

XQuery

Search Across a Variety of XML Data

XQuery

权威指南



O'REILLY®

Priscilla Walmsley 著
王银辉 译



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

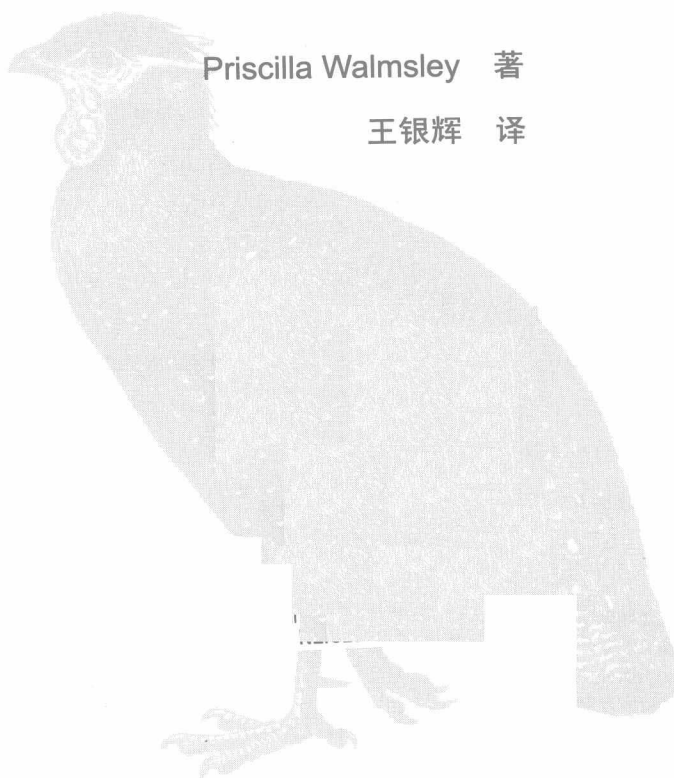
O'REILLY®

XQuery 权威指南

XQuery

Priscilla Walmsley 著

王银辉 译



電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

XQuery 是专门针对 XML 的查询语言,其作用类似于关系数据库的 SQL。

本书比较深入全面地介绍了 XQuery 相关知识,包括 XQuery 的概述和快速指南,如何使用 XQuery 编写简单和复杂的查询,如何对 XML 数据过滤、排序和分组,还讲述了 FLWOR 表达式、XPath 及提取、组合数据的 XQuery 工具。本书的附录提供了 XQuery 的函数、内置类型及错误信息,因此可以作为 XQuery 编程指南。

本书的特点是用简洁的语言深入浅出地介绍了 XQuery 方方面面的知识,在讲述有些内容时把 XQuery 与其他广为人知的技术如 SQL 等相对比,使得读者可以很快上手。阅读本书只须要了解一些基础的 XML 背景知识。本书除了用于指导编程外,还可用作提高与优化 XML 数据检索性能的参考书。

本书不仅适合那些须要操作大量 XML 数据的程序员(不管是使用 C#还是 Java 等语言的程序员都适合阅读本书),还适合管理 XML 数据库的管理员,以及那些非 IT 行业但须要频繁使用 XML 数据的业务人员。

这是一本适合作为 XQuery 学习教材和参考指南的书。

0-596-00634-9 XQuery © 2007 by O'Reilly Media, Inc. Simplified Chinese edition, jointly published by O'Reilly Media, Inc. and Publishing House of Electronics Industry, 2008. Authorized translation of the English edition, 2008 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

本书中文简体版专有出版权由 O'Reilly Media, Inc. 授予电子工业出版社,未经许可,不得以任何方式复制或抄袭本书的任何部分。

版权贸易合同登记号 图字: 01-2009-0876

图书在版编目(CIP)数据

XQuery 权威指南 / (美) 普里西拉, (美) 沃姆斯利著; 王银辉译. —北京: 电子工业出版社, 2009.4

书名原文: XQuery

ISBN 978-7-121-08434-8

I. X… II. ①普…②沃…③王… III. 可扩充语言, XML—十进制法—程序设计—指南 IV. TP312-62 TP393.092-62

中国版本图书馆 CIP 数据核字 (2009) 第 030103 号

责任编辑: 陈元玉

项目管理: 梁 晶

印 刷: 北京市天竺颖华印刷厂

装 订: 三河市鑫金马印装有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×980 1/16 印张: 32.5 字数: 550 千字

