



今日电子

美国计算机“宝典”丛书

Java, XML, and Web
Services Bible

丛书
累计印
73万册

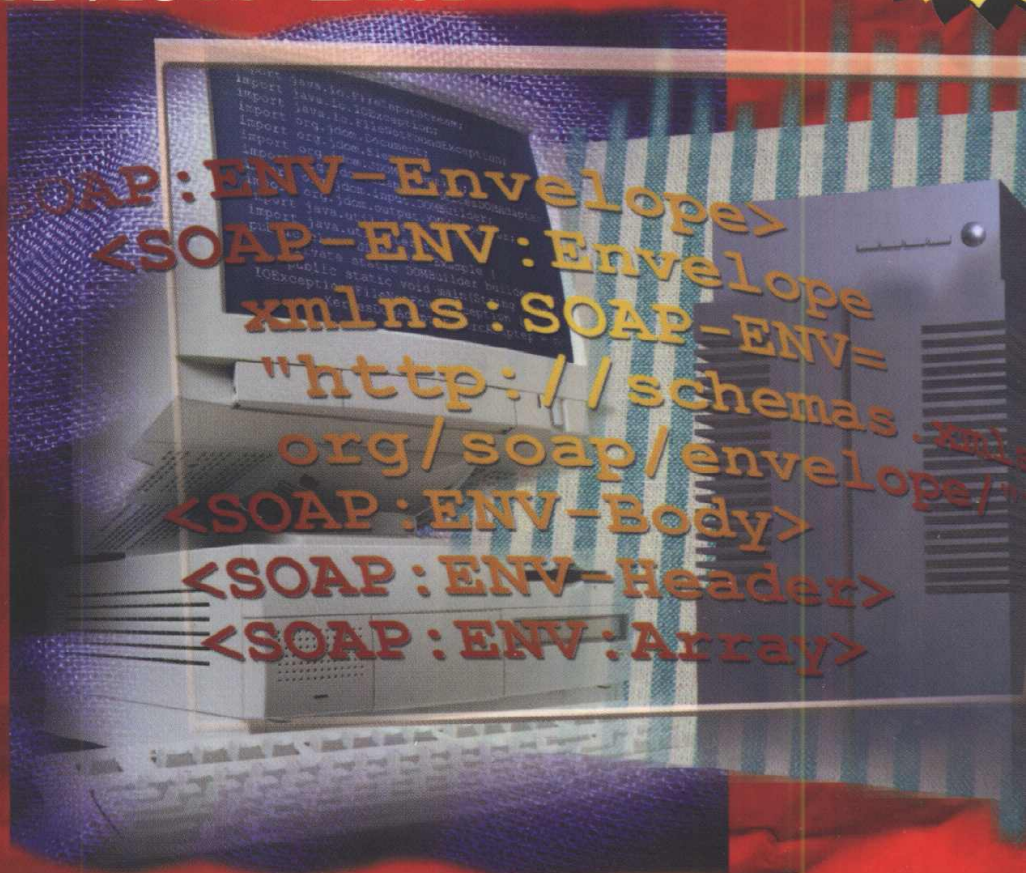
100%

内容丰富、权威

挖掘Java和XML这一
组合在多层应用程序
开发方面的最大潜能

掌握Java, XML与
SAX API, Castor的
绑定

构建面向业务与客户
的新一代Web服务



[美] Mike Jasnowski 等著 盖江南 王 勇 等译

Java, XML 和 Web 服务

宝典

深层覆盖

XML-RPC、
SOAP、即时消息
传递、SVG、应用程
序管理等更多内容



电子工业出版社

Publishing House of Electronics Industry

803

7P312JA
J32a1

美国计算机“宝典”丛书

Java, XML 和 Web 服务宝典

Java, XML, and Web Services Bible

[美] Mike Jasnowski 著

盖江南 王 勇 等译

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

出版说明

21 世纪是一个崭新的世纪，是催人奋进的世纪。在新世纪的第一乐章中，我们热忱地向广大读者、IT 人士推介这套全新改版的美国计算机“宝典”丛书。

丛书的出版宗旨

本着提高广大读者计算机专业技能的宗旨，我社从美国 Hungry Minds(原 IDG Books Worldwide) 公司引入了这套“宝典”丛书。该丛书在世界各地 51 个国家被译为 31 种文字，拥有几百万的读者。在我国自 1994 年引入这套丛书以来，截至 2001 年中期累计销量已近 60 余万册。得到了广大读者的认同，成为电子工业出版社的著名品牌之一。

