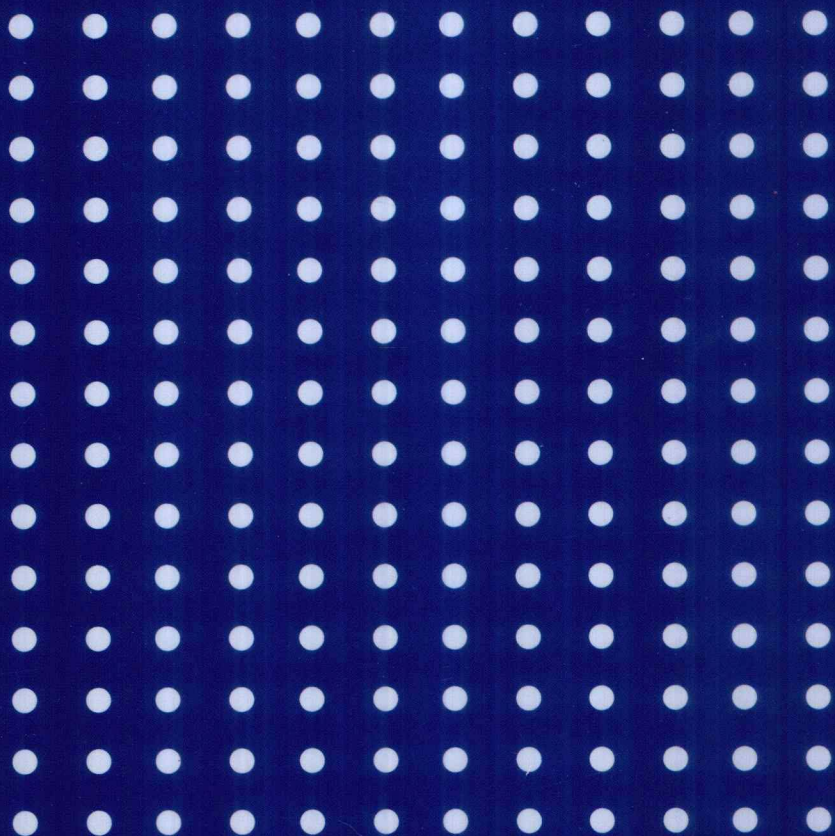


云南省“十二五”规划教材

重点大学计算机专业系列教材

XML及其相关技术

李 浩 主编
陆歌皓 刘宝龙 副主编



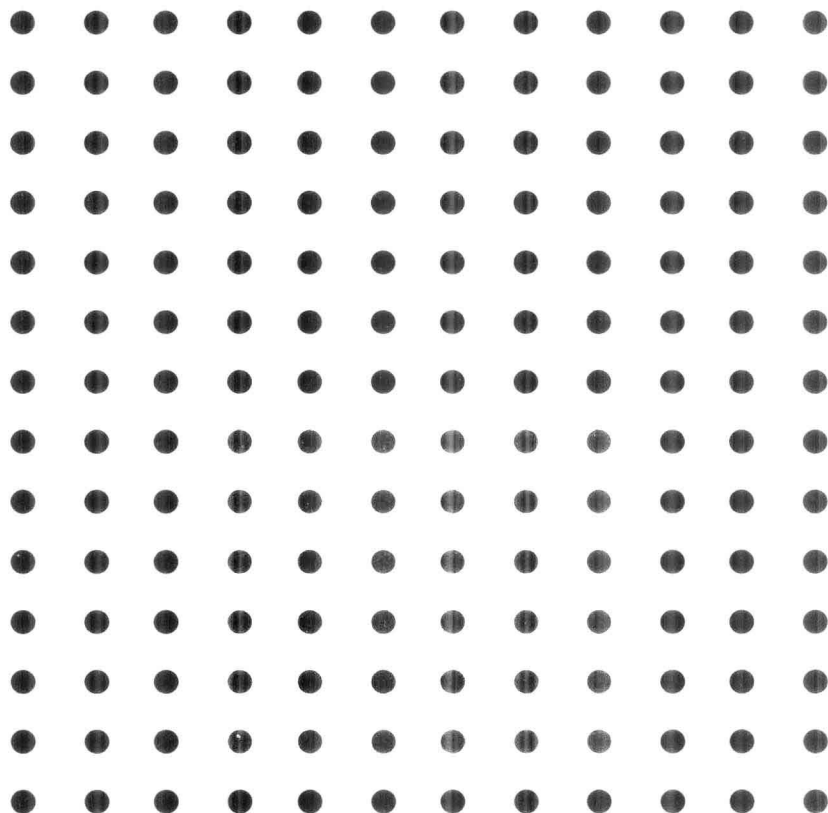
清华大学出版社

重点大学计算机专业系列教材

XML 及其相关技术

李浩 主编

陆歌皓 刘宝龙 副主编



清华大学出版社

北京

内 容 简 介

随着互联网的迅猛发展和广泛应用,人们通过互联网,从世界各地实时地接收和发送大量信息,但在信息交换的过程中存在着一个突出的问题,就是多种多样的数据格式,给信息的有效使用带来了障碍。人们期待着能够找到一种可以描述任何逻辑关系的数据格式来统一电子数据的存储。目前,能够担当此重任的就是 XML(eXtensible Markup Language,可扩展符号化语言)。可以说,XML 的出现给数据交换带来了一场革命,XML 成为下一代网络发展的基石。

本书共分为 14 章,全面介绍了 XML 从理论到实际应用的整个过程,除了常规的 XML 基础知识之外,还介绍了原生 XML 数据库、Web Service、SOA 等最新的构架技术及概念。本书的最后一章给出了一个完整的运用实例。

本书适合具有一定计算机基础知识的读者阅读,可作为计算机及相关专业本科 Web 设计或 XML 课程的教材,也可作为大专院校非计算机专业学生学习计算机基础的参考书和自学用书,还可供从事 Web 应用软件设计的科研人员参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

XML 及其相关技术/李浩主编. —北京:清华大学出版社,2012.3

(重点大学计算机专业系列教材)

ISBN 978-7-302-27095-9

I. ①X… II. ①李… III. ①可扩展语言,XML—程序设计—高等学校—教材 IV. ①TP312

中国版本图书馆 CIP 数据核字(2011)第 222692 号

责任编辑:梁 颖 薛 阳

封面设计:傅瑞学

责任校对:胡伟民

责任印制:何 芊

