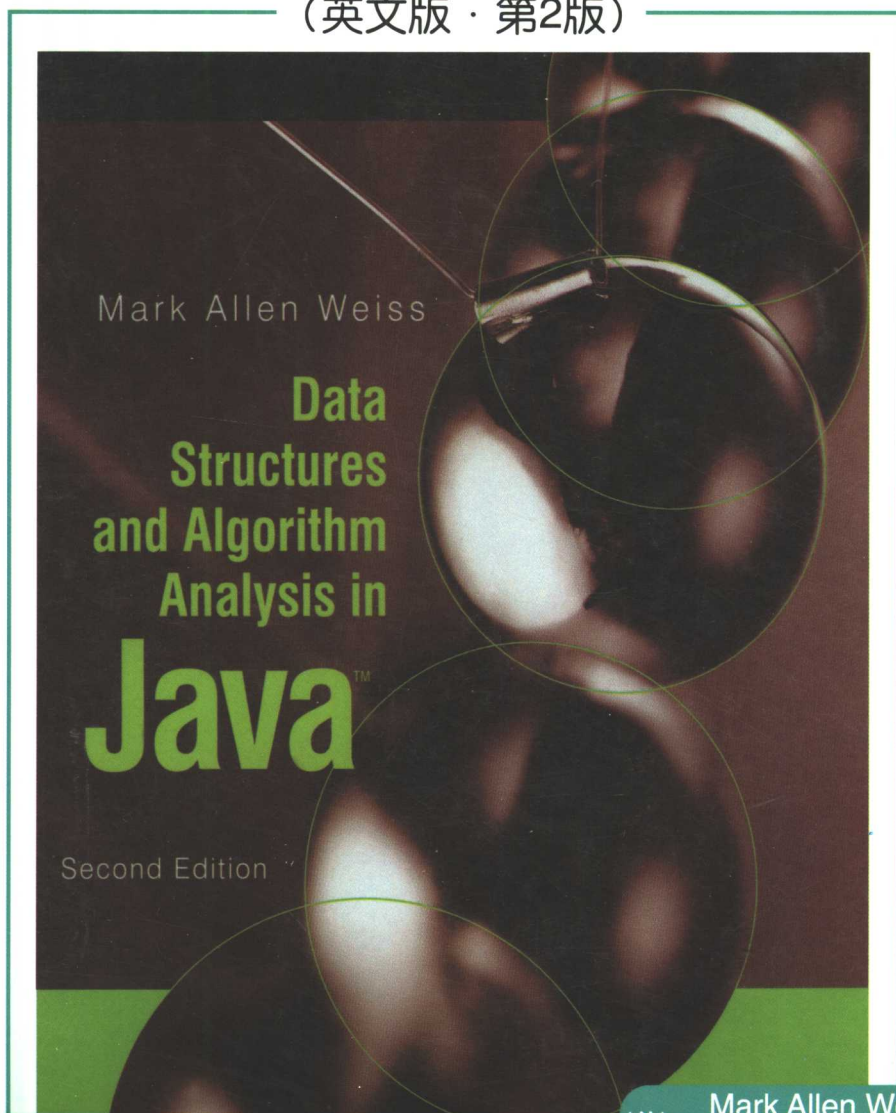


数据结构与算法分析 Java语言描述

(英文版·第2版)



(美) Mark Allen Weiss 著
佛罗里达国际大学

TP312
Y214=2

2007

经典原版书库

数据结构与算法分析

Java语言描述

(英文版·第2版)

Data Structures and Algorithm Analysis in Java
(Second Edition)

江苏工业学院图书馆
藏书章

(美) Mark Allen Weiss 著
佛罗里达国际大学



机械工业出版社
China Machine Press

English reprint edition copyright © 2006 by Pearson Education Asia Limited and China Machine Press.

Original English language title: *Data Structures and Algorithm Analysis in Java, Second Edition* (ISBN 0-321-37013-9) by Mark Allen Weiss, Copyright © 2007 by Pearson Education, Inc.

All rights reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Addison-Wesley.

For sale and distribution in the People's Republic of China exclusively (except Taiwan, Hong Kong SAR and Macau SAR).

