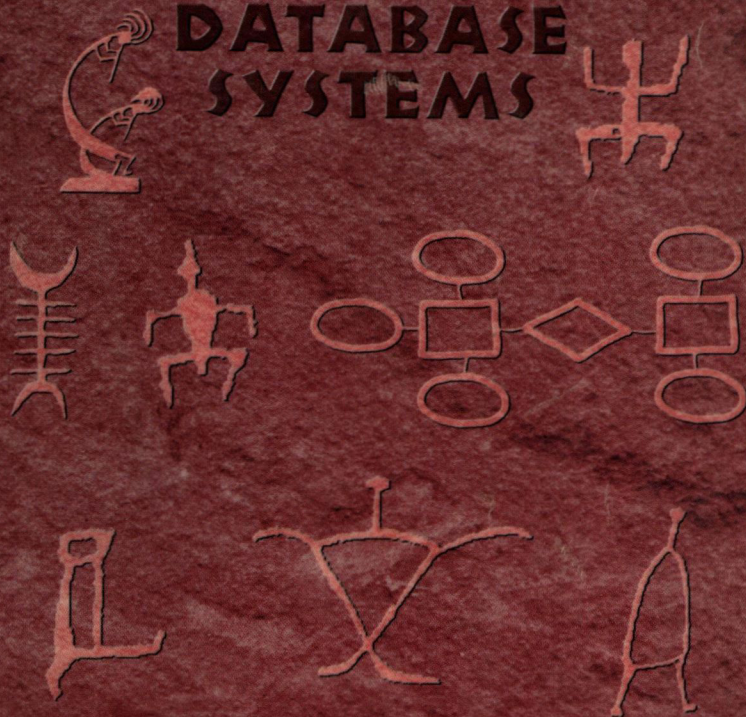


数据库系统基础教程

(英文版·第2版)

SECOND EDITION

A FIRST COURSE IN DATABASE SYSTEMS



JEFFREY D. ULLMAN
JENNIFER WIDOM

(美)

Jeffrey D. Ullman

斯坦福大学
Jennifer Widom

斯坦福大学

著



机械工业出版社
China Machine Press

数据库系统基础教程

(英文版·第2版)

A First Course in Database Systems

(Second Edition)

本书是《Database Systems: The Complete Book》(《数据库系统全书》, 由Hector Garcia-Molina、Jeffrey D. Ullman和Jennifer Widom合著, 中文版由机械工业出版社引进出版)一书的前十章, 由斯坦福大学知名计算机科学家Jeffrey D. Ullman和Jennifer Widom合作编写。作者采用一种易于理解的、面向用户的方式介绍了数据库系统, 主要侧重于介绍数据库的设计与使用以及实际数据库应用的实现。本书涵盖了最新数据库标准SQL-99、SQL PSM、SQL CLI和ODL等内容, 从数据库设计者、用户和应用程序员的角度深入浅出地介绍了数据库, 既可用做大学教科书, 也可作为数据库领域专业人员的参考书。

本书显著特色

- 通过使用人们普遍关注的、现实世界的例子提高可读性。
- 对SQL编程的许多方面展开了广泛而全面的讨论。
- 给出基于常见的E/R模型的设计原则。
- 用ODMG标准ODL介绍了面向对象设计, 用SQL-99标准介绍了对象-关系设计。
- 解释了SQL是如何通过JDBC (Java) 和SQL/CLI (ODBC) 与现代宿主语言交互作用的。
- 涵盖了对于数据库设计者和用户十分重要的若干高级论题, 包括完整性约束、触发器、事务、授权、SQL-99递归, 等等。
- 本书网站<http://www-db.stanford.edu/~ullman/fcdb.html>提供了更多的教学材料。

