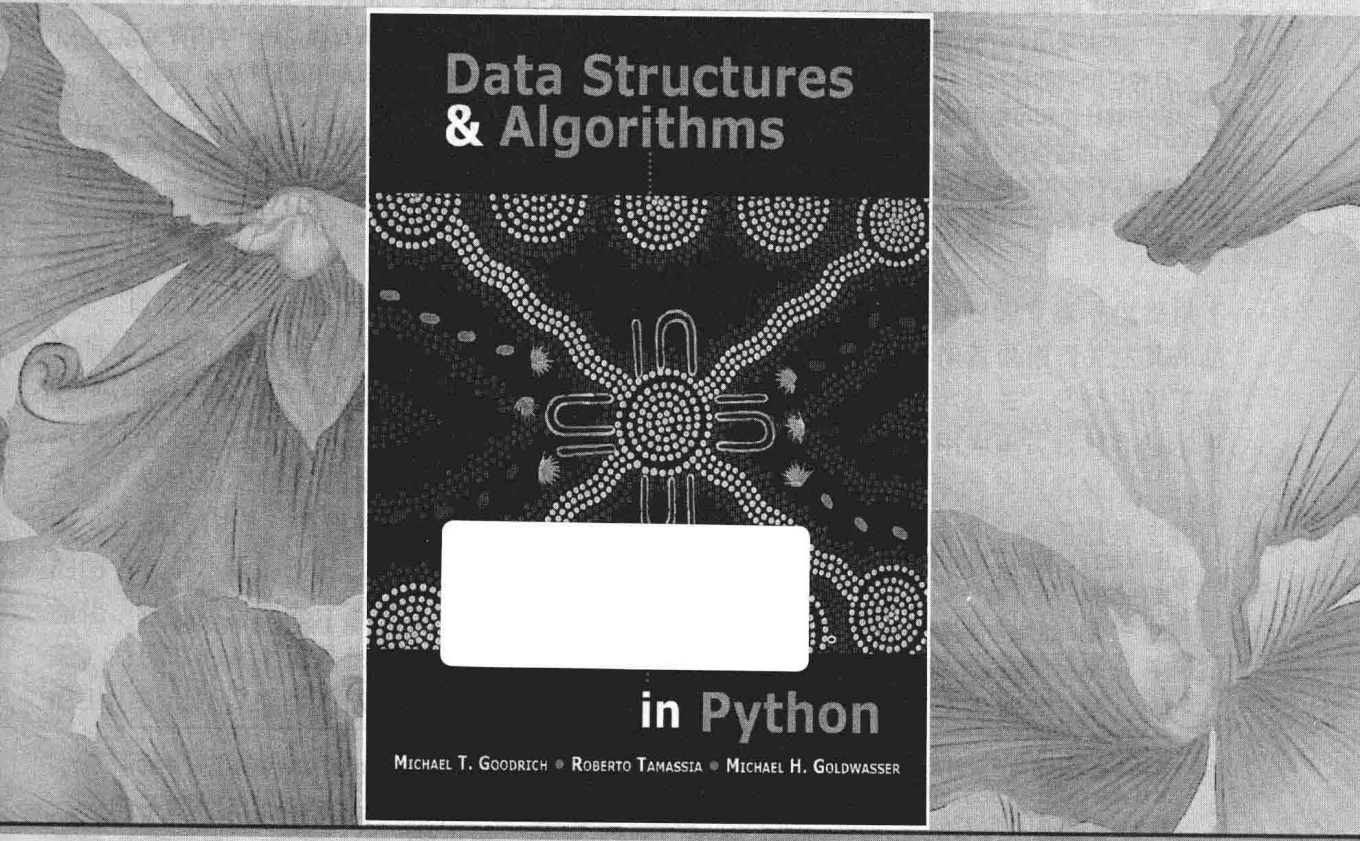
计 算 [redacted] 书

数据结构与算法

Python语言实现

迈克尔·T. 古德里奇（Michael T. Goodrich）
[美] 罗伯托·塔马西亚（Roberto Tamassia） 著
迈克尔·H. 戈德瓦瑟（Michael H. Goldwasser）
张晓 赵晓南 等译