丛书的涉及范围

“宝典”丛书的涉及范围甚广，既包括众多的流行软件、编程语言、图形图像，也包括数据库、网络等高端技术等方面的书籍。对于某些软件，我们还进行了本地化处理，按相应的中文版软件进行了调整，进一步贴近中国读者的需求。

每一本“宝典”共同贯彻的一项宗旨就是，对相应主题的介绍都非常全面、系统。使该软件或系统能做到的，读者通过本书的学习也能做到。

丛书的创作队伍

“宝典”丛书的作者都是某个计算机专业领域的专家、教授，有些还是某软件的特约测试者。比如 Deke McClelland、Alan Simpson 和 Ellen Finkelstein 等知名畅销计算机图书作家，在相关领域都具有很高的声望。他们拥有丰富的实践经验，所介绍的内容都是在工作中得到千锤百炼，具有一定的权威性。在他们所撰写的书籍当中，会介绍一些技巧，同时也会为读者提出某些忠告，以免犯同样的错误。

在中文版“宝典”中我们也本着同样的原则，所选的译者均经过严格筛选，他们大都是来自于高等院校的教授、学者，计算机领域的高手，不但具有高深的专业知识，同时也具备英语方面的深厚底蕴。我们的编辑队伍，同样是来自于计算机专业的高素质人才。通过这种严格的层层把关，我们相信最终奉献给读者的将是一部部精品。

丛书的新特性

进入新的世纪，“宝典”将以全新的面貌呈现在广大读者面前。无论是版式、用纸还是印刷质量，相关人员都颇费一番苦心，进行了很大改善。同时我们对于丛书的选题也进行了调整，使其更适合我国的计算机发展水平。对于原书中某些不适合中国国情以及过于调侃的内容进行了删减。我们将秉承“宝典”丛书一贯的“权威、全面、精益求精”的风格，力争使每一本书成为您探索计算机领域奥秘的“宝典”。

译者序

随着信息时代的来临,越来越多的信息需要在Internet上传播。信息种类的增多,使得具有固定标记集合的HTML标记语言越来越不能满足Internet的需求,而XML这种新型的可扩展标记语言兼取了HTML和SGML的优点,既通用全面,又简洁明了,从而适合各种网络应用的需要,成为Internet发展的必然趋势。通过XML的使用,展示了它较其他标记语言更大的可扩展性与灵活性,使用者不再局限于使用固定标记集合中的标记来描述文件中的各个元素,他们可以自定义更多适合要表达的元素标记,形成自己的标记库,从而使得文件的内容更加丰富、复杂以构成一个完整的信息体系。

在当今越来越多的应用中,已经将XML作为数据交换和分布式计算的基础,成为网络通用的语言。将XML作为描述信息的主要方式,用XML表达信息元素、传递数据,不仅跨越了平台,还跨越了空间,更跨越了设备之间的束缚。由此带来了XML的几大特点:平台无关性、数据和表现分离、良好的数据存储格式、可扩展性、高度结构化、便于网络传输,从而为基于Web的应用软件赋予了强大的功能和灵活性。

对于XML文档的处理是在基于树的DOM和基于事件的SAX基础上,而大多的XML工具是基于Java编写的。XML是一种与平台无关的中性格式,而Java是与平台无关的程序设计语言,两者的结合必然使得编写的应用程序具有更大的灵活性和可移植性,从而适应网络的需求。