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:北京国马印刷厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:18

字 数:450 千字

版 次:2012 年 3 月第 1 版

印 次:2012 年 3 月第 1 次印刷

印 数:1~3000

定 价:29.00 元

产品编号:044214-01

出版说明

随着国家信息化步伐的加快和高等教育规模的扩大,社会对计算机专业人才的需求不仅体现在数量的增加上,而且体现在质量要求的提高上,培养具有研究和实践能力的高层次的计算机专业人才已成为许多重点大学计算机专业教育的主要目标。目前,我国共有 16 个国家重点学科、20 个博士点一级学科、28 个博士点二级学科集中在教育部部属重点大学,这些高校在计算机教学和科研方面具有一定优势,并且大多以国际著名大学计算机教育为参照系,具有系统完善的教学课程体系、教学实验体系、教学质量保证体系和人才培养评估体系等综合体系,形成了培养一流人才的教学和科研环境。

重点大学计算机学科的教学与科研氛围是培养一流计算机人才的基础,其中专业教材的使用和建设则是这种氛围的重要组成部分,一批具有学科方向特色优势的计算机专业教材作为各重点大学的重点建设项目成果得到肯定。为了展示和发扬各重点大学在计算机专业教育上的优势,特别是专业教材建设上的优势,同时配合各重点大学的计算机学科建设和专业课程教学需要,在教育部相关教学指导委员会专家的建议和各重点大学的大力支持下,清华大学出版社规划并出版本系列教材。本系列教材的建设旨在“汇聚学科精英、引领学科建设、培育专业英才”,同时以教材示范各重点大学的优秀教学理念、教学方法、教学手段和教学内容等。

本系列教材在规划过程中体现了如下一些基本组织原则和特点。

(1) 面向学科发展的前沿,适应当前社会对计算机专业高级人才的培养需求。教材内容以基本理论为基础,反映基本理论和原理的综合应用,重视实践和应用环节。

(2) 反映教学需要,促进教学发展。教材要能适应多样化的教学需要,正确把握教学内容和课程体系的改革方向。在选择教材内容和编写体系时注意体现素质教育、创新能力与实践能力的培养,为学生知识、能力、素质协调发展创造条件。