Data Structures and Algorithms in Python



机械工业出版社
China Machine Press

图书在版编目（CIP）数据

数据结构与算法：Python 语言实现 /（美）迈克尔·T. 古德里奇（Michael T. Goodrich）等著；张晓等译 . —北京：机械工业出版社，2018.9
（计算机科学丛书）
书名原文：Data Structures and Algorithms in Python

ISBN 978-7-111-60660-4

I. 数… II. ①迈… ②张… III. ①数据结构 ②算法分析 ③软件工具 – 程序设计
IV. ① TP311.12 ② TP311.561

中国版本图书馆 CIP 数据核字（2018）第 183790 号

本书版权登记号：图字 01-2016-6251

本书采用 Python 语言讨论数据结构和算法，详细讲解其设计、分析与实现过程，是一本内容全面且特色鲜明的教材。书中将面向对象视角贯穿始终，充分利用 Python 语言优美而简洁的特点，强调代码的健壮性和可重用性，关注各种抽象数据类型以及不同算法实现策略的权衡。

本书适合作为高等院校初级数据结构或中级算法导论课程的教材，也适合相关工程技术人员阅读参考。

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：曲 熠　　责任校对：李秋荣
印　刷：北京市荣盛彩色印刷有限公司　　版　次：2018 年 9 月第 1 版第 1 次印刷
开　本：185mm × 260mm 1/16　　印　张：30.75
书　号：ISBN 978-7-111-60660-4　　定　价：109.00 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，Brian 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

Data Structures and Algorithms in Python

数据结构是计算机科学与技术专业的核心课程，是程序设计、编译原理、数据库等课程的基础。对于从事计算机应用尤其是软件开发的工程技术人员而言，掌握数据结构的相关知识和常用算法，对提高开发效率和编程质量都有着非常重要的作用。国内外有很多优秀的数据结构教材，而且有基于 C、C++、Java 等多种程序设计语言编写的版本，但是采用 Python 语言描述的并不多见。

Python 是一种面向对象的直译式计算机程序设计语言，语法简洁清晰，类库丰富强大。由于代码的平台无关性以及极简的编程思想，Python 近年来成为国内外各大科研院校和 IT 企业在教学活动、科学研究以及应用软件开发中频繁使用的程序设计语言。例如，卡内基·梅隆大学的编程基础课程、麻省理工学院的计算机科学及编程导论课程就使用 Python 语言讲授。我们西北工业大学也开设了 Python 程序设计的选修课，受到学生的热烈欢迎。在实践领域，NumPy、SciPy 等是利用 Python 语言开发的用于科学计算的工具包，著名的计算机视觉库 OpenCV、三维可视化库 VTK 等也是使用 Python 开发的。在 TIOBE 编程语言排行榜上，Python 排名第五，排在 Java、C、C++ 和 C# 之后。Python 语言也在大数据分析、网络爬虫、量化投资等新兴热点领域广泛使用。

对于计算机专业的学生和计算机应用行业的从业人员而言，从 Python 开始学习程序设计和数据结构入门门槛低，学习曲线平缓。国内已有大量介绍 Python 程序设计的书籍，但多局限在 Python 语法和特定软件包的使用方面。本书是难得的系统讲解如何使用 Python 语言设计并实现数据结构和基本算法的书籍。

本书作者 Goodrich 教授、Tamassia 教授等人先后撰写了《Data Structures and Algorithms in Java》和《Data Structures and Algorithms in C++》等书籍，对数据结构和常用算法的理解非常透彻。但本书并不是简单地将这些书籍中的代码描述部分替换成 Python 语言，而是充分利用 Python 语言的优势，以完整代码的方式实现了各种算法和数据结构。在此基础上，还大量介绍了 Python 语言内建的数据类型和一些常用基本库及接口的相关知识。

书中介绍的数据结构和算法包括完整的设计、分析和实现过程，非常适合作为初级数据结构课程的教材，同时也可以作为中级算法导论课程的教材，或作为计算机基础知识有限的工程技术人员的参考书。通过学习本书，读者能够更加灵活、高效地运用 Python 语言编写满足自己需求的程序。

非常荣幸能有机会翻译这样一本优秀的教材。对于书中的专业术语，我们尽量沿用了现有的习惯翻译。不过由于时间和水平所限，难免出现错误和不当之处，恳切希望广大读者不吝批评指正。