Java是一种Internet的编程语言,可用于创建安全、可移植、面向对象、多线程、交互式的程序。在处理XML文档时,比如解析和转换XML文档的时候,就可以利用Java编写的API进行具体的处理,特别是在编写Web服务应用程序方面,可以结合使用XML和Java两种技术,使得应用程序具有无可比拟的优越性。

本书主要讲述了XML文档的创建、处理、应用,其中涉及到Java技术以及其他更多有用的技术。主要内容包括XML的语法、规则、转换,XML文档到Java表示形式的转换,以及XML在用户界面、图形、分布式计算中的应用。另外,还介绍了在传统的存储技术中集成XML技术的内容。最后讲述了Web服务的相关内容和应用Java、XML两种技术编写Web服务应用程序的内容。

书中除了大量技术概念、使用方法的讲解外,还例举了很多代码示例,以便帮助读者理解XML、Java的应用。这是一本难得的参考书。

盖江南、王勇、尚红昕、韩旭、赵迎、傅长东、尚天任、王兰、郭雍、曾洋、李天乐、张力新、严明、宋强、许仲翔、徐亮、关天海、李辉明、周哲、郑暉等同志分别翻译了本书的各个章节或参加了部分审校工作。

由于译者水平有限,时间仓促,特别是一些首次出现的专业术语,译文中难免有不妥之处,恳请读者批评指正。

译者

2002年3月

目 录

前言	1
第 1 部分 XML 的使用	5
第 1 章 XML 概述	6
1.1 XML 简介	6
1.2 理解 XML 文档	8
1.3 文档类型定义 (DTD)	15
1.4 实体	23
1.5 XML 模式	25
1.6 小结	32
第 2 章 使用 DOM/SAX/JDOM/JAXP	33
2.1 文档对象模型 (DOM)	33
2.2 DOM API 应用	40
2.3 XML 简单应用程序编程接口	49
2.4 SAX API 应用	52
2.5 JDOM	54
2.6 JAXP	61
2.7 小结	63
第 3 章 使用 XML 进行配置	64
3.1 配置语言	64
3.2 XML 解决办法	67
3.3 使用 XML 的应用	68
3.4 用 XML 配置应用程序	71
3.5 小结	78
第 2 部分 转换语言的使用	79
第 4 章 转换语言概述	80
4.1 什么是 XPath	80
4.2 什么是 XSLT	87
4.3 小结	100
第 5 章 转换 XML	101
5.1 什么是 XSLT 处理器	101

5.2	Xalan Java XSLT 处理器	102
5.3	XML 到 HTML 的转换	103
5.4	节点到 HTML 的转换	108
5.5	使用 JAXP 实施转换	112
5.6	另一种 XML 到 HTML 的转换	113
5.7	另一种节点到 HTML 的转换	117
5.8	XML 到 XML 的转换	119
5.9	小结	122
第 6 章	XSLT 扩展	123
6.1	什么是 XSLT 扩展	123
6.2	编写扩展元素	125
6.3	编写扩展函数	133
6.4	调用静态方法	136
6.5	小结	137
第 7 章	Web 发布	138
7.1	Web 发布概述	139
7.2	Cocoon	142
7.3	转换和 Cocoon	147
7.4	可扩展服务器页面和 Cocoon	149
7.5	XSP 示例	159
7.6	小结	164
第 3 部分	XML 和分布式计算	165
第 8 章	分布式应用、传输协议和 XML	166
8.1	远程过程调用	166
8.2	传输协议	168
8.3	超文本传输协议	169
8.4	实体消息头	172
8.5	处理 HTTP 消息	173
8.6	HTTP 客户	184
8.7	HTTP 示例	186
8.8	小结	187
第 9 章	XML-RPC	189
9.1	XML-RPC 规范	189
9.2	以新的底层结构实现 XML-RPC	194
9.3	简单的 XML-RPC 示例	211
9.4	在现有底层结构中实现 XML-RPC	213
9.5	方法发现	215
9.6	小结	216

