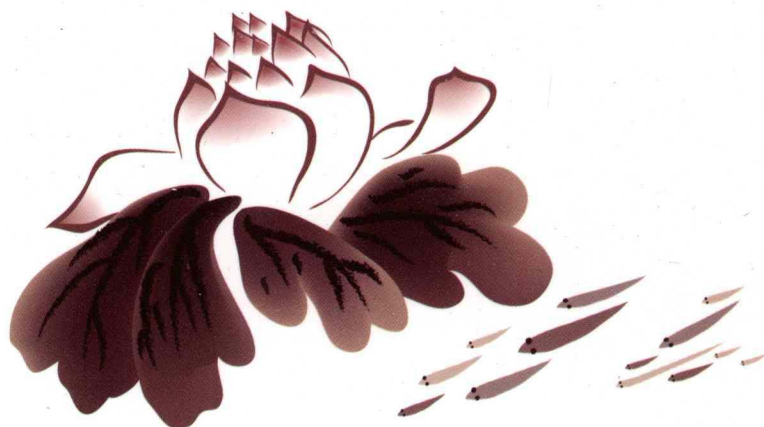


实用 应用开发技术

张宏展
王建成
蔡宗琰

等编著



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

实用XML 应用开发技术

张宏展
王建成
蔡宗琰
等编著



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

目前已经有讲解 XML 开发的书籍,但它们之中的大多数似乎更侧重于理论,实用性不是很强,对于初学者而言,显得有点深奥,很难轻松上手。

本书在讲解 XML 基本理论的基础上,讲解了 XML 编辑器和 XML 开发环境的构建,然后讲解了两种 XML 实用开发技术:一种是 Xerces 编程技术,在对 Xerces 功能进行分析并对 SAX, SAX2 和 DOM 等编程规范进行描述的基础上,设计实现了一个 Xerces 应用实例;另一种是 BDB XML 编程技术,在讲解 BDB XML 基本知识,对 XML 管理器和容器,在 BDB XML 中使用 XQuery,以及 BDB XML 索引和事务进行描述的基础上,设计实现了一个 BDB XML 应用实例。

本书适于初学 XML 编程的程序员阅读,也适于具有一定 XML 编程基础的程序员阅读,书中的实例对读者的实际应用开发有非常强的借鉴意义。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

实用 XML 应用开发技术 / 张宏展, 王建成, 蔡宗琰等编著. 北京: 电子工业出版社, 2008.7

ISBN 978-7-121-06675-7

I. 实… II. ①张…②王…③蔡… III. 可扩展语言 XML—程序设计 IV. TP312

中国版本图书馆 CIP 数据核字 (2008) 第 068386 号

责任编辑: 孙学瑛

印 刷: 北京智力达印刷有限公司

装 订: 北京中新伟业印刷有限公司

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: 787×1092 1/16 印张: 23 字数: 426 千字

印 次: 2008 年 7 月第 1 次印刷

印 数: 5000 册 定价: 49.00 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 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

推荐序

有幸在第一时间读到《实用 XML 应用开发技术》这本书的部分章节，也很乐意接受作者的邀请为本书的出版写几句话。

作者结合自己多年来对 XML 技术的研究和开发，将在研究开发过程中的体验和积累的知识整理出来与大家分享，从这一点上我觉得他们是有心人，我由衷地为他们感到高兴。这是系统讲解新的 XML 开发技术——DBXML 开发技术的中文书籍。看到我们上海贝尔阿尔卡特研发人员能有这样的科研成果，我也感到非常的自豪和高兴。

这本书的特点并不在于技术的精准或深奥，而在于它的切实可用性及新颖性。作者从事软件教学和软件平台研发工作多年，携其多年的理论和实践经验，厚积薄发，从一个应用者的角度诠释 XML 应用开发中的各个方面。

而站在开发者的角度，书中首先讲解 XML 技术，接着是对 XML 编辑器和开发环境的讲解，使开发者能够迅速熟悉并上手，最后是对两种 XML 开发技术的详细讲解。其中 Apache Xerces 开发技术比较流行，适合于想从底层开发功能强大的 XML 应用的开发者；而 DBXML 是由美加州大学伯克利分校开发的一种较新的 XML 开发技术，能够大大提高开发效率。