本书英文影印版由Pearson Education Asia Ltd.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。

仅限于中华人民共和国境内(不包括中国香港、澳门特别行政区和中国台湾地区)销售发行。

本书封面贴有Pearson Education(培生教育出版集团)激光防伪标签,无标签者不得销售。

版权所有,侵权必究。

本书法律顾问 北京市展达律师事务所

本书版权登记号: 图字: 01-2006-3994

图书在版编目(CIP)数据

数据结构与算法分析: Java语言描述(英文版·第2版)/(美)韦斯(Weiss, M. A.)著.
—北京:机械工业出版社,2007.1

(经典原版书库)

书名原文: *Data Structures and Algorithm Analysis in Java, Second Edition*

ISBN 7-111-19876-X

I. 数… II. 韦… III. ①数据结构—英文 ②算法分析—英文 ③JAVA语言—程序设计—英文 IV. TP311.12

中国版本图书馆CIP数据核字(2006)第106914号

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: 迟振春

北京京北制版印刷厂印刷·新华书店北京发行所发行

2007年1月第1版第1次印刷

170mm×242mm·36印张

定价: 55.00元

凡购本书,如有倒页、脱页、缺页,由本社发行部调换
本社购书热线:(010) 68326294

出版者的话

文艺复兴以降，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中取得了垄断性的优势；也正是这样的传统，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，不仅肇划了研究的范畴，还揭橥了学术的源变，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、从业人员较少的现状下，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图文信息有限公司较早意识到“出版要为教育服务”。自1998年开始，华章公司就将工作重点放在了遴选、移译国外优秀教材上。经过几年的不懈努力，我们与Prentice Hall, Addison-Wesley, McGraw-Hill, Morgan Kaufmann等世界著名出版公司建立了良好的合作关系，从它们现有的数百种教材中甄选出Tanenbaum, Stroustrup, Kernighan, Jim Gray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、研究及度藏。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襄助，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，有的还专程为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近百个品种，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，为进一步推广与发展打下了坚实的基础。

随着学科建设的初步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步入一个新的阶段。为此，华章公司将加大引进教材的力度，在“华章教育”的总规划之下出版三个系列的计算机教材：除“计算机科学丛书”之外，对影印版的教材，则单独开辟出“经典原版书库”；同时，引进全美通行的教学辅导书“Schaum's Outlines”系列组成“全美经典学习指导系列”。为了保证这三套丛书的权威性，同时也为了更好地为学校和老师服务，华章公司聘请了中国科学院、北京大学、清华大学、国防科技大学、复旦大学、上海交通大学、南京大学、浙江大学、中国科技大学、哈尔滨工业大学、西安交通大学、中国人民大学、北京航空航天大学、北京邮电大学、中山大学、解放军理工大学、郑州大学、湖北工学院、

中国国家信息安全测评认证中心等国内重点大学和科研机构在计算机的各个领域的著名学者组成“专家指导委员会”，为我们提供选题意见和出版监督。

这三套丛书是响应教育部提出的使用外版教材的号召，为国内高校的计算机及相关专业的教学度身订造的。其中许多教材均已为M. I. T., Stanford, U.C. Berkeley, C. M. U. 等世界名牌大学所采用。不仅涵盖了程序设计、数据结构、操作系统、计算机体系结构、数据库、编译原理、软件工程、图形学、通信与网络、离散数学等国内大学计算机专业普遍开设的核心课程，而且各具特色——有的出自语言设计者之手、有的历经三十年而不衰、有的已被全世界的几百所高校采用。在这些圆熟通博的名师大作的指引之下，读者必将在计算机科学的宫殿中由登堂而入室。

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，但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。教材的出版只是我们的后续服务的起点。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：

电子邮件: hzjsj@hzbook.com

联系电话: (010) 68995264

联系地址: 北京市西城区百万庄南街1号

邮政编码: 100037

专家指导委员会

(按姓氏笔画顺序)

尤晋元
石教英
张立昂
邵维忠
周克定
郑国梁
高传善
裘宗燕

王 珊
吕 建
李伟琴
陆丽娜
周傲英
施伯乐
梅 宏
戴 葵

冯博琴
孙玉芳
李师贤
陆鑫达
孟小峰
钟玉琢
程 旭

史忠植
吴世忠
李建中
陈向群
岳丽华
唐世渭
程时端

史美林
吴时霖
杨冬青
周伯生
范 明
袁崇义
谢希仁

To my one and only, Jill.