第 10 章 SOAP 协议	217
10.1 SOAP 与 XML-RPC	217
10.2 SOAP 消息	218
10.3 SOAP 串行化和数据类型	222
10.4 Apache SOAP 工具包	225
10.5 一个更复杂的服务	236
10.6 小结	241
第 11 章 XML 信息交换	242
11.1 什么是信息交换	242
11.2 Java 消息服务	243
11.3 小结	252
第 12 章 即时消息传递	253
12.1 即时消息传递系统	253
12.2 XML 和即时消息传递	257
12.3 客户	268
12.4 消息格式	274
12.5 小结	278
第 4 部分 Java 绑定和 XML	279
第 13 章 使用 SAX API 创建 Java 对象	280
13.1 Java 绑定案例	280
13.2 什么是 Java 绑定	281
13.3 小结	294
第 14 章 Castor	295
14.1 什么是 Castor	295
14.2 源代码生成器	306
14.3 编组和反编组	312
14.4 小结	320
第 15 章 用 XML 表示 Swing 组件的永久性	321
15.1 串行化 Java 对象	321
15.2 串行化一个 <code>JavaBean</code>	323
15.3 小结	327
第 5 部分 用户界面和图形	329
第 16 章 可缩放矢量图形	330
16.1 什么是可缩放矢量图形	330
16.2 可缩放矢量图形规范	332

16.3	SVG 文档的结构和功能	335
16.4	用 SVG 样式化对象	349
16.5	用 SVG 和 ECMAScript 编写脚本	351
16.6	链接到资源: <a>元素	355
16.7	使用图像: <image>元素	356
16.8	动画: <animate>元素和<animateMotion>元素	356
16.9	小结	359
第 17 章	与 Java 一起使用的 SVG	360
17.1	Apache Batik SVG 工具包	360
17.2	SVG 文档和服务端	368
17.3	SVG 贺卡应用程序	372
17.4	小结	375
第 18 章	用 XML 创建基于客户的 Java 用户界面	376
18.1	设计时与运行时	376
18.2	使用 XML	378
18.3	小结	395
第 19 章	用 XML 创建基于 Web 的用户界面	396
19.1	HTML 表单	396
19.2	不同类型的表单元素	400
19.3	小结	404
第 6 部分	应用程序管理	405
第 20 章	Java 管理扩展	406
20.1	什么是 JMX	406
20.2	JMX 体系结构	407
20.3	小结	416
第 21 章	基于 XML 的应用程序管理	417
21.1	基于 XML 的协议适配器	417
21.2	两个 MBean: Simple 和 XML	422
21.3	小结	436
第 7 部分	XML 和数据库	437
第 22 章	XML 数据库和工具	438
22.1	将 XML 文档存储在数据库中	438
22.2	数据库产品中的 XML 支持	441
22.3	小结	449

第 23 章 将 XML 文档映射到关系数据库	450
23.1 开始映射	450
23.2 查看 DTD	451
23.3 小结	458
第 24 章 用 XIS 管理本地 XML 数据	459
24.1 XIS 概述	459
24.2 处理 XIS 中的 XML 文档	464
24.3 小结	473
第 8 部分 XML 在三层应用程序中的使用	475
第 25 章 在 servlet/JSP 中使用 XSLT	476
25.1 基础技术概述	476
25.2 内容和表示形式	487
25.3 性能问题	501
25.4 小结	505
第 26 章 标记库的使用	506
26.1 什么是标记库	506
26.2 标记处理程序的概念和规则	514
26.3 小结	516
第 27 章 使用 servlet, JSP 和 XML 创建可以在 Web 上传输的数据	517
27.1 多个层	517
27.2 GameTrader.com	520
27.3 小结	530
第 9 部分 Web 服务	531
第 28 章 服务描述	532
28.1 Web 服务简介	532
28.2 接口与实现	534
28.3 Web 服务描述语言	535
28.4 在服务器 (程序) 上使用 WSDL	543
28.5 在客户上使用 WSDL	548
28.6 按程序设计使用 WSDL	557
28.7 小结	562
第 29 章 发现服务	563
29.1 目录服务	563
29.2 UDDI 标准	564