本书章节组织结构清晰，前半部分都着墨于理论阐述，后半部分深入讲解两种开发技术，并带有应用实例。可见作者在编写过程中确实下了很大的功夫。

这是一本实用性很强的书籍，非常适合于从事 XML 应用的读者，书中包含大量的实用 API 函数，同时又不失为一本很好的参考书籍。

其实，企业科研人员大多工作在开发新产品、新技术的前沿，他们往往有很多实际的感悟和经验，这种经验的传播和共享，就是从个人智慧到集体智慧的转化，这是现在很流行的企业知识管理中的一个基本规则。

如果将此概念扩展到企业的伙伴，乃至整个社会，那就真正成为了人类进步的阶梯。作者能将他工作学习中的第一手经验来与广大读者分享，这种精神值得广大科技工作者学习和发扬。

李春亭

上海贝尔阿尔卡特移动研发部总监

2008-04-30

前言

目前已经有很多讲解 XML 开发的书籍，但它们之中的大多数似乎更侧重于理论，实用性不是很强，对于初学者而言，显得有点深奥，很难轻松上手。

读者通过学习本书，可以掌握关于 XML 开发方面的基本知识和技巧。本书分为两个部分：前半部分讲解 XML 的基本理论和相关工具，以及如何基于 Xerces 开发 XML 应用；后半部分讲解如何快速使用 BDB XML 开发 XML 应用。为了提高读者的学习效率，这两部分都采取循序渐进，理论与实用相结合的方式组织各个章节，在最后都给出了一个应用实例，方便读者快速上手。

本书读者对象

如果您完全不懂 XML，这本书恰恰是您所需要的。本书几乎覆盖了 XML 开发的所有层面。它包含了 XML 开发语言的每一方面的丰富的信息，而且本书的设计风格会使您很容易跟着学下去。

本书包含完备的 Xerces 参考，可以成为您的一本 Xerces 编程手册。如果您想使用 BDB XML 进行开发的话，本书也可以引导您快速上手。本书包含的两个应用实例已经在 UNIX/Linux 下调试通过，您完全可以拿来直接运行。

在这本书中，您可以找到使您的开发工作变得更加容易的例子和解释。

怎样使用这本书

本书共有 13 章，可以分为两个部分：第 1 部分包括 1 至 7 章，主要讲解 XML 的基本理论知识和 Xerces 基本编程；第 2 部分包括 8 至 13 章，讲解 BDB XML 编程。

如果您是初次接触 XML 编程，建议您先从第 1 部分读起，这有助于您熟悉并了解 XML 规范及基本编程机制。第 1 章详细讲解 XML 的基本概念，包括 XML 技术标准，DTD 和 XML Schema 以及它们的优缺点。第 2 章在回顾 DTD 语法的基础上，对 XML

Schema 进行了详细描述。第 3 章讲解 XML 开发时使用的编辑器，并对 XML Spy 的功能进行了简单讲解。第 4 章讲解构造 XML 开发环境。第 5 章对常用的 Xerces 开发包的功能进行分析。第 6 章是 Xerces 编程指导，包括 SAX，SAX2 和 DOM 的编程指导，第 7 章给出了一个 Xerces-C++编程实例，使读者能够快速上手。

如果您已经拥有 XML 的基本知识并了解 Xerces 编程，只是想采用更有效的方法编写 XML 应用程序，或提高工作效率，则您可以直接开始读本书的第 2 部分。这部分的第 8 章讲解 Berkeley DB XML 的基本知识，包括 XPath 和 XQuery。第 9 章讲解 Xml 管理器和容器。第 10 章讲解如何在 BDB XML 中使用 XQuery。第 11 章讲解 BDB XML 索引。第 12 章讲解使用 BDB XML 事务。第 13 章给出了一个 BDB XML 应用实例，使读者能够快速熟悉 BDB XML 编程。

本书中的约定

注：本书中的注，提供了一些额外的信息，这些信息可以帮助您避免一些错误或是提醒您在使用所描述的特色时会遇到或应该考虑的一些问题。

本书的源代码

请访问以下网址：<http://www.broadview.com.cn>，在那里您可以找到本书中的两个应用实例的源代码，以及一些辅助资料。

致谢

编著本书花费了近一年的时间。可以很自豪地说这是一本很好的书，本书的出版得到很多人的帮助。

感谢部门总监李春亭先生为本书写序，感谢我的朋友吴欣，感谢我的同事宫宁、沈震生、Ni Eric、邓浩，以及上海的叶剑锋、李春旭、蒋智宁等，本书的出版得到了他们的支持和帮助。