Purpose/Goals

This new Java edition describes *data structures*, methods of organizing large amounts of data, and *algorithm analysis*, the estimation of the running time of algorithms. As computers become faster and faster, the need for programs that can handle large amounts of input becomes more acute. Paradoxically, this requires more careful attention to efficiency, since inefficiencies in programs become most obvious when input sizes are large. By analyzing an algorithm before it is actually coded, students can decide if a particular solution will be feasible. For example, in this text students look at specific problems and see how careful implementations can reduce the time constraint for large amounts of data from 16 years to less than a second. Therefore, no algorithm or data structure is presented without an explanation of its running time. In some cases, minute details that affect the running time of the implementation are explored.

Once a solution method is determined, a program must still be written. As computers have become more powerful, the problems they must solve have become larger and more complex, requiring development of more intricate programs. The goal of this text is to teach students good programming and algorithm analysis skills simultaneously so that they can develop such programs with the maximum amount of efficiency.

This book is suitable for either an advanced data structures (CS7) course or a first-year graduate course in algorithm analysis. Students should have some knowledge of intermediate programming, including such topics as object-based programming and recursion, and some background in discrete math.

Approach

Although the material in this text is largely language independent, programming requires the use of a specific language. As the title implies, we have chosen Java for this book.

Java is a relatively new language that is often examined in comparison with C++. Java offers many benefits, and programmers often view Java as a safer, more portable, and easier-to-use language than C++. As such, it makes a fine core language for discussing and implementing fundamental data structures. Other important parts of Java, such as threads and its GUI, although important, are not needed in this text and thus are not discussed.

Complete versions of the data structures, in both Java and C++, are available on the Internet. We use similar coding conventions to make the parallels between the two languages more evident.

Summary of the Most Significant Changes in the Second Edition

The second edition incorporates numerous bug fixes, and many parts of the book have undergone revision to increase clarity of presentation. In addition:

- Throughout the text, the code has been updated as appropriate to use modern features from Java 5.0.
- Chapter 3 has been significantly revised and contains a discussion of the use of the standard `ArrayList` and `LinkedList` classes (and their iterators) as well as an implementation of the standard `ArrayList` and `LinkedList` classes.
- Chapter 4 has been revised to include a discussion of the `TreeSet` and `TreeMap` classes, along with an extensive example illustrating their use in the design of efficient algorithms. Chapter 9 also includes an example that makes use of the standard `TreeMap` to implement a shortest path algorithm.
- Chapter 7 contains a discussion of the standard sort algorithms, including an illustration of the techniques involved in implementing the overloaded standard sort algorithms.

Overview

Chapter 1 contains review material on discrete math and recursion. I believe the only way to be comfortable with recursion is to see good uses over and over. Therefore, recursion is prevalent in this text, with examples in every chapter except Chapter 5. Chapter 1 also presents material that serves as a review of inheritance in Java. Included is a discussion of Java 5 generics.

Chapter 2 deals with algorithm analysis. This chapter explains asymptotic analysis and its major weaknesses. Many examples are provided, including an in-depth explanation of logarithmic running time. Simple recursive programs are analyzed by intuitively converting them into iterative programs. More complicated divide-and-conquer programs are introduced, but some of the analysis (solving recurrence relations) is implicitly delayed until Chapter 7, where it is performed in detail.

Chapter 3 covers lists, stacks, and queues. This chapter has been significantly revised from prior editions. It now includes a discussion of the Collections API `ArrayList` and `LinkedList` classes, and it provides implementations of a significant subset of the collections API `ArrayList` and `LinkedList` classes.

Chapter 4 covers trees, with an emphasis on search trees, including external search trees (B-trees). The UNIX file system and expression trees are used as examples. AVL trees and splay trees are introduced. More careful treatment of search tree implementation details is found in Chapter 12. Additional coverage of trees, such as file compression and game trees, is deferred until Chapter 10. Data structures for an external medium are considered as the final topic in several chapters. New to this edition is a discussion of the Collections API `TreeSet` and `TreeMap` classes, including a significant example that illustrates the use of three separate maps to efficiently solve a problem.

Chapter 5 is a relatively short chapter concerning hash tables. Some analysis is performed, and extendible hashing is covered at the end of the chapter.