29.3	发布到 UDDI 注册表	569
29.4	查询 UDDI 注册表	581
29.5	服务广告	588
29.6	小结	588
第 30 章	高级 Web 服务	589
30.1	有状态的服务	589
30.2	附件	593
30.3	安全性	599
30.4	互操作性	608
30.5	各厂商提供的体系结构	609
30.6	小结	610
第 31 章	Web 服务和企业对企业的信息交换	611
31.1	商务交易建模语言	611
31.2	采办和供应链管理	616
31.3	在线市场	641
31.4	小结	643
第 32 章	为客户提供的 Web 服务	644
32.1	客户服务	644
32.2	对等技术	658
32.3	小结	672
附录 A	Java, XML 和 Web 服务的网上资源	673
附录 B	DOM Level 2 Events 和 Traversal 模块, 解析非 XML 数据	675
附录 C	JAXM, SOAP 和 XML-RPC	701

前 言

Java, XML 和 Web 服务分别有不同的含义。Java 已经成为一种强大的平台中性程序设计语言, 而 XML 则作为相应的平台中性数据格式。

在过去的几年, XML 受到了人们的广泛关注。人们发现, XML 是一种非常好的格式, 可以用于应用程序的各个层。但是, 既然 XML 有如此强大的功能, 人们又如何利用它呢? 我们经常听到有人声称 XML 是最终的文件格式, 并且可用于转换数据。但是如何应用呢?

本书不仅介绍技术层面的内容, 而且介绍如何在实践中结合使用 Java 和 XML 两种技术。在这本书中, 不仅讲述了如何处理和创建 XML, 还讲述了在各种应用中如何使用 XML 技术的内容。读者从本书中还能够了解到如何使用 Java 和 XML 创建使用 Web 服务的新型应用分支。此外, 还探讨了各种可用的技术, 比如 XML, SOAP, Java servlet 和 JSP 技术, 使用这些技术可以编写出交换 XML 编码的事务消息的应用程序。

本书的读者对象

本书适合使用 Java 和 XML 创建应用程序的人群, 此外有兴趣学习如何结合使用这两种技术的人群也适合阅读这本书。书中向人们提供了一个构造和功能化 Java 和 XML 应用的窗口, 并且着重介绍了一些商业性产品以及开放源代码产品。另外, 本书还描述了一些底层技术和协议, 有了这些技术和协议, 才可能在分布式应用中使用 XML 提供 XML-RPC 或 Web 服务的功能。

XML 初学者

如果你是使用 XML 的新手, 那么这本书将帮助你在 XML 方面快速入门。XML 作为一种企业注册表语言, 要求用户遵循一些基本的规则, 这些规则非常容易学会。这本书将向你介绍有关 XML 的规则和语法, 如果要使得所构造的 XML 文档有效并且格式良好, 就必须使用这些规则和语法。另外, 这本书还介绍了软件模块的使用, 比如解析 XML 文档的 XML 解析器和转换 XML 文档的 XSLT 处理器。

Java 程序员

本书是针对在 Java 程序设计语言方面有一定基础的读者编写的。所以, 如果读者还不熟悉 Java 语言, 应该先学会如何编译和创建基本的 Java 类, 而且还要熟悉面向对象程序设计的概念, 比如继承性。但是, 读者不一定要是这方面的专家。

Java 和 XML 高手

如果你了解类的内部运作过程并且熟练使用 Emacs 编写 XML, 那么这本书也适合你的口味。你将从这本书中了解到用于解析和转换 XML 的最新的 Java API 信息。另外还可以了解到源代码开放的产品信息。

专业的 Microsoft 开发人员

如果你还没有转换到 Sun 系统并加入 Java 开发人员的队伍, 那么可能会不了解 XML 和 Java 之间的关系。这本书将向你讲述 Java 和 XML 结合使用的方法。