(3) 实施精品战略,突出重点,保证质量。规划教材建设的重点依然是专业基础课和专业主干课;特别注意选择并安排了一部分原来基础比较好的优秀教材或讲义修订再版,逐步形成精品教材;提倡并鼓励编写体现重点大学

计算机专业教学内容和课程体系改革成果的教材。

(4) 主张一纲多本,合理配套。专业基础课和专业主干课教材要配套,同一门课程可以有多本具有不同内容特点的教材。处理好教材统一性与多样化的关系;基本教材与辅助教材以及教学参考书的关系;文字教材与软件教材的关系,实现教材系列资源配套。

(5) 依靠专家,择优落实。在制订教材规划时要依靠各课程专家在调查研究本课程教材建设现状的基础上提出规划选题。在落实主编人选时,要引入竞争机制,通过申报、评审确定主编。书稿完成后要认真实行审稿程序,确保出书质量。

繁荣教材出版事业,提高教材质量的关键是教师。建立一支高水平的以老带新的教材编写队伍才能保证教材的编写质量,希望有志于教材建设的教师能够加入到我们的编写队伍中来。

教材编委会

序

XML 作为网络时代最重要的计算机语言之一,已经被广泛应用到网络数据交互的各个领域。正是因为有了 XML,才有了今天飞速发展的网络应用。由于 XML 本身的自描述性及平台无关性,使得它成为几乎所有现代网络交换的标准及平台。如果某一公司推出的系统或产品不支持 XML 及其相关标准,那么这些系统和产品也就会失去生命力。

由于 XML 技术自身的特点,使其几乎在所有的计算机应用领域中都扮演着重要角色。在网络服务领域,XML 有利于信息的表达和结构化组织,从而使数据搜索更有效。XML 可以使用 URL 别名使 Web 的维护更方便,也使 Web 的应用更稳定。在 EDI(Electronic Data Interchange)领域,传统的 EDI 标准缺乏灵活性和可扩展性;而使用 XML 编写的交换数据,应用程序则容易理解。在交换数据中,XML 可以表示商务数据及概念,并根据明确的商务规则来进行数据处理。在电子商务领域,XML 的丰富置标信息可以用来描述不同类型的单据,例如,信用证、保险单、索赔单,以及各种发票等。结构化的 XML 文档发送至 Web 的数据可以被加密,并且很容易附加上数字签名。在数据库领域,XML 文档可以定义数据结构,代替数据字典,用程序输出建库脚本。运用“元数据模型”技术,对数据源中不同格式的文件数据,按照预先定义的 XML 模板,以格式说明文档结构统一描述并提取数据并做进一步的处理,最后转换为 XML 格式输出。XML-数据库-网页或文档中的表格可以相互转换。在 Agent(智能体)领域,倘若送到 Agent 的是 XML 结构化数据,Agent 就能很容易地理解这些数据的含义及它拥有知识的关系。基于 XML 的数据交换对于解决 Agent 的交互性问题具有重要作用。在软件设计元素交换上,XML 也可以用来描述软件设计中的有关设计元素,如对象模型,甚至能描述最终设计出来的软件。另外,XML 及其相关技术使得软件的分发及更新在 Web 上更容易实现。

从技术上讲,XML 只是一种简单的信息描述语言。但从应用角度上说,XML 的价值就远不止是一种信息的表达工具。事实上,借助 XML,可以准确地表示几乎所有类型的数字化信息,可以清晰地解释信息的内涵和信息之间的关联,可以在最短的时间内准确定位需要的信息资源。

本书作者长期从事 XML 及其相关领域的研究和教学工作,承担了包括国家自然科学基金在内的大量研究和工程项目,拥有丰富的工程实践经验。全书内容丰富,行文严谨流畅,便于阅读。由于 XML 技术的实践性很强,本书力图做到理论联系实际,在计算机教学及软件工程实践中发挥引导作用。相信这本书的出版将对计算机相关专业教学及业界的人才培养起到一定的促进作用。