Chapter 6 is about priority queues. Binary heaps are covered, and there is additional material on some of the theoretically interesting implementations of priority queues. The Fibonacci heap is discussed in Chapter 11, and the pairing heap is discussed in Chapter 12.

Chapter 7 covers sorting. It is very specific with respect to coding details and analysis. All the important general-purpose sorting algorithms are covered and compared. Four algorithms are analyzed in detail: insertion sort, Shellsort, heapsort, and quicksort. External sorting is covered at the end of the chapter.

Chapter 8 discusses the disjoint set algorithm with proof of the running time. This is a short and specific chapter that can be skipped if Kruskal's algorithm is not discussed.

Chapter 9 covers graph algorithms. Algorithms on graphs are interesting, not only because they frequently occur in practice but also because their running time is so heavily dependent on the proper use of data structures. Virtually all of the standard algorithms are presented along with appropriate data structures, pseudocode, and analysis of running time. To place these problems in a proper context, a short discussion on complexity theory (including *NP*-completeness and undecidability) is provided.

Chapter 10 covers algorithm design by examining common problem-solving techniques. This chapter is heavily fortified with examples. Pseudocode is used in these later chapters so that the student's appreciation of an example algorithm is not obscured by implementation details.

Chapter 11 deals with amortized analysis. Three data structures from Chapters 4 and 6 and the Fibonacci heap, introduced in this chapter, are analyzed.

Chapter 12 covers search tree algorithms, the k -d tree, and the pairing heap. This chapter departs from the rest of the text by providing complete and careful implementations for the search trees and pairing heap. The material is structured so that the instructor can integrate sections into discussions from other chapters. For example, the top-down red-black tree in Chapter 12 can be discussed along with AVL trees (in Chapter 4).

Chapters 1–9 provide enough material for most one-semester data structures courses. If time permits, then Chapter 10 can be covered. A graduate course on algorithm analysis could cover Chapters 7–11. The advanced data structures analyzed in Chapter 11 can easily be referred to in the earlier chapters. The discussion of *NP*-completeness in Chapter 9 is far too brief to be used in such a course. You might find it useful to use an additional work on *NP*-completeness to augment this text.

Exercises

Exercises, provided at the end of each chapter, match the order in which material is presented. The last exercises may address the chapter as a whole rather than a specific section. Difficult exercises are marked with an asterisk, and more challenging exercises have two asterisks.

References

References are placed at the end of each chapter. Generally the references either are historical, representing the original source of the material, or they represent extensions and improvements to the results given in the text. Some references represent solutions to exercises.

Supplements

The following supplements are available to all readers at www.aw.com/cssupport:

- Source code for example programs

In addition, the following material is available only to qualified instructors at Addison-Wesley's Instructor Resource Center (www.aw.com/irc). Visit the IRC or contact your campus A-W representative for access.

- Solutions to selected exercises
- Figures from the book

Acknowledgments

Many, many people have helped me in the preparation of books in this series. Some are listed in other versions of the book; thanks to all.

As usual, the writing process was made easier by the professionals at Addison-Wesley. I'd like to thank my editor, Michael Hirsch, and production editor, Marilyn Lloyd. I'd also like to thank Paul Anagnostopoulos and his staff at Windfall Software for their fine work putting the final pieces together. My wonderful wife Jill deserves extra special thanks for everything she does.

Finally, I'd like to thank the numerous readers who have sent e-mail messages and pointed out errors or inconsistencies in earlier versions. My World Wide Web page www.cis.fiu.edu/~weiss will also contain updated source code (in Java, C++, and C), an errata list, and a link to submit bug reports.

M.A.W.
Miami, Florida

Preface vii

Chapter 1 Introduction

1

1.1	What's the Book About?	1
1.2	Mathematics Review	2
1.2.1	Exponents	3
1.2.2	Logarithms	3
1.2.3	Series	4
1.2.4	Modular Arithmetic	5
1.2.5	The <i>P</i> Word	6
1.3	A Brief Introduction to Recursion	7
1.4	Implementing Generic Components Pre Java 5	11
1.4.1	Using <code>Object</code> for Genericity	12
1.4.2	Wrappers for Primitive Types	12
1.4.3	Using Interface Types for Genericity	13
1.4.4	Compatibility of Array Types	15
1.5	Implementing Generic Components Using Java 5 Generics	16
1.5.1	Simple Generic Classes and Interfaces	16
1.5.2	Autoboxing/Unboxing	17
1.5.3	Wildcards with Bounds	18
1.5.4	Generic Static Methods	19
1.5.5	Type Bounds	20
1.5.6	Type Erasure	21
1.5.7	Restrictions on Generics	22
1.6	Function Objects	23
	Summary	25
	Exercises	25
	References	26