本书的组织结构

这本书向读者介绍了 XML 和 Java 的各种应用, 提供了结合使用两种技术的应用范例, 并讲述了一些新技术, 比如可缩放矢量图形、即时消息传递和应用管理方面的技术。

第1部分 XML 的使用

第1部分介绍了可扩展标记语言 (XML)。其中, 包括编写 XML 文档所使用的语法规则, 从头创建 XML 文档以及使用 XML 解析器的相关内容。第1章是 XML 语言的简介部分, 其中讨论了 XML 文档中的元素、属性以及其他语法上的构件。第2章讨论了 DOM 和 SAX API 在 XML 文档中的使用。另外, 还介绍了 JDOM 和 JAXP 技术, JDOM 可以代替 DOM 和 SAX 使用, 它和 DOM 技术密切相关, JAXP 是用于 XML 处理的一种新型的 Java API。第3章讨论了 XML 作为配置语言的使用。

第2部分 转换语言的使用

第2部分讲述了如何提取 XML 文档并将其转换为其他标记语言的内容。第4章概述了执行转换的基本内容。第5章首先介绍了一个应用模块——XSLT 处理器, 它主要负责处理 XSLT 样式表和生成转换的结果。然后讲述了 Xalan 这种 Java XSLT 处理器的使用以及 JAXP 在 XML 转换中的用法。第6章介绍了如何使用 XSLT 扩展扩充 XSLT 样式表的功能。另外, 还阐述了如何创建扩展元素和函数的方法。第7章讨论了如何使用 XML 发布基于 Web 的内容, 介绍了开放源代码的发布框架 Cocoon。

第3部分 XML 和分布式计算

第3部分介绍了 XML 和 Java 在分布式计算中的应用。第8章讨论了超文本传输协议 (HTTP)、远程过程调用 (RPC)。以及其他传输协议。第9章讨论了 XML-RPC 并介绍了使用 XML 作为 RPC 的一种机制的内容。在这一章中将通过 HTTP 协议使用 XML-RPC 执行远程过程调用。第10章介绍了简单对象存取协议 (SOAP) 及其作为实现 RPC 的机制的使用情况。还讨论了 SOAP 消息和消息头的格式和语法, 最后描述了 Apache SOAP 工具包的情况。第11章讨论了 XML 在信息交换的应用中使用情况, 并讲述了如何使用 Java 消息服务 (JMS) 交换 XML 编码形式的消息。第12章介绍了即时消息传送客户程序和服务器应用程序, 另外还介绍了如何交换基于 XML 的消息。

第4部分 Java 绑定和 XML

第4部分介绍了如何将 XML 文档绑定到 Java 表示形式。第13章介绍了除 org.w3c.dom.Document 接口提供的方法以外将 XML 文档表示为 Java 对象的其他传统方法, 还讨论了如何表

示XML文档中的实体以及修改或创建实体的内容。另外,讲述了如何使用SAX API根据XML表示形式创建Java对象。第14章介绍了更高级的XML绑定,其中用到了Castor工具。还讨论了Castor的各种组件,另外还讲述了如何将Java类编组为XML形式以及相反的反编组过程。第15章讲述了如何使用Swing GUI工具包中的永久JavaBean组件的持久特性,从而在XML形式中插入持久性的GUI组件。

第5部分 用户界面和图形

第5部分探讨了XML在图形化用户界面中的使用。第16章讨论了可缩放矢量图形(SVG)语言及其语法和规则。在创建SVG文档时,用SVG语法和规则表示文本和形状可以增加对ECMAScript的脚本支持。第17章讨论了SVG在Java应用中的使用,讲述了如何使用Apache Batik SVG工具包的组件。此外,还讲述了如何使用JSP动态创建SVG文档,并例举了贺卡应用程序的例子。第18章和第19章讲述了如何使用XML为基于Java和基于Web的两类应用动态创建用户界面。