印 次: 2009 年 4 月第 1 次印刷

定 价: 88.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

O'Reilly Media, Inc.介绍

为了满足读者对网络 and 软件技术知识的迫切需求，世界著名计算机图书出版机构O'Reilly Media, Inc.授权电子工业出版社，翻译出版一批该公司久负盛名的英文经典技术专著。

O'Reilly Media, Inc.是世界上在Unix、X、Internet和其他开放系统图书领域具有领导地位的出版公司，同时也是在线出版的先锋。

从最畅销的《The Whole Internet User's Guide & Catalog》(被纽约公共图书馆评为20世纪最重要的50本书之一)到GNN(最早的Internet 门户和商业网站)，再到WebSite(第一个桌面 PC 的Web服务器软件)，O'Reilly Media, Inc.一直处于Internet发展的最前沿。

许多书店的反馈表明，O'Reilly Media, Inc.是最稳定的计算机图书出版商——每一本书都一版再版。与大多数计算机图书出版商相比，O'Reilly Media, Inc.具有深厚的计算机专业背景，这使得O'Reilly Media, Inc.形成了一个非常不同于其他出版商的出版方针。O'Reilly Media, Inc.所有的编辑人员以前都是程序员，或者是顶尖级的技术专家。O'Reilly Media, Inc.还有许多固定的作者群体——他们本身是相关领域的技术专家、咨询专家，而现在编写著作，O'Reilly Media, Inc.依靠他们及时地推出图书。因为O'Reilly Media, Inc.紧密地与计算机业界联系着，所以O'Reilly Media, Inc.知道市场上真正需要什么图书。

译者序

XQuery 是什么？译者认为它是个好东西！您能想象在没有 SQL 的情况下操作关系数据库吗？既然 SQL 对关系数据库来说是必不可少的一部分，那么对如日中天的 XML 来说，XQuery 也是至关重要的一部分。简单地说，XQuery 就类似关系数据库的 SQL，用于操作 XML 数据。

如果您的程序须要频繁地与 XML 数据打交道，或者须要处理大量 XML 数据，那么强烈推荐丢掉传统的方式转而使用 XQuery。

目前，主流数据库 MS SQL Server、Oracle、db2 等都支持 XQuery。

XML 相关的书籍与资料已经较为普及。由于 XML 的方便性，越来越多的程序员把数据以 XML 形式存储。随着技术的进一步发展，以前按关系数据存放大量数据的形式可能逐步变成以 XML 形式存储，应用程序须要频繁地与 XML 打交道，存取 XML 数据。译者早年曾经在 Linux 下用 C 语言操作 XML 数据，深感 XML 的检索与存取颇费周折，效率非常低。幸好现在的程序员可以不用把大量时间耗费在这种效率低下的工作上了。与关系数据库的 SQL 对应，XML 也演化出了自己的检索语言——XQuery，使得操作 XML 数据像关系数据库里面使用 SQL 一样快捷方便。市面上已经有比较成熟的 XQuery 开源系统，如 Saxon 等，可以使程序员以较低成本和较高效率完成大量 XML 数据的处理。

目前国内还没有一本完整的 XQuery 中文书籍，相关中文资料也很少，这限制了 XQuery 的推广与普及。大多数相关程序员还在重复译者当年的劳动。有鉴于此，译者翻译了这本在国外受欢迎的 XQuery 英文书，希望为广大相关人员提供帮助。

由于附录提供了 XQuery 的内置函数、类型和错误提示，因此本书不仅可以作为供普通读者了解 XQuery 的综述性书籍，也可以作为专业程序员的 XQuery 手册。

由于译者水平有限，错误在所难免，敬请读者指正，以期提高翻译水平，对此深表谢意。

王银辉

2008 年 12 月于重庆

前言

Preface

本书涵盖了 2007 年 1 月 W3C 制订的 XQuery 1.0 规范的全部内容。除此之外，本书还介绍了与编写 XML 查询语句相关的背景知识，包括命名空间、schema、内置类型和正则表达式。

本书适合具有一定 XML 基础知识并须要用到 XML 查询的开发人员，但不须要精通 XML 相关技术。通过仔细阅读，本书可作为学习 XQuery 的指南；通过使用索引和附录，本书亦可作为 XQuery 的参考书。