最后，感谢作者为我们呈现了一本优秀的教材；感谢出版社的信任，将这项有趣而又有意义的工作交给我们完成；还要感谢所有参与本书翻译和校对工作的教师和研究生，他们是赵晓南、王蕾、赵楠、陈震、卜海龙、柳春懿、孔兰昕、朱顺意等。

张晓

2018 年 4 月

于启真湖畔

高效数据结构的设计与分析，长期以来一直被认为是计算领域的一个重要主题，同时也是计算机科学与计算机工程本科教学中的核心课程。本书介绍数据结构和算法，包括其设计、分析和实现，可在初级数据结构或中级算法导论课程中使用。我们随后会更详细地讨论如何在这些课程中使用本书。

为了提高软件开发的健壮性和可重用性，我们在本书中采取一致的面向对象的视角。面向对象方法的一个主要思想是数据应该被封装，然后提供访问和修改它们的方法。我们不能简单地将数据看作字节和地址的集合，数据对象是*抽象数据类型*（Abstract Data Type，ADT）的实例，其中包括可在这种类型的数据对象上执行的操作方法的集合。我们强调的是对于一个特定的 ADT 可能有几种不同的实现策略，并探讨这些选择的优点和缺点。我们几乎为书中的所有数据结构和算法都提供了完整的 Python 实现，并介绍了将这些实现组织为可重用的组件所需的重要的面向对象设计模式。

通过阅读本书，读者可以：

- 对常见数据集合的抽象有一定了解（如栈、队列、表、树、图）。
- 理解生成常用数据结构的高效实现的算法策略。
- 通过理论方法和实验方法分析算法的性能，并了解竞争策略之间的权衡。
- 明智地利用编程语言库中已有的数据结构和算法。
- 拥有大多数基础数据结构和算法的具体实现经验。
- 应用数据结构和算法来解决复杂的问题。

为了达到最后一个目标，我们在书中提供了数据结构的很多应用实例，包括：文本处理系统，结构化格式（如 HTML）的标签匹配，简单的密码技术，文字频率分析，自动几何布局，霍夫曼编码，DNA 序列比对，以及搜索引擎索引。

本书特色

本书主要基于由 Goodrich 和 Tamassia 所著的《Data Structures and Algorithms in Java》，以及由 Goodrich、Tamassia 和 Mount 所著的《Data Structures and Algorithms in C++》编写而成。然而，我们并不是简单地用 Python 语言实现以上书籍的内容。为了充实内容，我们重新设计了本书：

- 对全部代码进行了重新设计，以充分利用 Python 的优势，如使用生成器迭代集合的元素。
- 在 Java 和 C++ 版本中，我们提供了很多伪代码，而本书则提供了 Python 实现的完整代码。
- 在一般情况下，ADT 被定义为与 Python 内建数据类型和 Python 的 collections 模块具有一致的接口。
- 第 5 章深入探讨了 Python 中基于动态数组的内置数据结构，如 list、tuple 和 str 类。新增的附录 A 提供了关于 str 类功能的进一步讲解。
- 重新绘制或修改了超过 450 幅插图。
- 经过新增和修订，练习的总数达到 750 个。

在线资源[⊖]

本书提供一系列丰富的在线资源，可访问以下网站获取：

www.wiley.com/college/goodrich

鼓励学生在学习本书时使用这个网址，以更有效地进行练习并提高对所学知识的认识。也欢迎教师使用本网站来帮助规划、组织和展示他们的课程材料。对于教师和学生而言，网站中包含一系列与本书主题相关的教学资源，由于它们是有附加价值的，所以一些网上资源受密码保护。

对于所有的读者，尤其是学生，我们有以下资源：

- 书中所有 Python 程序的源代码。
- 提供给教师的 PDF 讲义版 PPT（每页四张）。
- 保存所有练习提示的数据库，以练习的编号为索引。

对于使用本书的教师，我们有以下额外的教学辅助资源：

- 本书练习的答案。
- 书中所有图形和插图的彩色版本。
- PPT 和 PDF 版本的幻灯片，其中 PDF 版本为每页一张。

PPT 是完全可编辑的，教师可根据自己的课程需求进行修改。在教师使用本书作为教材时，所有的在线资源不收取额外费用。

内容和组织

书中各章节的内容循序渐进，适于教学。从 Python 编程和面向对象设计的基础开始，然后逐渐增加如算法分析和递归之类的基础技术。在本书的主体部分中，我们展示了基本的数据结构和算法，并且包括对内存管理的讨论（也是数据结构的架构基础）。本书的章节安排如下：