第6部分 应用程序管理

第6部分讨论了XML作为一种应用管理协议的使用情况。第20章介绍了Java管理扩展(JMS)。第21章讲述了如何创建一个基于XML的协议适配器,以便为各种客户提供应用方面的管理。

第7部分 XML和数据库

第7部分讲述了在传统的存储技术中如何集成XML技术,传统的存储方式包括关系数据库管理系统(RDBMS)和面向对象的数据库。第22章讲述了如何在关系数据库和面向对象数据库中使用XML技术,还介绍了在这类数据库中存储XML文档的传统技术。第23章讨论了在将XML文档存储到关系数据库之前需要做哪些准备以及如何映射的问题。第24章讨论了在本地XML数据库中如何使用XML的内容,还讲述了如何使用eXcelon可扩展信息服务器(XEIS)的内容,XEIS将用于管理本地XML数据。

第8部分 XML在三层应用程序中的使用

目前,在很多应用中都可以应用XML技术。第8部分讨论了在遵循多层体系结构设计的应用中XML的使用问题。第25章讲述了在servlet中如何应用XSLT转换来自后台或任何数据源的XML文档。第26章讨论了类似XML的标记库,用其中的自定义标记扩展JSP。第27章阐述了如何使用servlet、JSP和XML的组合作为企业的数据入口的问题。

第9部分 Web服务

第9部分介绍了Web服务,这是XML的最新应用。Web服务使用XML作为向各种客户端展示应用功能的中介,不管应用程序是新版本还是老版本,都可以这么做。第28章介绍了Web服务并讨论了各种服务以及Web服务描述语言(WSDL)。第29章讲述了如何查找服务和发布服务,还讨论了有关注册表和通用描述、发现和集成(UDDI)技术的内容。第30章深入

探讨了现有分布式技术的持久性、安全性和相互可操作性,这些技术包括EJB,RMI和CORBA。第31章和第32章讨论了如何将Web服务应用于企业和客户的内容。第31章着重在企业应用方面,其中讨论了商务模型化语言,比如电子商务可扩展标记语言(ebXML)和 BizTalk。而第32章着重在客户端应用方面,其中讲述了客户如何与Web服务交互的内容。另外,还讲述了关于分布式计算的对等模型的Web服务应用。

附录 A Java, XML 和 Web 服务的网上资源

附录 A 中罗列了各种工具的网址,其中包括商业性的工具和源代码开放的工具。

附录 B DOM Level 2 Events 和 Traversal 模块, 解析非 XML 数据

附录 B 主要介绍 DOM, SAX 和 JAXP API 的高级特性。其中,讲述了 DOM Level 2 和解析非 XML 数据的内容,阐述了如何配置 XML 解析器以及用 XML 解析器对 XML 文档进行有效性验证的问题。

附录 C JAXM, SOAP 和 XML-RPC

附录 C 介绍了用于 XML 信息交换的 Java API (JAXM) 技术、SOAP 消息以及在分布式应用中使用 XML 的内容。还讲述了如何使用 JSP 产生和处理 SOAP 消息以及 XML-RPC 调用的内容。本附录还谈到了在 J2EE 应用中如何使用 XML-RPC 的问题。另外,介绍了如何编写 SOAP 自定义串行转换器来处理 SOAP RPC 调用中传递的各种类型的数据。

本书中的约定



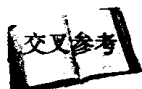
表示一种技巧,可以节省读者的时间。



表示与主题相关的实际情况或者有关的意见。



警告读者,如果不仔细就会出现问題。



表示可以在本书其他章节中找到更详细的解释信息。

XML 的使用

第 1 部分



本部分包括：

第 1 章

XML 概述

第 2 章

使用 DOM/SAX/JDOM/
JAXP

第 3 章

使用 XML 进行配置



XML 概述

第 1 章



本章包括：

理解 XML

识别 XML 文档的各个部分

使用 DTD 将约束置于 XML 文档

创建简单的 XML 文档

理解 XML 模式



在最近几年，可扩展标记语言（XML，eXtensible Markup Language）已经成为信息技术中最引人注目的技术之一。不管使用 XML 的是哪种平台或者应用程序，它都可以用来描述信息。