作者简介

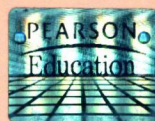
Jeffrey D. Ullman 斯坦福大学计算机科学教授。他独立或与其他人合作出版了16本著作, 其中包括《Elements of ML Programming》(Prentice Hall, 1998)。他的研究兴趣包括数据库理论、数据库集成、数据挖掘和利用信息基础设施进行教育。他获得了Guggenheim Fellowship等多种奖励, 并且是美国国家工程院成员。他还荣获Sigmod贡献奖、Karl V. Karlstrom杰出教育家奖和Knuth奖。



Jennifer Widom 于1987年在康奈尔大学获得计算机科学博士学位, 现为斯坦福大学计算机科学与电气工程教授。她是ACM Fellow、Guggenheim Fellow和美国国家工程院成员, 并且是多个编辑委员会、程序委员会和顾问委员会的成员。她的研究兴趣包括半结构化数据的数据库系统和XML、数据仓库以及主动数据库系统。



中文版
ISBN 7-111-10095-6
定价: 32.00元



www.PearsonEd.com

上架指导: 计算机/数据库



封面设计: 吴刚



华章图书

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

华章网站 <http://www.hzbook.com>

网上购书: www.china-pub.com

投稿热线: (010) 88379604

购书热线: (010) 68995259, 68995264

读者信箱: hzsj@hzbook.com

ISBN 7-111-19344-X

定价: 59.00元



数据库系统基础教程

A First Course in Database Systems

(美)

Jeffrey D. Ullman
Jennifer Widom

著

业出版社
Machine Press

经典原版书库

数据库系统基础教程

(英文版·第2版)

A First Course in Database Systems
(Second Edition)

(美) Jeffrey D. Ullman
斯坦福大学
Jennifer Widom
斯坦福大学

著



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: *A First Course in Database Systems, Second Edition* (ISBN 0-13-035300-0) by Jeffrey D. Ullman and Jennifer Widom, Copyright © 2002, 1997 by Prentice Hall.

All rights reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Prentice Hall.

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:01-2006-3111

图书在版编目(CIP)数据

数据库系统基础教程(英文版·第2版)/ (美)厄尔曼(Ullman, J. D.)等著. -北京:机械工业出版社, 2006. 7

(经典原版书库)

书名原文: *First Course in Database Systems, Second Edition*

ISBN 7-111-19344-X

I. 数… II. 厄… III. 数据库系统-教材-英文 IV. TP311.13

中国版本图书馆CIP数据核字(2006)第062558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

北京瑞德印刷有限公司印刷·新华书店北京发行所发行

2006年7月第1版第1次印刷

170mm × 242mm · 33.25印张

定价:59.00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，我们的联系方式如下：

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



Preface

At Stanford, we are on the quarter system, and as a result, our introductory database instruction is divided into two courses. The first, CS145, is designed for students who will use database systems but not necessarily take a job implementing a DBMS. It is a prerequisite for CS245, which is the introduction to DBMS implementation. Students wishing to go further in the database field then take CS345 (theory), CS346 (DBMS implementation project), and CS347 (transaction processing and distributed databases).

Starting in 1997, we published a pair of books. *A First Course in Database Systems* was designed for CS145, and *Database System Implementation* was for CS245 and parts of CS346. Because many schools are on the semester system or combine the two kinds of database instruction into one introductory course, we felt that there was a need to produce the two books as a single volume, which we call *Database Systems: The Complete Book*.

However, because many more students need to know how to use database systems than to implement them, we have continued to package the material originally in *A First Course in Database Systems* as the present book, giving it the new material on modeling and programming in the first ten chapters of *Database Systems: The Complete Book*. This new material includes object-relational data, SQL/PSM (stored programs), SQL/CLI (the emerging standard for the C/SQL interface), and JDBC (the same for Java/SQL). At the same time, we have reorganized the material, separating the treatment of object-oriented models from coverage of the entity-relationship model, and separating the discussion of logical queries from that of relational algebra. The latter has been expanded to include operators not in the traditional relational algebra but necessary to reflect the semantics of SQL.