Chapter 2 Algorithm Analysis

29

- 2.1 Mathematical Background 29
- 2.2 Model 32
- 2.3 What to Analyze 32
- 2.4 Running Time Calculations 35
 - 2.4.1 A Simple Example 35
 - 2.4.2 General Rules 36
 - 2.4.3 Solutions for the Maximum Subsequence Sum Problem 38
 - 2.4.4 Logarithms in the Running Time 44
 - 2.4.5 Checking Your Analysis 48
 - 2.4.6 A Grain of Salt 48
- Summary 50
- Exercises 50
- References 55

Chapter 3 Lists, Stacks, and Queues

57

- 3.1 Abstract Data Types (ADTs) 57
- 3.2 The List ADT 58
 - 3.2.1 Simple Array Implementation of Lists 58
 - 3.2.2 Simple Linked Lists 59
- 3.3 Lists in the Java Collections API 60
 - 3.3.1 Collection Interface 61
 - 3.3.2 Iterators 62
 - 3.3.3 The List Interface, ArrayList, and LinkedList 63
 - 3.3.4 Example: Using remove on a LinkedList 65
 - 3.3.5 ListIterators 66
- 3.4 Implementation of ArrayList 67
 - 3.4.1 The Basic Class 68
 - 3.4.2 The Iterator and Java Nested and Inner Classes 68
- 3.5 Implementation of LinkedList 75
- 3.6 The Stack ADT 82
 - 3.6.1 Stack Model 82
 - 3.6.2 Implementation of Stacks 83
 - 3.6.3 Applications 83
- 3.7 The Queue ADT 91
 - 3.7.1 Queue Model 91
 - 3.7.2 Array Implementation of Queues 91
 - 3.7.3 Applications of Queues 94
- Summary 95
- Exercises 95

Chapter 4 Trees

101

- 4.1 Preliminaries 101
 - 4.1.1 Implementation of Trees 102
 - 4.1.2 Tree Traversals with an Application 103
- 4.2 Binary Trees 107
 - 4.2.1 Implementation 108
 - 4.2.2 An Example: Expression Trees 109
- 4.3 The Search Tree ADT—Binary Search Trees 112
 - 4.3.1 contains 113
 - 4.3.2 findMin and findMax 115
 - 4.3.3 insert 115
 - 4.3.4 remove 117
 - 4.3.5 Average-Case Analysis 120
- 4.4 AVL Trees 123
 - 4.4.1 Single Rotation 125
 - 4.4.2 Double Rotation 128
- 4.5 Splay Trees 135
 - 4.5.1 A Simple Idea (That Does Not Work) 135
 - 4.5.2 Splaying 137
- 4.6 Tree Traversals (Revisited) 143
- 4.7 B-Trees 145
- 4.8 Sets and Maps in the Standard Library 150
 - 4.8.1 Sets 151
 - 4.8.2 Maps 151
 - 4.8.3 Implementation of TreeSet and TreeMap 152
 - 4.8.4 An Example That Uses Several Maps 152
- 4.9 Summary 157
 - Exercises 159
 - References 165

Chapter 5 Hashing

169

- 5.1 General Idea 169
- 5.2 Hash Function 170
- 5.3 Separate Chaining 172
- 5.4 Hash Tables Without Linked Lists 177
 - 5.4.1 Linear Probing 177
 - 5.4.2 Quadratic Probing 179
 - 5.4.3 Double Hashing 181
- 5.5 Rehashing 186

5.6	Hash Tables in the Standard Library	187
5.7	Extendible Hashing	190
	Summary	193
	Exercises	194
	References	198

Chapter 6 Priority Queues (Heaps)