本书内容

Contents of This Book

本书由 6 部分组成。

1. 第 1 章和第 2 章给出了 XQuery 的全面概述。
2. 第 3 章至第 9 章将教会读者如何在不被类型、命名空间和 schema 的细节困扰的情况下，写出老练的查询语句。
3. 第 10 章至第 15 章向希望利用模块、命名空间、类型和 schema 的用户介绍一些高级概念。
4. 第 16 章至第 23 章提供了使用数据类型的指导，例如数字、字符、日期、URI 和处理指令。
5. 第 24 章和第 25 章介绍了 XQuery 与 SQL、XSLT 之间的关系。
6. 附录 A、B 和 C 提供了按字母顺序排列的完整的内置函数、类型和出错信息的参考。

阅读语法图

Reading the Syntax Diagrams

本书给想更直观了解 XQuery 表达式语法的读者提供了语法图。每一个语法图都配有说明文字和例子。图 P-1 以 schema 导入为例，说明了一个语法图的组成部分。

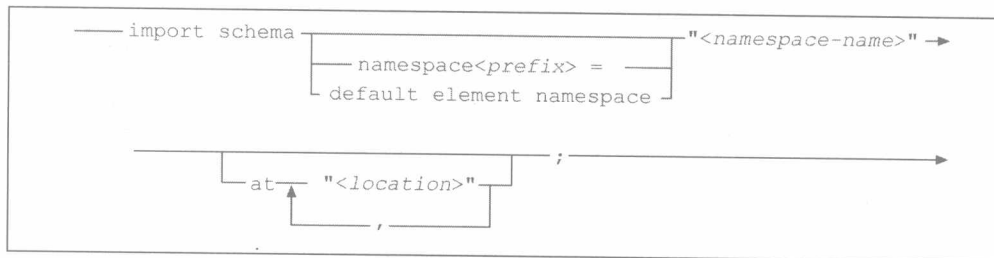


图 P-1: 语法图示例

语法图的解释规则如下。

- 图中等宽字体的部分表示字面值。在图 P-1 中, `import schema` 和 `at` 应该原样出现在查询中。
- 出现在语法图中的引号也必须出现在查询中。如图 P-1 所示, `<namespace-name>` 必须加上引号而 `<prefix>` 不能加引号。XQuery 中可以使用单引号或双引号, 但是为了简便, 图中只包含双引号。
- 在用户可以对其设置值的地方, 如名称、值的描述名称等, 使用等宽斜体字并且加上了尖括号。如图 P-1 所示, 可以将 `<namespace-name>`、`<prefix>` 和 `<location>` 中的斜体字替换成自己的值。
- 图中的平行线表示可选项。如图 P-1 所示, 可以指定一个命名空间前缀或使用默认元素命名空间。
- 表达式的可选部分由主箭头旁边的箭头表示。在图 P-1 中, 不必包含 `namespace` `<prefix>` 或 `default element namespace` 关键词。
- 表达式的重复部分由返回起点的箭头表示。如图 P-1 所示, 可以指定多个 `<location>` (在逗号之后) 作为 `at` 子句的一部分。

- 出现在语法图中的引号也必须出现在查询中。如图 P-1 所示, `<namespace-name>` 必须加上引号而 `<prefix>` 不能加引号。XQuery 中可以使用单引号或双引号, 但是为了简便, 图中只包含双引号。

- 在用户可以对其设置值的地方，如名称、值的描述名称等，使用等宽斜体字并且加上了尖括号。如图 P-1 所示，可以将<namespace-name>、<prefix>和<location>中的斜体字替换成自己的值。

- 图中的平行线表示可选项。如图 P-1 所示，可以指定一个命名空间前缀或使用默认元素命名空间。

- 表达式的可选部分由主箭头旁边的箭头表示。在图 P-1 中，不必包含 namespace `<prefix>=`或 default element namespace 关键词。

- 表达式的重复部分由返回起点的箭头表示。如图 P-1 所示，可以指定多个<location>（在逗号之后）作为 at 子句的一部分。

排版约定

Conventions Used in This Book

等宽字体 (Constant width):

- 示例代码和代码片段。
- 任何可能出现在 XML 文档中的东西，包括元素名称、标签、属性值和处理说明。
- 任何可能出现在程序中的东西，包括关键字、操作符、方法名、类名和文字。