- Warning: because this book is the first part of a larger book, it contains certain cross-references to chapters not present in this volume. If you are uninterested in DBMS implementation, the lack of the referenced sections will not impede your understanding of the modeling and programming issues considered here. However, if you are considering buying this book, and anticipate a later study of implementation, you should consider buying *Database Systems: The Complete Book* instead.

Use of the Book

There is adequate material in this volume for a one-semester course on database modeling and programming. For a one-quarter course, you will probably have to omit some of the topics. We regard Chapters 2–7 as the core of the course. The remaining three chapters contain material from which it is safe to select at will, although we believe that every student should get some exposure to the issues of embedding SQL in standard host languages from one of the sections in Chapter 8.

If, as we do in CS145, you give students a substantial database-application design and implementation project, then you may have to reorder the material somewhat, so that SQL instruction occurs earlier in the book. You may wish to defer material such as dependencies, although students need normalization for design.

Prerequisites

We have used the book at the “mezzanine” level, in a course taken both by undergraduates and beginning graduate students. The formal prerequisites for the course are Sophomore-level treatments of: (1) Data structures, algorithms, and discrete math, and (2) Software systems, software engineering, and programming languages. Of this material, it is important that students have at least a rudimentary understanding of such topics as: algebraic expressions and laws, logic, basic data structures, object-oriented programming concepts, and programming environments. However, we believe that adequate background is acquired by the Junior year of a typical computer science program.

Exercises

The book contains extensive exercises, with some for almost every section. We indicate harder exercises or parts of exercises with an exclamation point. The hardest exercises have a double exclamation point.

Some of the exercises or parts are marked with a star. For these exercises, we shall endeavor to maintain solutions accessible through the book’s web page. These solutions are publicly available and should be used for self-testing. Note that in a few cases, one exercise B asks for modification or adaptation of your solution to another exercise A . If certain parts of A have solutions, then you should expect the corresponding parts of B to have solutions as well.

Support on the World Wide Web

The book’s home page is

<http://www-db.stanford.edu/~ullman/fcdb.html>

There are solutions to starred exercises, errata as we learn of them, and backup materials. We are making available the notes for each offering of CS145 as we teach it, including homeworks, projects and exams.

Acknowledgements

A large number of people have helped us, either with the initial vetting of the text for this book and its predecessors, or by contacting us with errata in the books and/or other Web-based materials. It is our pleasure to acknowledge them all here.

Marc Abromowitz, Joseph H. Adamski, Brad Adelberg, Gleb Ashimov, Donald Aingworth, Jonathan Becker, Margaret Benitez, Larry Bonham, Phillip Bonnet, David Brokaw, Ed Burns, Karen Butler, Christopher Chan, Sudarshan Chawathe, Per Christensen, Ed Chang, Surajit Chaudhuri, Ken Chen, Rada Chirkova, Nitin Chopra, Bobbie Cochrane, Arturo Crespo, Linda DeMichiel, Tom Dienstbier, Pearl D'Souza, Oliver Duschka.

Also Xavier Faz, Greg Fichtenholtz, Bart Fisher, Jarl Friis, John Fry, Chipping Fu, Tracy Fujieda, Manish Godara, Meredith Goldsmith, Luis Gravano, Gerard Guillemette, Rafael Hernandez, Antti Hjelt, Ben Holtzman, Steve Huntsberry, Leonard Jacobson, Thulasiraman Jeyaraman, Dwight Joe, Seth Katz, Yeong-Ping Koh, Gyorgy Kovacs, Phillip Koza, Brian Kulman, Sang Ho Lee, Olivier Lobry, Lu Chao-Jun, Arun Marathe, Le-Wei Mo, Fabian Modoux, Peter Mork, Mark Mortensen.

Also Ramprakash Narayanaswami, Hankyung Na, Marie Nilsson, Torbjorn Norbye, Chang-Min Oh, Mehul Patel, Bert Porter, Limbek Reka, Prahash Ramanan, Ken Ross, Tim Roughgarten, Mema Roussopoulos, Richard Scherl, Catherine Tornabene, Anders Uhl, Jonathan Ullman, Mayank Upadhyay, Vasilis Vassalos, Qiang Wang, Kristian Widjaja, Janet Wu, Sundar Yamunachari, Takeshi Yokukawa, Min-Sig Yun, Torben Zahle, Sandy Zhang. The remaining errors are ours, of course.