第 1 章　Python 入门
第 2 章　面向对象编程
第 3 章　算法分析
第 4 章　递归
第 5 章　基于数组的序列
第 6 章　栈、队列和双端队列
第 7 章　链表
第 8 章　树
第 9 章　优先级队列
第 10 章　映射、哈希表和跳跃表
第 11 章　搜索树
第 12 章　排序与选择
第 13 章　文本处理
第 14 章　图算法
第 15 章　内存管理和 B 树

⊖ 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。需要的读者可访问华章网站 www.hzbook.com 下载 PPT、练习答案和源代码。如果需要其他资源，可向约翰·威立出版公司北京代表处申请，电话 010-84187869，电子邮件 sliang@wiley.com。——编辑注

附录 A　Python 中的字符串
附录 B　有用的数学定理

预备知识

我们假设读者至少接触过一种高级语言，如 C、C++、Python 或 Java，可以理解相关高级语言的主要概念，包括：

- 变量和表达式。
- 决策结构（if 语句和 switch 语句）。
- 迭代结构（for 循环和 while 循环）。
- 函数（无论是过程式方法还是面向对象方法）。

对于已经熟悉这些概念但还不清楚如何在 Python 中应用的读者，我们建议将第 1 章作为 Python 语言的入门。这本书主要讨论数据结构，而不是讲解 Python，因此并没有详尽介绍 Python。

直到第 2 章才开始使用 Python 中的面向对象编程，这一章对于那些 Python 新手以及熟悉 Python 但不熟悉面向对象编程的人都是有用的。

就数学背景而言，我们假定读者多少熟悉些高中数学知识。即便如此，在第 3 章中，我们先讨论了算法分析的 7 个最重要的功能。若所涉及的内容超出了这 7 个功能，则作为可选章节，用星号（*）标记。附录 B 对其他有用的数学定理做了总结，包括初等概率等。

计算机科学课程的设计

为了帮助教师在 IEEE/ACM 2013 的框架下设计教学课程，下表描述了本书涵盖的知识要点。

知识要点	相关章节
AL/ 基本分析	第 3 章，4.2 节，12.2.4 节
AL/ 算法策略	12.2.1 节，13.2.1 节，13.3 节，13.4.2 节
AL/ 基本数据结构与算法	4.1.3 节，5.5.2 节，9.4.1 节，9.3 节，10.2 节，11.1 节，13.2 节，第 12 章，第 14 章的大部分内容
AL/ 高级数据结构	5.3 节，10.4 节，11.2 ～ 11.6 节，12.3.1 节，13.5 节，14.5 节，15.3 节
AR/ 内存系统组织和架构	第 15 章
DS/ 集合、关系和功能	10.5.1 节，10.5.2 节，9.4 节
DS/ 证明技巧	3.4 节，4.2 节，5.3.2 节，9.3.6 节，12.4.1 节
DS/ 基础计数	2.4.2 节，6.2.2 节，12.2.4 节，8.2.2 节，附录 B
DS/ 图和树	第 8 章和第 14 章的大部分内容
DS/ 离散概率	1.11 节，10.2 节，10.4.2 节，12.3.1 节
PL/ 面向对象编程	本书的大部分内容，特别是第 2 章以及 7.4 节、9.5.1 节、10.1.3 节和 11.2 节
PL/ 函数式编程	1.10 节
SDF/ 算法和设计	2.1 节，3.3 节，12.2.1 节
SDF/ 基本编程概念	第 1 章，第 4 章
SDF/ 基本数据结构	第 6 章，第 7 章，附录 A，1.2.1 节，5.2 节，5.4 节，9.1 节，10.1 节
SDF/ 开发方法	1.7 节，2.2 节
SE/ 软件设计	2.1 节，2.1.3 节

致　谢

Data Structures and Algorithms in Python

许多人帮助我们完成了本书。首先要感谢的是 Wiley 这个优秀的团队，感谢我们的编辑 Beth Golub 从始至终对这个项目的热情支持。从最初阶段的提议到通过广泛同行评审的过程中，Elizabeth Mills 和 Katherine Willis 的努力是推动项目持续前进的关键动力。我们非常感谢专注于细节的 Julie Kennedy，她也是本书的文字编辑。最后，非常感谢 Joyce Poh 对于最后几个月的生产过程的管理。