目录

第 1 篇 XML 基本理论和 Xerces 基本编程

第1章 XML文档 2

1.1 什么是 XML	2
1.2 XML 技术标准	3
1.2.1 XML 相关标准	3
1.2.2 XML 的主要技术特点	4
1.2.3 XML Schema	5
1.3 XML Schema 和 DTD	6
1.3.1 XML 模式与 XML 格式	6
1.3.2 数据类型	7
1.3.3 元素顺序的支持	8
1.3.4 命名空间	9
1.3.5 对 API 的支持	10
1.3.6 对于属性的定义	11
1.3.7 注释	11
1.3.8 对数据库的支持	11
1.4 关于 XML 的几点说明	12

第2章 XML Schema 17

2.1 DTD 语法	17
2.1.1 工作示例	18
2.1.2 关于 DTD 的几点说明	20

2.2	内置数据类型.....	20
2.2.1	数据类型分类	20
2.2.2	数据类型继承图	23
2.2.3	基本数据类型	23
2.2.4	派生数据类型	24
2.2.5	常用数据类型描述	25
2.3	命名空间.....	29
2.3.1	什么是命名空间	30
2.3.2	XML 中的命名空间.....	31
2.3.3	对命名空间进行命名	31
2.3.4	定义命名空间	33
2.3.5	使用命名空间	34
2.3.6	schemaLocation 属性	37
2.3.7	XML Schema 和 XML Schema Instance 命名空间.....	38
2.3.8	elementFormDefault 和 attributeFormDefault 属性	39
2.4	元素定义.....	40
2.4.1	元素定义	40
2.4.2	元素命名规则	41
2.4.3	使用子元素还是使用属性	41
2.5	自定义数据类型.....	42
2.5.1	简单数据类型	42
2.5.2	简单数据类型	48
2.6	结论	61

第 3 章 XML 编辑器 62

3.1	XML 编辑器综述.....	62
3.1.1	Morphon	63
3.1.2	<oXygen/>	65
3.1.3	XML Editor.....	66
3.1.4	XMLwriter	67
3.1.5	EditML Pro	68
3.1.6	XML Spy	70
3.1.7	XMetal 3	71
3.1.8	Eclipse XML 编辑器.....	73
3.1.9	其他产品	74

3.2 XML Spy 功能介绍	75
3.2.1 XMLSpy 用户界面	75
3.2.2 创建一个 XML Schema 文件	76
3.2.3 高级的 XML Schema 定义	86
3.3 创建一个 XML 文档	94
3.4 验证文档	101
3.5 工程管理	105

第 4 章 构造XML开发环境 108

4.1 XML Parser 综述	108
4.1.1 XML Parser 概述	108
4.1.2 DOM	109
4.1.3 SAX	111
4.1.4 总结	112
4.2 XML Parser 的比较	113
4.2.1 Apache Xerces	114
4.2.2 JAXP	115
4.2.3 Microsoft MSXML	116
4.2.4 Oracle XML Parser for Java	118
4.2.5 James Clark XP	119
4.3 构造 Apache Xerces 开发环境	120
4.4 Xerces API 文档	122
4.4.1 文档规范	122
4.4.2 常见问题	123
4.4.3 编程指导	127

第 5 章 Xerces-C++功能分析 129

5.1 Xerces-C++解析器	129
5.2 建立 Xerces-C++示例程序	130
5.3 Xerces-C++功能分析	132
5.3.1 SAXCount	132
5.3.2 SAXPrint	133
5.3.3 DOMCount	135
5.3.4 DOMPrint	135

5.3.5	MemParse	137
5.3.6	Redirect	139
5.3.7	PParse	139
5.3.8	StdInParse	140
5.3.9	EnumVal	141
5.3.10	CreateDOMDocument	142
5.3.11	Sax2Count	142
5.3.12	SAX2Print	143
5.3.13	SEnumVal	145
5.3.14	PSVIWriter	147
5.3.15	SCMPrint	148

第 6 章 Xerces-C++编程指导 150

6.1	Xerces-C++解析器	150
6.2	SAX 编程指导	151
6.2.1	使用 SAX API	151
6.2.2	SAXParser	152
6.3	SAX2 编程指导	158
6.3.1	使用 SAX2 API	158
6.3.2	SAX2XMLReader	159
6.4	DOM 编程指导	167
6.4.1	使用 DOM API	167
6.4.2	XercesDOMParser	171
6.4.3	DOMBuilder	178
6.4.4	DOMWriter	189
6.5	DOM Level 3 支持	194