J. D. U.

J. W.

Stanford, CA

July, 2001



Table of Contents

1	The Worlds of Database Systems	1
1.1	The Evolution of Database Systems	2
1.1.1	Early Database Management Systems	2
1.1.2	Relational Database Systems	4
1.1.3	Smaller and Smaller Systems	5
1.1.4	Bigger and Bigger Systems	6
1.1.5	Client-Server and Multi-Tier Architectures	7
1.1.6	Multimedia Data	8
1.1.7	Information Integration	8
1.2	Overview of a Database Management System	9
1.2.1	Data-Definition Language Commands	10
1.2.2	Overview of Query Processing	10
1.2.3	Storage and Buffer Management	12
1.2.4	Transaction Processing	13
1.2.5	The Query Processor	14
1.3	Outline of Database-System Studies	15
1.3.1	Database Design	16
1.3.2	Database Programming	17
1.3.3	Database System Implementation	17
1.3.4	Information Integration Overview	19
1.4	Summary of Chapter 1	19
1.5	References for Chapter 1	20
2	The Entity-Relationship Data Model	23
2.1	Elements of the E/R Model	24
2.1.1	Entity Sets	24
2.1.2	Attributes	25
2.1.3	Relationships	25
2.1.4	Entity-Relationship Diagrams	25
2.1.5	Instances of an E/R Diagram	27
2.1.6	Multiplicity of Binary E/R Relationships	27
2.1.7	Multiway Relationships	28
2.1.8	Roles in Relationships	29

2.1.9	Attributes on Relationships	31
2.1.10	Converting Multiway Relationships to Binary	32
2.1.11	Subclasses in the E/R Model	33
2.1.12	Exercises for Section 2.1	36
2.2	Design Principles	39
2.2.1	Faithfulness	39
2.2.2	Avoiding Redundancy	39
2.2.3	Simplicity Counts	40
2.2.4	Choosing the Right Relationships	40
2.2.5	Picking the Right Kind of Element	42
2.2.6	Exercises for Section 2.2	44
2.3	The Modeling of Constraints	47
2.3.1	Classification of Constraints	47
2.3.2	Keys in the E/R Model	48
2.3.3	Representing Keys in the E/R Model	50
2.3.4	Single-Value Constraints	51
2.3.5	Referential Integrity	51
2.3.6	Referential Integrity in E/R Diagrams	52
2.3.7	Other Kinds of Constraints	53
2.3.8	Exercises for Section 2.3	53
2.4	Weak Entity Sets	54
2.4.1	Causes of Weak Entity Sets	54
2.4.2	Requirements for Weak Entity Sets	56
2.4.3	Weak Entity Set Notation	57
2.4.4	Exercises for Section 2.4	58
2.5	Summary of Chapter 2	59
2.6	References for Chapter 2	60
3	The Relational Data Model	61
3.1	Basics of the Relational Model	61
3.1.1	Attributes	62
3.1.2	Schemas	62
3.1.3	Tuples	62
3.1.4	Domains	63
3.1.5	Equivalent Representations of a Relation	63
3.1.6	Relation Instances	64
3.1.7	Exercises for Section 3.1	64
3.2	From E/R Diagrams to Relational Designs	65
3.2.1	From Entity Sets to Relations	66
3.2.2	From E/R Relationships to Relations	67
3.2.3	Combining Relations	70
3.2.4	Handling Weak Entity Sets	71
3.2.5	Exercises for Section 3.2	75
3.3	Converting Subclass Structures to Relations	76
3.3.1	E/R-Style Conversion	77