李 彤

二级教授,博士生导师

2011 年 8 月

前言

云南省教育厅为促进教学内容与课程体系改革,加强精品课程、特色专业建设,全面提高教材建设的水平,开展了云南省“十二五规划”教材建设工作。《XML 及其相关技术》就是受到资助出版的“规划”教材之一。本书是编者在云南省省级教学质量工程基础之上完成的一个教学总结。

XML 技术自出现以来发展非常迅速,在许多领域内得到广泛的支持而且有着广阔的应用前景。例如电子数据交换、电子商务等更是将 XML 作为一种基础性、支柱性的技术。为了满足软件工程及其相关人员的需求,我们编写了本书。编者着眼于编程人员在牢固掌握相关 XML 知识的基础上,结合实际运用,使得本书可以在教学和工程开发中发挥积极的作用。

本书由云南大学软件学院及西安工业大学计算机科学与技术学院的老师联合编写。在编写过程中,编者本着客观、真实、完整、简练的原则,进行了有针对性的摘选和撰写,在内容排列顺序上参考了近年来出版及发表的相关材料以及最新的国际标准。本书共有 14 章,从理论和实践出发,结合软件开发实际,以 XML 理论为主线,辅以开发工作中需要的基础知识、常用软件和开发实例作为充实内容。

本书在风格上力求文字精练,图表丰富,脉络清晰,版式明快,便于在教学过程中教师对教学内容灵活选取和学生的理解。在策划时特别设计了一些特色板块,比如,章前导读、学习重点、本章总结以及习题与实训。

本书可作为计算机人员培训和考试的学习用书,也可供软件工程、计算机科学等专业的学生作为教材,以及从事软件开发工程人员参考。

云南大学软件学院郭威、章祯超、李晓霖、赵倩、孙冕、朱桀骜、姚志斌、纳霞、宋天龙等参与了本书的资料收集及部分编写工作,李婷婷参与了本书的整理、编辑及规范工作。本书的编写得到了云南大学软件学院李彤教授,姚绍文教授以及英国 Huddersfield 大学 Dr.Joan Lu 的鼓励和支持,他们提出了许多宝贵的修改意见。在此,我们一并对给予本书指导和帮助的所有同志表示深深的谢意。

我们将根据实际工作过程中技术的变化和新标准、规范的出台,不断地修订和完善本书。由于内容较多、时间仓促,我们所掌握的材料和自身水平所限,书中难免存在不足和缺憾,如有疏漏、不当之处,敬请读者批评指正。

编 者

2011 年 8 月 10 日

目录

第 1 章 XML 的发展背景及原因	1
1.1 XML 诞生的背景	1
1.1.1 SGML 介绍	2
1.1.2 HTML 的出现	2
1.1.3 XML 的诞生	2
1.2 XML 的特点及技术	3
1.2.1 XML 可以作为电子数据交换的统一格式	3
1.2.2 XML 的数据描述特点	4
1.2.3 XML 的不足之处	5
1.2.4 XML 的主要相关技术分类	5
1.3 XML 的发展前景	7
1.4 HTML 介绍	7
1.4.1 什么是 HTML	7
1.4.2 HTML 的作用及主要构成	8
1.4.3 HTML 工作协议基础	10
1.5 XML 与 HTML 的主要差异	11
本章小结	11
习题与实训	11
第 2 章 XML 的基本组成	12
2.1 创建一个 XML 文档	12
2.2 XML 文档的基本构成元素	13
2.2.1 关于文件头的声明	13
2.2.2 XML 文件的元素	14
2.2.3 元素内容	14

2.2.4 元素的属性	15
2.3 XML 文档的语法规则及限制	15
本章小结	16
习题与实训	16
第3章 XML 的命名空间	17
3.1 为什么需要命名空间	18
3.2 命名空间的语法	18
3.2.1 命名空间的定义	18
3.2.2 统一资源标识符	19
3.2.3 前缀与本地部分、限定名称	19
3.2.4 依赖 URI, 而不依赖前缀	20
3.2.5 默认命名空间	20
3.2.6 通用前缀	21
3.3 命名空间的作用域	21
3.4 命名空间和验证文件	22
3.4.1 命名空间和 DTD	22
3.4.2 命名空间和 XML Schema	23
3.5 解析器对命名空间的支持	24
本章小结	25
习题与实训	25
第4章 XML 文档类型定义——DTD	26
4.1 什么是 DTD	26
4.2 为什么要使用 DTD	27
4.3 XML DTD 声明	27
4.3.1 内部 DTD	27
4.3.2 外部 DTD	28
4.4 DTD 的语法	29
4.4.1 元素声明	29
4.4.2 属性声明	31
4.4.3 实体声明	33
4.5 实例	34
本章小结	35
习题与实训	35
第5章 XML Schema	36
5.1 XML Schema 基础知识	36