第 7 章 Xerces-C++开发实例 196

7.1	应用例子描述	196
7.1.1	需求描述	196
7.1.2	需求分析	201
7.2	应用例子设计	206
7.2.1	XmlMsgRouter 类	206
7.2.2	XmlMain 类	208

7.3 XML 文档查询操作	214
7.4 XML 文档增加节点操作	215
7.4.1 doSetActionToMainTree()方法	215
7.4.2 getSetOption()方法	216
7.4.3 doCreateAction()方法	217
7.4.4 XmlSetResponse()方法	222
7.5 XML 文档修改节点操作	224
7.6 XML 文档删除节点操作	225
7.7 路径打印和清除操作	226
7.7.1 printPathList()方法	226
7.7.2 clearPathList ()方法	226

第 2 篇 BDB XML 编程

第 8 章 Berkeley DB XML基础 228

8.1 BDB XML 概述	228
8.1.1 功能描述	228
8.1.2 主要特性	229
8.2 XPath 路径语言	231
8.2.1 XPath 中的节点类型	232
8.2.2 文档顺序	233
8.2.3 XPath 表达式介绍	234
8.2.4 XPath 中的定位路径	235
8.2.5 定位路径实例讲解 (1)	238
8.2.6 定位路径实例讲解 (2)	240
8.2.7 XPath 核心函数	242
8.3 XQuery 数据查询语言	243
8.3.1 XQuery 概述	243
8.3.2 XQuery 查询应用	244
8.3.3 XQuery 转换应用	245
8.3.4 XQuery 更新应用	247
8.3.5 XQuery 与其他语言的比较	247
8.4 更多 DBD XML 资源	248

第 9 章 XML 管理器和容器 250

9.1	XmlManager	250
9.1.1	Berkeley DB 环境	250
9.1.2	XmlManager 实例和释放	253
9.2	管理容器	255
9.2.1	容器标记	255
9.2.2	容器类型	256
9.2.3	删除和重命名容器	257
9.3	容器中装载 XML 文档	258
9.3.1	输入流和字符串	258
9.3.2	添加文档	259
9.3.3	设置元数据	261
9.4	管理容器中的文档	263
9.4.1	删除文档	263
9.4.2	替换文档	264
9.4.3	修改文档	265

第 10 章 在 BDB XML 中使用 XQuery 274

10.1	XQuery 简单介绍	274
10.1.1	使用 XQuery 提取部分文档	275
10.1.2	判断语	275
10.1.3	上下文	276
10.1.4	通配符	278
10.1.5	操作函数	279
10.1.6	在 BDB XML 中使用 FLWOR	280
10.2	使用 XQuery 回取 BDB XML 文档	281
10.2.1	Query 上下文	281
10.2.2	执行查询	284
10.3	检查 XQuery 结果	286
10.3.1	检查文档值	287
10.3.2	检查元数据	289
10.3.3	使用事件读入器	290

第 11 章 使用 BDB XML 索引 293

11.1 索引及其类型	293
11.1.1 唯一性	294
11.1.2 路径类型	294
11.1.3 节点类型	295
11.1.4 键类型	295
11.2 语法类型	296
11.3 指定索引策略	296
11.3.1 使用字符串指定索引	296
11.3.2 使用枚举类型指定索引	298
11.4 指定索引节点	300
11.5 索引处理注解	301
11.6 管理 BDB XML 索引	302
11.6.1 增加索引	302
11.6.2 删除索引	303
11.6.3 替换索引	304
11.6.4 检查容器索引	305
11.6.5 使用默认索引	305
11.6.6 查寻索引文档	306
11.7 使用 Query 计划验证索引	308
11.7.1 查询计划	308
11.7.2 使用 dbxml Shell 检查查询计划	310

第 12 章 使用事务 313

12.1 BDB XML 事务	313
12.2 初始化事务子系统	314
12.3 事务保护容器操作	316
12.4 事务考虑	318

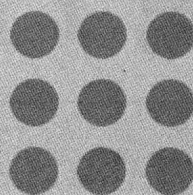
第 13 章 BDB XML 应用实例 320

13.1 应用描述	320
13.1.1 需求描述	320
13.1.2 需求分析	327

13.2	应用设计	337
13.2.1	exampleLoadContainer()方法	337
13.2.2	loadMyExamplesData.sh 命令	341
13.3	应用实现	343
13.3.1	XML 装入容器的实现	343
13.3.2	doQueryIE()方法	344
13.3.3	getEvtNameAttrValue()方法	348
13.3.4	doQueryResults()方法	349
13.3.5	其他方法和源代码	350