201

6.1	Model	201
6.2	Simple Implementations	202
6.3	Binary Heap	202
	6.3.1 Structure Property	203
	6.3.2 Heap Order Property	205
	6.3.3 Basic Heap Operations	205
	6.3.4 Other Heap Operations	210
6.4	Applications of Priority Queues	214
	6.4.1 The Selection Problem	214
	6.4.2 Event Simulation	215
6.5	d -Heaps	216
6.6	Leftist Heaps	217
	6.6.1 Leftist Heap Property	217
	6.6.2 Leftist Heap Operations	218
6.7	Skew Heaps	225
6.8	Binomial Queues	227
	6.8.1 Binomial Queue Structure	228
	6.8.2 Binomial Queue Operations	229
	6.8.3 Implementation of Binomial Queues	232
6.9	Priority Queues in the Standard Library	237
	Summary	237
	Exercises	239
	References	243

Chapter 7 Sorting

247

7.1	Preliminaries	247
7.2	Insertion Sort	248
	7.2.1 The Algorithm	248
	7.2.2 Analysis of Insertion Sort	248
7.3	A Lower Bound for Simple Sorting Algorithms	249
7.4	Shellsort	250
	7.4.1 Worst-Case Analysis of Shellsort	252

7.5	Heapsort	254
7.5.1	Analysis of Heapsort	256
7.6	Mergesort	258
7.6.1	Analysis of Mergesort	260
7.7	Quicksort	264
7.7.1	Picking the Pivot	264
7.7.2	Partitioning Strategy	266
7.7.3	Small Arrays	268
7.7.4	Actual Quicksort Routines	268
7.7.5	Analysis of Quicksort	271
7.7.6	A Linear-Expected-Time Algorithm for Selection	274
7.8	A General Lower Bound for Sorting	276
7.8.1	Decision Trees	276
7.9	Bucket Sort	278
7.10	External Sorting	279
7.10.1	Why We Need New Algorithms	279
7.10.2	Model for External Sorting	279
7.10.3	The Simple Algorithm	279
7.10.4	Multiway Merge	281
7.10.5	Polyphase Merge	282
7.10.6	Replacement Selection	283
	Summary	284
	Exercises	285
	References	290

Chapter 8 The Disjoint Set Class

293

8.1	Equivalence Relations	293
8.2	The Dynamic Equivalence Problem	294
8.3	Basic Data Structure	295
8.4	Smart Union Algorithms	299
8.5	Path Compression	301
8.6	Worst Case for Union-by-Rank and Path Compression	303
8.6.1	Analysis of the Union/Find Algorithm	304
8.7	An Application	309
	Summary	312
	Exercises	312
	References	314

Chapter 9 Graph Algorithms

317

- 9.1 Definitions 317
 - 9.1.1 Representation of Graphs 318
- 9.2 Topological Sort 320
- 9.3 Shortest-Path Algorithms 323
 - 9.3.1 Unweighted Shortest Paths 325
 - 9.3.2 Dijkstra's Algorithm 329
 - 9.3.3 Graphs with Negative Edge Costs 338
 - 9.3.4 Acyclic Graphs 338
 - 9.3.5 All-Pairs Shortest Path 342
 - 9.3.6 Shortest-Path Example 342
- 9.4 Network Flow Problems 344
 - 9.4.1 A Simple Maximum-Flow Algorithm 344
- 9.5 Minimum Spanning Tree 349
 - 9.5.1 Prim's Algorithm 351
 - 9.5.2 Kruskal's Algorithm 353
- 9.6 Applications of Depth-First Search 355
 - 9.6.1 Undirected Graphs 357
 - 9.6.2 Biconnectivity 358
 - 9.6.3 Euler Circuits 361
 - 9.6.4 Directed Graphs 366
 - 9.6.5 Finding Strong Components 367
- 9.7 Introduction to NP-Completeness 369
 - 9.7.1 Easy vs. Hard 369
 - 9.7.2 The Class NP 370
 - 9.7.3 NP-Complete Problems 371
- Summary 373
- Exercises 373
- References 381

Chapter 10 Algorithm Design Techniques

385

- 10.1 Greedy Algorithms 385
 - 10.1.1 A Simple Scheduling Problem 386
 - 10.1.2 Huffman Codes 389
 - 10.1.3 Approximate Bin Packing 395
- 10.2 Divide and Conquer 403
 - 10.2.1 Running Time of Divide and Conquer Algorithms 404
 - 10.2.2 Closest-Points Problem 406
 - 10.2.3 The Selection Problem 411
 - 10.2.4 Theoretical Improvements for Arithmetic Problems 414