5.1.1	什么是 XML Schema	36
5.1.2	为什么使用 Schema	37
5.1.3	如何使用 Schema	38
5.2	XSD 的数据类型	39
5.2.1	简单数据类型	39
5.2.2	复杂数据类型	40
5.3	XSD 的语法	41
5.3.1	XSD 的元素声明	41
5.3.2	XSD 的属性声明	42
5.3.3	XSD 复合类型指示器	43
5.3.4	在一个 XML 模式中使用另一个模式	45
5.4	XML Schema 实例	46
	本章小结	52
	习题与实训	52

第 6 章 XSL 转换 53

6.1	XSL 概述	53
6.1.1	链接 XSL 到 XML	54
6.1.2	XML 文档结构树	55
6.1.3	XSLT 样式单的创建和输出	55
6.1.4	在 XML 文档中使用多个样式单	57
6.2	XSLT 模板	57
6.2.1	模板的定义元素	57
6.2.2	关于 XSL 样式单中的 HTML 置标	58
6.3	节点的访问	60
6.3.1	访问单个节点	60
6.3.2	访问多个节点	60
6.3.3	XML 结构树中的节点类型和节点值	61
6.4	节点的选择方式	61
6.4.1	通用匹配符	62
6.4.2	直接使用元素名	64
6.4.3	路径选择方式	64
6.4.4	选择元素属性	66
6.4.5	为模板选择多个元素	67
6.4.6	为选择的元素添加条件	67
6.4.7	使用节点类型函数选择节点	70
6.5	对输出结果的排序	70

6.6 对输出节点的选择	71
本章小结	73
习题与实训	74
第 7 章 层叠样式表	75
7.1 CSS 简介	75
7.1.1 初识 CSS	75
7.1.2 为什么要使用 CSS	77
7.1.3 CSS 规范	78
7.1.4 HTML 现有技术对样式表的支持	79
7.2 基本 CSS 语法	80
7.2.1 CSS 使用	80
7.2.2 CSS 基础语法	81
7.2.3 选择符	81
7.2.4 CSS 属性	83
7.3 CSS 盒模型	88
7.3.1 基本 BOX 概念	88
7.3.2 BOX 模型属性	90
7.3.3 BOX 的类型	92
7.3.4 可视性	93
7.4 CSS 高级应用	93
7.4.1 CSS 打印支持	93
7.4.2 声音层叠样式表	95
7.4.3 CSS 滤镜与渐变	97
7.5 CSS 3.0	101
7.5.1 边框	102
7.5.2 背景	102
7.5.3 文字效果	102
7.5.4 颜色	103
本章小结	103
习题与实训	103
第 8 章 XPath、XLink、XPointer、XInclude 和 XBase	105
8.1 XML Path Language	105
8.1.1 节点	106
8.1.2 定位路径	107
8.1.3 节点集操作符和函数	110

8.2 XML 指针语言	112
8.2.1 绝对位置	115
8.2.2 相对位置	116
8.2.3 其他位置	119
8.3 XML 链接语言	120
8.3.1 简单 XLink	120
8.3.2 扩展 XLink	122
8.4 XML Inclusions	124
8.5 XML Base	124
8.5.1 与 RFC 2396 的关系	125
8.5.2 基准 URI 信息的粒度	125
8.5.3 URI 与基准 URI 的匹配	126
本章小结	126
习题与实训	126
第 9 章 XML 文档处理	128
9.1 XML DOM	128
9.1.1 DOM 接口中的主要对象	130
9.1.2 XML 文档的加载	132
9.1.3 XML 文档的遍历	133
9.1.4 XML 元素的添加、删除和修改	134
9.2 SAX	135
9.2.1 SAX 简介	135
9.2.2 SAX 的常用接口介绍	136
9.2.3 创建 SAX 解析器	137
9.3 JDOM	139
9.3.1 JDOM 包概览	140
9.3.2 JDOM 类说明	140
9.3.3 JDOM 的主要使用方法	141
9.3.4 JDOM 应用	144
9.4 DOM4J	148
9.4.1 DOM4J 概念	148
9.4.2 DOM4J 使用简介 1	150
9.4.3 DOM4J 使用简介 2	153
本章小结	158
习题与实训	158