真心感谢评审人员和广大读者，他们丰富的评论、邮件和具有建设性的批评对我们写作本书价值很大。我们要感谢以下评审人员：Claude Anderson（Rose Hulman Institute of Technology），Alistair Campbell（Hamilton College），Barry Cohen（New Jersey Institute of Technology），Robert Franks（Central College），Andrew Harrington（Loyola University Chicago），Dave Musicant（Carleton College），Victor Norman（Calvin College）。特别感谢 Claude 非常负责地给我们提供了 400 条详细的建议。

感谢 David Mount（University of Maryland）慷慨地分享了他从 C++ 版本中获得的经验。感谢 Erin Chambers 和 David Letscher（Saint Louis University）在多年数据结构教学中的默默奉献，以及基于本书早期 Python 代码版本的评论。感谢 David Zampino（Loyola University Chicago 的学生），他在使用本书草稿独立学习后反馈了有益的建议，还要感谢 Andrew Harrington 一直督促着 David 完成学习。

很多同行和助教为本书先前的 C++ 和 Java 版本提供了帮助，那些贡献同样对本书有益，再次感谢他们。

最后，我们要由衷地感谢 Susan Goldwasser、Isabel Cruz、Karen Goodrich、Giuseppe Di Battista、Franco Preparata、Ioannis Tollis 以及我们的父母，他们在本书的不同准备阶段给予我们建议、鼓励和支持。我们还要感谢 Calista 和 Maya Goldwasser 关于许多插图的建议，这些建议提升了图片的艺术水准。更重要的是，有些人不断提醒着我们——生活中不止写书这一件有意义的事。没错，谢谢他们。

Michael T. Goodrich
Roberto Tamassia
Michael H. Goldwasser

Michael Goodrich 于 1987 年从普渡大学获得计算机科学博士学位，目前是加州大学欧文分校计算机科学系校长讲席教授。他之前是约翰·霍普金斯大学的教授。他是富布莱特学者，美国科学促进会（AAAS）、计算机协会（ACM）以及电气和电子工程师学会（IEEE）的会士。他还是 IEEE 计算机协会技术成就奖、ACM 卓越服务奖以及 Pond 本科教学优秀奖的获得者。

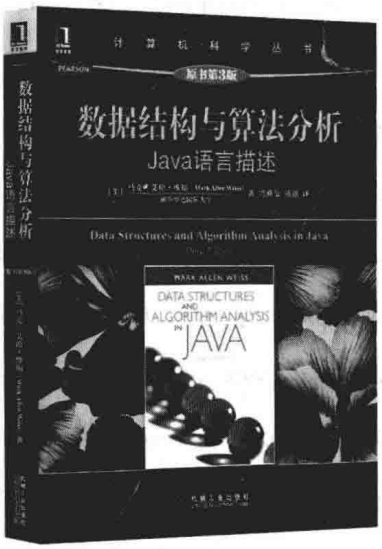
Roberto Tamassia 于 1988 年从伊利诺伊大学厄巴纳－香槟分校获得电子与计算机工程博士学位，目前是布朗大学计算机科学系 Plastech 教授，并担任系主任，同时兼任布朗大学几何计算中心主任。他的研究方向涵盖信息安全、密码学、统计学、算法的设计和实现、图形绘制以及计算几何学。他是 AAAS、ACM 和 IEEE 的会士。他也是 IEEE 计算机协会技术成就奖的获得者。

Michael Goldwasser 于 1997 年从斯坦福大学获得计算机科学博士学位，目前是圣路易斯大学数学和计算机科学系教授，同时兼任计算机科学项目主任。之前，他在芝加哥罗耀拉大学计算机科学系任教。他的研究方向为算法的设计与实现以及计算几何学，同时他还活跃在各种计算机科学的教育社区。

这些作者的其他著作

- M.T. Goodrich and R. Tamassia, *Data Structures and Algorithms in Java*, Wiley.
- M.T. Goodrich, R. Tamassia, and D.M. Mount, *Data Structures and Algorithms in C++*, Wiley.
- M.T. Goodrich and R. Tamassia, *Algorithm Design: Foundations, Analysis, and Internet Examples*, Wiley.
- M.T. Goodrich and R. Tamassia, *Introduction to Computer Security*, Addison-Wesley.
- M.H. Goldwasser and D. Letscher, *Object-Oriented Programming in Python*, Prentice Hall.