- 任何可能出现在 XML 文档中的东西，包括元素名称、标签、属性值和处理说明。

- 任何可能出现在程序中的东西，包括关键字、操作符、方法名、类名和文字。

等宽粗体 (**Constant width bold**):

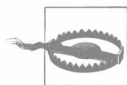
- 对某些示例代码和代码片段中的内容进行强调。

斜体 (*Italic*):

- 首次定义的新术语。
- 正文主体中的重点。
- 路径名、文件名和程序名。
- 主机名和域名。



这个图标表示一个提示、建议，或者一般性的注解。



这个图标表示一个警告或警示。



这个图标表示兼容性问题可能会引起不可预料的结果。

重要的代码片段、完整程序和文档通常都单独成段，就像这样：

```
for $prod as element(*,ProductType) in doc("catalog.xml")/catalog/*  
order by $prod/name  
return $prod/name
```

关于示例代码

Using Code Examples

本书旨在帮助读者完成手头工作。通常情况下，读者可以在自己的程序和文档中使用本书中的代码。除非原样引用大量的代码，否则无须征得我们的许可。例如，在编写程序时引用本书的若干代码片段是不用许可的。销售或发行 O'Reilly 图书的示例光盘则须要得到许可。引用本书内容或示例代码来回答问题不须要获得许可。而将书中的大量示例代码加入到读者的产品文档中则须要许可。

如果读者在引用时注明出处，我们将不胜感激，但这并非是必须的。引用通常包含了标题、作者、出版商，以及 ISBN 号。例如：“*XQuery* by Priscilla Walmsley. Copyright 2007 Priscilla Walmsley, 978-0-596-00634-1.”

如果读者发现自己对示例代码的使用有失公允或违反了上述条款，敬请通过 permission@oreilly.com 与我们联系。

常用函数

Useful Functions

本书包含一系列标记为“常用函数”的示例代码。之所以把这些示例与常规示例分开，是因为这些示例很有可能对读者编写查询语句产生直接的帮助。这些常用函数涵盖了如 `substring-after-last` 和 `replace-first` 这样的字符函数，以及修改元素和属性的函数，例如 `add-attribute`。

本书中包含的常用函数只是被称为 FunctX XQuery 函数库中大量的 XQuery 函数的一部分，读者可以从 <http://www.xqueryfunctions.com> 上得到这些代码。这个函数库包含大量可以按类别搜索或浏览的各种各样的可重用 XQuery 函数，同样也包含详细的说明和示例函数调用。

联系方式

We'd Like to Hear from You

如果你想就本书发表评论或有任何疑问，敬请联系出版社：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707-829-0104 (fax)

奥莱理软件（北京）有限公司
北京市 西城区 西直门南大街 2 号 成铭大厦 C 座 807 室
邮政编码：100055
网页：<http://www.oreilly.com.cn>
E-mail：info@mail.oreilly.com.cn

北京博文视点资讯有限公司（武汉分部）
湖北省 武汉市 洪山区 吴家湾 邮科院路特 1 号 湖北信息产业科技大厦 1402 室
邮政编码：430074
电话：(027) 87690813 传真：(027) 87690595
网页：<http://bv.csdn.net>

读者服务信箱：
reader@broadview.com.cn（读者信箱）
bvtougao@gmail.com（投稿信箱）

我们为本书开设了一个站点，在那里放置了勘误表、示例代码和其他附加信息。读者可以访问以下网址：

<http://www.oreilly.com/catalog/9780596006341>

<http://www.oreilly.com.cn/book.php?bn=978-7-121-08434-8> (中文版)

有关本书的评论或询问技术问题，请发送电子邮件到：

bookquestions@oreilly.com

关于我们的书籍、会议、资源中心和 O'Reilly 网络的更多信息请登录我们的网址：

<http://www.oreilly.com>

致谢

Acknowledgments

我深深地感激 Michael Kay，不仅因为他详细地审查了本书，还因为他杰出的 Saxon——XQuery 的实现（译注 1）。没有 Saxon，我不可能可靠地测试这些示例代码。

Ron Bourret、Bob DuCharme、Tim Finney、Ashok Malhotra、Darin McBeath、Peter Meggitt、Shannon Shiflett 和 Bruno J. Walmsley（我的父亲）提供给我很多关于本书非常有用的意见和帮助。