3.3.2	An Object-Oriented Approach	78
3.3.3	Using Null Values to Combine Relations	79
3.3.4	Comparison of Approaches	79
3.3.5	Exercises for Section 3.3	80
3.4	Functional Dependencies	82
3.4.1	Definition of Functional Dependency	83
3.4.2	Keys of Relations	84
3.4.3	Superkeys	86
3.4.4	Discovering Keys for Relations	87
3.4.5	Exercises for Section 3.4	88
3.5	Rules About Functional Dependencies	90
3.5.1	The Splitting/Combining Rule	90
3.5.2	Trivial Functional Dependencies	92
3.5.3	Computing the Closure of Attributes	92
3.5.4	Why the Closure Algorithm Works	95
3.5.5	The Transitive Rule	96
3.5.6	Closing Sets of Functional Dependencies	98
3.5.7	Projecting Functional Dependencies	98
3.5.8	Exercises for Section 3.5	100
3.6	Design of Relational Database Schemas	102
3.6.1	Anomalies	103
3.6.2	Decomposing Relations	103
3.6.3	Boyce-Codd Normal Form	105
3.6.4	Decomposition into BCNF	107
3.6.5	Recovering Information from a Decomposition	112
3.6.6	Third Normal Form	114
3.6.7	Exercises for Section 3.6	117
3.7	Multivalued Dependencies	118
3.7.1	Attribute Independence and Its Consequent Redundancy	118
3.7.2	Definition of Multivalued Dependencies	119
3.7.3	Reasoning About Multivalued Dependencies	120
3.7.4	Fourth Normal Form	122
3.7.5	Decomposition into Fourth Normal Form	123
3.7.6	Relationships Among Normal Forms	124
3.7.7	Exercises for Section 3.7	126
3.8	Summary of Chapter 3	127
3.9	References for Chapter 3	129
4	Other Data Models	131
4.1	Review of Object-Oriented Concepts	132
4.1.1	The Type System	132
4.1.2	Classes and Objects	133
4.1.3	Object Identity	133
4.1.4	Methods	133
4.1.5	Class Hierarchies	134

4.2	Introduction to ODL	135
4.2.1	Object-Oriented Design	135
4.2.2	Class Declarations	136
4.2.3	Attributes in ODL	136
4.2.4	Relationships in ODL	138
4.2.5	Inverse Relationships	139
4.2.6	Multiplicity of Relationships	140
4.2.7	Methods in ODL	141
4.2.8	Types in ODL	144
4.2.9	Exercises for Section 4.2	146
4.3	Additional ODL Concepts	147
4.3.1	Multiway Relationships in ODL	148
4.3.2	Subclasses in ODL	149
4.3.3	Multiple Inheritance in ODL	150
4.3.4	Extents	151
4.3.5	Declaring Keys in ODL	152
4.3.6	Exercises for Section 4.3	155
4.4	From ODL Designs to Relational Designs	155
4.4.1	From ODL Attributes to Relational Attributes	156
4.4.2	Nonatomic Attributes in Classes	157
4.4.3	Representing Set-Valued Attributes	158
4.4.4	Representing Other Type Constructors	160
4.4.5	Representing ODL Relationships	162
4.4.6	What If There Is No Key?	164
4.4.7	Exercises for Section 4.4	164
4.5	The Object-Relational Model	166
4.5.1	From Relations to Object-Relations	166
4.5.2	Nested Relations	167
4.5.3	References	169
4.5.4	Object-Oriented Versus Object-Relational	170
4.5.5	From ODL Designs to Object-Relational Designs	172
4.5.6	Exercises for Section 4.5	172
4.6	Semistructured Data	173
4.6.1	Motivation for the Semistructured-Data Model	173
4.6.2	Semistructured Data Representation	174
4.6.3	Information Integration Via Semistructured Data	175
4.6.4	Exercises for Section 4.6	177
4.7	XML and Its Data Model	178
4.7.1	Semantic Tags	178
4.7.2	Well-Formed XML	179
4.7.3	Document Type Definitions	180
4.7.4	Using a DTD	182
4.7.5	Attribute Lists	183
4.7.6	Exercises for Section 4.7	185
4.8	Summary of Chapter 4	186