推荐阅读



数据结构与算法分析：Java语言描述（原书第3版）

作者：Mark Allen Weiss ISBN：978-7-111-52839-5 定价：69.00元

本书是国外数据结构与算法分析方面的经典教材，使用卓越的Java编程语言作为实现工具，讨论数据结构（组织大量数据的方法）和算法分析（对算法运行时间的估计）。

随着计算机速度的不断增加和功能的日益强大，人们对有效编程和算法分析的要求也不断增长。本书将算法分析与最有效率的Java程序的开发有机结合起来，深入分析每种算法，并细致讲解精心构造程序的方法，内容全面，缜密严格。

算法设计与应用

作者：Michael T. Goodrich等 ISBN：978-7-111-58277-9 定价：139.00元

这是一本非常棒的著作，既有算法的经典内容，也有现代专题。我期待着在我的算法课程试用此教材。我尤其喜欢内容的广度和问题的难度。

——Robert Tarjan，普林斯顿大学

Goodrich和Tamassia编写了一本内容十分广泛而且方法具有创新性的著作。贯穿本书的应用和练习为各个领域学习计算的学生提供了极佳的参考。本书涵盖了超出一学期课程可以讲授的内容，这给教师提供了很大的选择余地，同时也给学生提供了很好的自学材料。

——Michael Mitzenmacher，哈佛大学

目 录

Data Structures and Algorithms in Python

第 1 章

Data Structures and Algorithms in Python

Python 入门

1.1 Python 概述

构建数据结构和算法需要我们了解计算机中详细的指令。一种很好的方法是使用高级计算机语言描述这个了解的过程，如 Python。Python 编程语言最初是由 Guido van Rossum 于 20 世纪 90 年代初开发的，并已在工业和教育领域成为一门十分重要的语言。Python 语言的第二个主要版本 Python 2 于 2000 年发布，而第三个主要版本 Python 3 于 2008 年发布。Python 2 和 Python 3 之间不兼容。本书是基于 Python 3 编写的（更具体地说，基于 Python 3.1 或更高版本）。Python 语言的最新版本及其文档和教程可在 www.python.org 免费获取。

在本章中，我们对 Python 编程语言进行了概述，并将在下一章继续讨论面向对象原则。我们假设这本书的读者已有一定的编程经验，但不一定局限于 Python。本书不提供 Python 语言的完整描述（有许多语言可以参考用于实现这一目的），但它的确介绍了使用的代码片段里所用语言的方方面面。

1.1.1 Python 解释器

Python 是一种解释语言。命令通常在被称为 Python 解释器的软件中执行。Python 解释器接收到一条命令，然后评估该命令，最后返回该命令的结果。解释器可以交互使用（尤其是在调试时），程序员通常提前定义一系列命令，然后把这些命令保存为纯文本文件，这些程序被称为源代码或脚本。对于 Python，源代码通常存储在一个扩展名为 .py 的文件中（例如 demo.py）。

在大多数操作系统中，Python 解释器可以通过在命令行中输入“python”启动。在默认情况下，解释器在交互模式下使用新的工作空间启动。执行命令时，从保存在文件中的一个预定义脚本（例如 demo.py）中把文件名作为调用解释器执行的一个参数（例如 python demo.py），或使用一个额外的 -i 标志来执行脚本，然后进入交互模式（例如 python -i demo.py）。

许多集成开发环境（Integrated Development Environments，IDE）为 Python 提供了更加丰富的软件开发平台，包括一个拥有标准 Python 发行版的 IDLE。IDLE 提供了一个嵌入式的文本编辑器（可显示和编辑 Python 代码），以及一个基本调试器（允许逐步执行程序，以便检查关键变量的值）。

1.1.2 Python 程序预览

下面的代码段 1-1 是一个 Python 程序，用户输入字母表示学生的成绩等级，而后程序由输入数据计算学生平均绩点（Grade-Point Average，GPA）。这个例子所采用的许多技术将在本章的其余部分讨论。这时，我们注意到一些高层次的问题，尤其是对于那些初次接触 Python 这门编程语言的读者。