没有 Simon St.Laurent 就不可能有这个项目，他给我提供编辑上的指导并且支持由 O'Reilly 出版本书。谢谢你，Simon！

最后，我要感谢我的搭档，我的好朋友，Doug。感谢他在我忙于本书写作时对我自始至终的支持和鼓励。

译注 1：本文中的 XQuery 实现指的是一种 XQuery 实现工具，如某种 XQuery 产品，读者阅读后面的内容时会注意到此概念。

作者简介

Priscilla Walmsley 多年深入从事 XQuery 和 XML Schema 相关工作。她是 W3C XML Schema 工作组 1999 年到 2004 年的成员。此外，她还编写了广受好评的《Definitive XML Schema》（Prentice Hall 出版社）（译者注：中文名《XML 模式权威教程》）一书。最近 Priscilla 担任 Datypic (www.datypic.com) 的常务董事，在那里她专注于 XML 和面向服务架构 (SOA) 相关的咨询和培训。

封面介绍

封面的动物是一只红胸角雉，野鸡家族中的一员，5 种角雉物种之一。这种鸟有时被称为红色角雉，产自喜马拉雅山。它的两个名字由雄性个体的两种外观而得来——眼睛上面的凸出物，看起来像角，以及鲜红色的羽毛。羽毛和角在求爱的时候尤其显著。

角雉以昆虫、树叶、嫩芽及种子为食，只有一个配偶。虽然雌性负责孵化，答辩雄性也会帮助照顾幼鸟。大部分角雉适合饲养，能较好地适应低温气候并且性格温顺。

由于栖息地的破坏，5 种角雉物种中的 4 种处于濒危灭绝状态。与大部分家禽不同，角雉分布在从海拔 925 米到 3 650 米的高海拔地带。冬季它们出没于松树灌木丛中，但在交配季节它们进入森林深处。找寻较高的枝条，雄性角雉建立自己的领地，每隔 5 分钟发出求偶呼唤。这种呼唤声有时被描述成类似鹅或小羊的声音，一英里之外都可以听到这种声音。

目录

Table of Contents

前言	1
第 1 章 XQuery 介绍	1
1.1 什么是 XQuery	1
1.2 转向 XQuery	3
1.3 路径表达式	5
1.4 FLWOR 表达式	6
1.5 添加 XML 元素和属性	7
1.6 函数	10
1.7 连接	10
1.8 求和与分组	11
第 2 章 XQuery 基础	12
2.1 XQuery 语言的设计	12
2.2 XQuery 与其他标准的关系	13
2.3 处理查询	15
2.4 XQuery 数据模型	17
2.5 类型	24
2.6 命名空间	24
第 3 章 表达式: XQuery 的组成部分	26
3.1 表达式种类	26
3.2 关键字和名称	27
3.3 查询中的空白符	27
3.4 字面值	28
3.5 变量	28
3.6 函数调用	29

3.7	注释.....	29
3.8	计算顺序和括号.....	30
3.9	比较表达式.....	30
3.10	条件 (if-then-else) 表达式.....	35
3.11	逻辑表达式.....	37
第 4 章	使用路径定位输入文件.....	39
4.1	路径表达式.....	39
4.2	谓词.....	46
4.3	动态路径.....	52
4.4	输入文件.....	52
4.5	深入了解上下文.....	55
第 5 章	在结果中增加元素和属性.....	57
5.1	包含输入文件中的元素和属性.....	57
5.2	直接元素构造器.....	58
5.3	计算构造器.....	68
第 6 章	使用 FLWOR 表达式查询和连接.....	72
6.1	使用路径表达式查询.....	72
6.2	FLWOR 表达式.....	72
6.3	量化表达式.....	79
6.4	选择不同的值.....	81
6.5	连接.....	81
第 7 章	排序和分组.....	85
7.1	XQuery 中的排序.....	85
7.2	分组.....	93
7.3	对值求聚合.....	94
第 8 章	函数.....	99
8.1	内置函数与用户自定义函数.....	99
8.2	调用函数.....	99
8.3	用户自定义函数.....	103
第 9 章	高级查询.....	110
9.1	复制并修改输入文件.....	110
9.2	使用位置和序列编号.....	115
9.3	合并结果.....	118
9.4	使用中间 XML 文档.....	119