现在，XML 几乎用于每一个可能的行业中，比如，医疗、汽车制造以及多数应用（如信息交换、数据库管理和内容管理）中。另外，XML 还引起一些创建、编辑和转换它的新的应用产生。

在这一章中，将讨论 XML 及其特征。还将讨论如何创建 XML 文档以及 XML 语法和规则。

1.1 XML 简介

XML 非常简单，它是一种描述信息的标记语言，与使用它的应用程序无关；而其他的标记语言则需要特定的应用程序才可以使用。

HTML，VoiceXML，Word 和 Excel 都是使用标记语言所需的特定应用程序的文件格式或类型。需要特定应用程序的原因之一是这些标记语言的结构已经定义好并且不能改变。尽管 HTML 和 VoiceXML 的文件结构比 Word 和 Excel 的灵活性要高，但是仍然不能添加自定义的元素类型；事实上，类似 Word 这样的文件类型，不能改变定义文档如何布局的结构。



有一些标记语言是基于 XML 的，比如 VoiceXML，WML 和 XSL。在 VoiceXML 和 WML 中，XML 文档分别转换成 VoiceXML 或 WML 文档。在本书的第 2 部分将介绍有关转换的内容。

另一方面，XML 允许用户创建自己的结构。可以定义自己的标记和属性，从而以一种更有意义的方式描述信息。XML 可以与用逗号分隔的值（CSV，Comma Separated Value）做不精确的对比。CSV 用带有描述每个字段的标题的行和列构造数据。CSV 远非 XML 那样灵活或强大。例如，在 XML 中，可以描述不同元素之间的关系。不过，CSV 也有类似的用途：它们以一种中性格式提供信息。

为了进一步了解 XML，使用下面的样例。假设已经有一个有关计算机游戏（名为 XML Invaders）信息的属性集合。使用 XML 创建如下的 XML 文档：

```
<?xml version="1.0"?>
<games>
  <game year="2001">
    <name>XML Invaders</name>
  </game>
</games>
```

现在，先不用着急明白这段代码的所有含义。注意上述代码包含 3 个标记：games，game 和 name。其中，game 元素有一个属性 year，该属性的值为 2001。name 元素的值为 XML Invaders，该元素包含在 game 元素的起始和结束标记之间。

1.1.1 标记

标记描述代表数据的字符顺序。上述标记（包括看似古怪的以 <? 开始的标记）都被用于 XML 文档。但是还不止这些！可以更简洁地编写 XML 文档，代码如下所示：

```
<?xml version="1.0"?>
<games>
  <game year="2001" name="XML Invaders"/>
</games>
```

上述例子突出显示了 XML 有别于其他语言的主要优势之一：按照任何希望的方式构造数据的能力。如果公司的一些程序制作出 XML 文件，那么任何程序都可以读取和理解这些文件。这样，更方便很多人的工作，特别是应用程序开发人员，而且使得应用程序之间拥有前所未有的协作能力。例如，如果字处理程序制作出 XML 文件，那么很可能不用费神命名一堆文件格式，比如 Word，Star Office 或 Word Perfect。

1.1.2 XML 规则

XML 即使拥有这种强大的功能以及灵活性，但是仍然有一些规则负责管理 XML 文档的构造。这些规则允许应用程序模块（比如 XML 解析器）以一种方式读取它们。这一点类似于 Word 理解 Word 文档如何布局的方式。没有这些规则，XML 将会失去它的强大功能以及灵活性。

所有这些规则在 XML 规范中都有介绍，XML 规范由万维网联盟（W3C，World Wide Web Consortium）维护，W3C 的网址为：<http://www.w3c.org>。XML 规范规定了创建 XML 文档的规则和语法。

现在，读者对 XML 多少有了一些了解，并且大概看了一下 XML 文档。之后，我们将介绍管理 XML 文档构造的一些规则。