1

第 1 篇



XML 基本理论和 Xerces 基本编程

- 第1章 XML文档 2
- 第2章 XML Schema 17
- 第3章 XML编辑器 62
- 第4章 构造XML开发环境 108
- 第5章 Xerces-C++功能分析 129
- 第6章 Xerces-C++编程指导 150
- 第7章 Xerces-C++开发实例 196

第 1 章

XML文档

XML 是一种比较新的数据储存格式，也可以被看成是一种超越 HTML 能力范围的新语言。XML 具有简单、开放、可扩展性、灵活性和自描述性等特点，给 Internet 的应用注入了新的活力。作为数据表示的一个开放标准，XML 为数据集成带来了新的机遇。用 XML 作为数据交换的中介，给系统的实现带来了极大的灵活性。系统可以屏蔽掉后台的多种数据源，将统一的 XML 格式的数据呈现给用户。

DTD/XML Schema 用于验证 XML 数据的有效性。数据接收方可以根据 DTD/XML Schema 对 XML 数据进行任意的处理，如分解出其中需要的数据或是以不同的样式来显示。

1.1 什么是 XML

XML 是一种比较新的数据储存格式。它使用树状结构的标签对数据进行格式化。这种方法和传统的 HTML 非常地类似，但 HTML 是格式、超文本和图形用户界面语法的混合体，要同时发展这些混合在一起的功能是很困难的。而 XML 提供了一种结构化的数据表示方式，使得用户界面与结构化数据相分离，并且 XML 的标签是可由用户自定义的，甚至可以使用中文等双字节编码文字作为 XML 标签。比如，我们可以建立这样一个 XML 文档：

```
<学生>
  <学号 N="100001">
    <姓名>顾雪明</姓名>
    <性别>男</性别>
  </学号>
  <学号 N="1000002">
    <姓名>韩志勇</姓名>
    <性别>男</性别>
  </学号>
</学生>
```

对于这样一个文档, 标签具有自我描述性, 即使是不懂网络知识的人也可以轻易地从中找到大部分的信息。同时, XML 在结合了 CSS, XSL, ASP 和 PHP 后更是在信息检索方面展现了自己的特长。这几年来, XML 越来越受到了大家的青睐, 其应用领域也越来越广泛, 如已经应用于通信、航空、教育和医疗等领域。

下面的例子(清单 1-1)来自 www.moreover.com, 这个发布新闻信息的系统是关于发行/订阅结构的相当好的例子。Moreover 组织收集了来自全球超过 1 800 个站点的新闻条目并对其进行处理, 以使用户订阅。订阅用户只能够下载自己感兴趣的新闻。

清单 1-1 XML 编码的新闻

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE moreovernews SYSTEM "http://p.moreover.com/xml_dtds/moreovernews.dtd">
<moreovernews>
  <article id="_13177343">
    <url>http://c.moreover.com/click/here.pl?x13177340</url>
    <headline_text>Microsoft Signs Pact with Electronic Arts to Develop Games
    for Its X-Box Console</headline_text>
  </article>
  <article id="_13176679">
    <url>http://c.moreover.com/click/here.pl?x13176678</url>
    <headline_text>Attorney general says children have easy access to violent
    video games </headline_text>
  </article>
  <article id="_13171770">
    <url>http://c.moreover.com/click/here.pl?x13171767</url>
    <headline_text>Nintendo generation piling on the kilos</headline_text>
  </article>
</moreovernews>
```

这里使用 DTD (Document Type Define, 文档类型定义) 类型文件 `moreovernews.dtd` 对 XML 文档进行验证, 即 XML 文档的元素、元素间嵌套关系、元素内容类型、取值范围等都要符合 `moreovernews.dtd` 的定义。另一种对 XML 文档进行验证的方法是 XML Schema。我们将在后面分别对这两种方法进行介绍。

1.2 XML 技术标准

1.2.1 XML 相关标准

XML 不仅仅是一种语言, 广义的 XML 是一系列标准的集合。图 1.1 说明了 XML 及其相关标准。图中, 中间一层是 XML 基础标准, 最下层是 XML 的应用标准。

XML 基础标准是为 XML 的进一步实用化制定的标准, 它规定了采用 XML 制定标准时的一些公用特征、方法或规则。XML Schema 描述了更加严格定义 XML 文档的方