第 10 章	命名空间与 XQuery.....	123
10.1	命名空间	123
10.2	命名空间和 XQuery	127
10.3	查询中的命名空间声明	128
10.4	在结果中控制命名空间声明	135
第 11 章	深入了解类型	141
11.1	XQuery 类型系统	141
11.2	内置类型	143
11.3	类型、节点和原子值	145
11.4	XQuery 中的类型检查	146
11.5	自动类型转换	147
11.6	序列类型	151
11.7	构造器和转换	155
第 12 章	查询、序言和模块	160
12.1	查询的结构：序言和主体	160
12.2	集中使用多个模块中的查询	163
12.3	变量声明	166
12.4	声明外部函数	168
第 13 章	在 XQuery 中使用 Schema	170
13.1	什么是 Schema	170
13.2	为什么在查询中使用 Schema	171
13.3	W3C XML Schema:概述	172
13.4	作用域内的 Schema 定义	175
13.5	Schema 校验和类型分配	178
13.6	序列类型和 Schema	183
第 14 章	静态类型	185
14.1	什么是静态类型	185
14.2	typeswitch 表达式	187
14.3	treat 表达式	189
14.4	类型声明	190
14.5	zero-or-one, one-or-more 和 exactly-one 函数	192
第 15 章	查询设计的原则	193
15.1	查询设计的目标	193
15.2	清晰化	193
15.3	模块化	196

15.4	健壮性	196
15.5	错误处理	199
15.6	性能	201
第 16 章	操作数值	204
16.1	数值类型	204
16.2	创建数值	205
16.3	比较数值	206
16.4	算术操作符	207
16.5	数值函数	211
第 17 章	操作字符串	213
17.1	xs:string 类型	213
17.2	创建字符串	213
17.3	比较字符串	214
17.4	子字符串	216
17.5	得到字符串的长度	217
17.6	字符串连接和拆分	218
17.7	操作字符串	220
17.8	空白符和字符串	222
17.9	国际化考虑	223
第 18 章	正则式	226
18.1	正则式的结构	226
18.2	表示单个字符	228
18.3	表示任意字符	229
18.4	表示字符的组	230
18.5	字符类表达式	233
18.6	勉强模式	235
18.7	锚	236
18.8	向后引用	237
18.9	使用标记	238
18.10	使用带替换变量的子表达式	239
第 19 章	操作日期、时间和时间段	242
19.1	日期和时间类型	242
19.2	时间段类型	246
19.3	提取日期、时间和时间段的组成部分	248
19.4	使用算术操作符处理日期、时间和时间段	249
19.5	日期组成部分类型	252

第 20 章	操作限定名, URI 和 ID.....	254
20.1	操作限定名.....	254
20.2	操作 URI.....	259
20.3	操作 ID.....	264
第 21 章	操作其他 XML 组件.....	267
21.1	XML 组件.....	267
21.2	处理指令.....	269
21.3	文档.....	272
21.4	文本节点.....	274
21.5	XML 实体和字符引用.....	278
21.6	CDATA 段.....	280
第 22 章	其他 XQuery 相关的标准.....	282
22.1	序列化.....	282
22.2	XQueryX.....	284
22.3	XQuery 修改功能.....	285
22.4	全文检索.....	285
22.5	用于 Java (XQJ)的 XQuery API.....	287
第 23 章	实现工具特有的特征.....	289
23.1	一致性.....	289
23.2	XML 版本支持.....	290
23.3	设置查询上下文.....	290
23.4	选项声明和扩展表达式.....	291
23.5	指定序列化参数.....	293
第 24 章	SQL 用户的 XQuery.....	294
24.1	关系模型与 XML 数据模型.....	294
24.2	比较 SQL 语法和 XQuery 语法.....	296
24.3	结合 SQL 与 XQuery.....	303
24.4	SQL/XML.....	306
第 25 章	XSLT 用户的 XQuery.....	307
25.1	XQuery 和 XPath.....	307
25.2	XQuery 与 XSLT.....	307
25.3	XQuery 1.0/XPath 2.0 和 XPath 1.0 的差别.....	314

附录 A 内置函数参考	319
附录 B 内置类型	411
附录 C 错误汇总	440
索引	465