第 10 章 资源描述框架	159
10.1 RDF 简介	159
10.2 RDF 规则	162
10.2.1 RDF 资源、属性和属性值	162
10.2.2 RDF 陈述	163
10.3 RDF 实例	163
10.4 RDF 的主要元素及容器	165
10.5 RDF Schema	168
本章小结	170
习题与实训	171
第 11 章 Native XML Database	172
11.1 Native XML Database 简介	172
11.2 Native XML Database 的优势	174
11.3 了解 eXist	174
11.3.1 序言	174
11.3.2 预备知识	175
11.3.3 安装 eXist	176
11.3.4 编写与 eXist 交互的程序	176
11.3.5 插入文档	177
11.3.6 删除文档	178
11.3.7 更新文档	178
11.4 了解 Apache Xindice	179
11.4.1 Xindice 简介	179
11.4.2 现状与前景	179
11.4.3 Xindice 主要功能	180
11.4.4 Xindice 数据库管理指南	180
11.4.5 开发人员指南	185
本章小结	194
习题与实训	194
第 12 章 Web Service	195
12.1 Web Service 简介	195
12.2 SOAP 标签	196
12.2.1 SOAP 简介	196
12.2.2 SOAP 语法	196

12.2.3	SOAP 元素	196
12.2.4	SOAP 实例	197
12.3	RESTful 概念	200
12.4	Web Service 开发实例	200
12.4.1	开发工具	200
12.4.2	开发环境	201
12.4.3	Web Service 实例——图书管理系统	201
12.5	RESTful 架构例子	219
12.5.1	图书管理系统——增加功能	219
12.5.2	图书管理系统——修改功能	220
12.5.3	图书管理系统——删除功能	221
12.5.4	图书管理系统——查询功能	222
	本章小结	223
	习题与实训	223
第 13 章	SOA	224
13.1	SOA 的基本概念及思想	224
13.1.1	概念	224
13.1.2	SOA 的发展及其思想	226
13.2	SOA 和企业业务逻辑的绑定	228
13.2.1	步骤 1: 剖析业务	228
13.2.2	步骤 2: 决定哪些业务是企业的独特之处, 哪些是一般性的 共性业务	231
13.2.3	步骤 3: 决定如何投资	232
13.3	SOA 主要技术	233
13.3.1	什么是企业服务总线	233
13.3.2	ESB 的适用场景及要素	234
13.3.3	实践	235
13.4	通过 Web Service 实现 SOA	236
13.4.1	SOA 相关技术标准分类	237
13.4.2	SOA 相关技术标准比较说明	237
13.4.3	总结	241
	本章小结	241
	习题与实训	241
第 14 章	XML 及其相关技术应用实例	242
14.1	应用实例背景	242

14.1.1	新闻发布系统概述	242
14.1.2	新闻发布系统功能展示	242
14.2	功能需求	249
14.2.1	系统需求	249
14.2.2	权限矩阵	250
14.2.3	业务流程	250
14.2.4	新闻状态变化分析	250
14.3	开发指南	251
14.3.1	开发流程	251
14.3.2	系统架构的设计	252
14.3.3	开发平台	253
14.3.4	数据库的设计	253
14.3.5	XML 文档设计	253
14.3.6	包截图	253
14.3.7	核心类的设计	255
14.3.8	系统集成	255
14.4	编码实现	256
14.4.1	XML 的核心代码	256
14.4.2	其他业务类代码	261
14.5	相关技术集成	264
14.5.1	表示层的其他集成路线	264
14.5.2	持久层的其他集成路线	264
14.5.3	与热门框架的集成应用	265
	本章小结	265
	习题与实训	265
	参考文献	267