



ORACLE®

Oracle 技术系列 丛书

Oracle

XML 开发手册

Oracle XML
Handbook

Oracle Press™ 授权出版

内容涉及
Oracle 8i, 8, 7x

(美) Ben Chang 等著

高波 王琰 等译



机械工业出版社
China Machine Press

OSBORNE

Oracle 技术系列丛书

Oracle XML 开发手册

(美) Ben Chang 等著

高波 王琰 等译



机械工业出版社
China Machine Press

本书介绍如何在Oracle环境下创建强大的XML应用程序。本书由Oracle XML 开发小组编写，权威性地阐述了如何创建XML文档以及与电子商务创建者集成数据系统。示例丰富、内容全面。随书赠送的光盘包含JDeveloper 3.1和Oracle Business Components。

Ben Chang, et al: Oracle XML Handbook (ISBN 0-07-212489-x).

Original edition copyright © 2000 by McGraw-Hill. All rights reserved.

Chinese edition copyright © 2001 by China Machine Press. All rights reserved.

本书中文简体字版由美国麦格劳-希尔公司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。

版权所有，侵权必究。

本书版权登记号：图字：01-2000-3125

图书在版编目（CIP）数据

Oracle XML使用手册/（美）常（Chang, B.）著，高波等译。—北京：机械工业出版社，2001.1

（Oracle技术系列丛书）

书名原文：Oracle XML Handbook

ISBN 7-111-08608-2

I. O… II. ①常… ②高… III. 可扩充语言, XML-应用-关系数据库-数据库管理系统, Oracle-手册 IV. TP311.138-62

中国版本图书馆CIP数据核字（2001）第00687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：瞿静华

北京昌平第二印刷厂印刷·新华书店北京发行所发行

2001年1月第1版第1次印刷

787mm × 1092mm 1/16 · 15印张

印数：0 001-5 000册

定价：45.00元（附光盘）

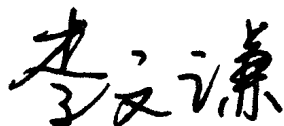
凡购本书，如有倒页、脱页、缺页，由本社发行部调换

贺 辞

祝贺中文版 ORACLE 8 系列丛书的出版。希望她能为中国的广大 ORACLE 用户和对数据库技术感兴趣的读者提供最先进的 ORACLE 技术知识。

ORACLE 软件系统有限公司希望通过中文版 ORACLE 8 系列丛书的出版，更好地建立起与广大用户和数据库工作者之间技术交流的桥梁。

Oracle 中国有限公司总经理

Handwritten signature in black ink, reading '李文谦' (Li Wenqian).

1998.6

序

数据库技术的发展，使它已经成为现代信息技术的重要组成部分，成为现代计算机信息系统和计算机应用系统的基础和核心。可以说，如果没有数据库技术的发展，没有优秀的数据库产品的推出和应用，社会信息化的进程将是难以实现的。因此，在衡量一个国家信息化的程度时，其数据库的建设规模、数据库信息量的大小和使用程度也就成为重要的标志之一。

我国引进数据库技术始于70年代末，从微型计算机上运行的数据库到当前的大型数据库系统的引入和应用，已经有20多年的历史。20多年来，虽然在微型计算机数据库知识的普及和应用上取得了很大的进展和成绩，在大型数据库系统的开发和应用上也取得了进步，甚至还有了国产化的数据库软件，但如果从对数据库系统的应用效果和对数据库技术的掌握上来比较，则与发达国家之间仍然存在较大的差距，特别是在大型数据库系统的开发、建设和应用水平方面差距更大。这种差距主要表现在两个方面：第一是数据库的数量及所收集信息的数量和质量与发达国家相比存在很大的差距；第二是对数据库技术和产品的掌握和应用上更有待于提高和加强，应用人才急需培养，经验有待积累和总结。前者的改善应依靠于对数据库应用基础工作的加强，如重视基础数据的收集和整理，即重视数据工程的建设，并制定相应的数据政策；而后者的改善则会更多地依赖于人们对数据库技术和对数据库产品的掌握。为此，我们必须首先占有充足的资料并加以消化。如果数据库厂家和出版机构能提供较完整的、质量较高的技术资料 and 书籍，并为较多的数据库技术人员和应用人员所掌握，无疑将会促进国内数据库技术人才的成长并推动数据库应用水平的提高。

喜闻机械工业出版社华章公司与ORACLE出版公司合作，为配合ORACLE 8在中国的发行，由机械工业出版社买断了ORACLE出版公司出版的ORACLE 8系列丛书的中文简体字版的出版权，并组织国内从事ORACLE应用开发的科技人员和教学人员进行翻译出版，还邀请国内数据库专家对译稿进行了审定，以保证丛书在技术

上的权威性。无疑，这对大型数据库系统特别是ORACLE数据库系统的开发和应用将起到很好的推动作用。

众所周知，ORACLE公司推出的ORACLE 8是一种面向网络计算的数据库(the database for Network Computing)并支持对象关系模型的数据库产品。该系列丛书全面地介绍了ORACLE 8的功能和技术，具体书目见封底。

该丛书内容丰富，涵盖了大型数据库应用开发中的全部技术内容，有的资料，如《Oracle 8数据仓库分析、构建实用指南》、《新版Oracle 8故障解决手册》等都是在国内首次面世，很值得数据库技术工作者参考和阅读。

无疑，这套丛书应该有广泛的读者，它可供大型数据库系统，特别是ORACLE系统应用的系统管理员、应用程序员、系统分析员和设计人员以及广大用户学习和参考，也可做为ORACLE系统培训的教材和高等学校本科和研究生的学习参考书，同时也是掌握大型数据库系统理论和实践的好材料。

应感谢机械工业出版社的领导和组织翻译丛书的先生们和女士们，辛勤的劳动将为我国信息化事业的发展增加动力。ORACLE中国有限公司对丛书的出版给予极大的关注与支持，李文谦总经理并为丛书的出版题词。

预祝丛书的出版和发行获得成功，并得到读者的欢迎。

罗晓沛

中国科学院研究生院 教授
ORACLE大学(中国) 校长

作者简介

Ben Chang 在Oracle公司有10年资深经历，他是CORE与XML开发组的领导者。除了在Oracle6到Oracle8i的发行版本上的工作外，他任职时间最长的职务是Oracle8.0的开发发布经理，在任期期间，发布了五个版本。他还在Oracle的C语言编码标准委员会担任过3年的主席职务，其间发布了该标准的两个主要的修正版；同时，他还是W3C DOM工作组委员会的成员。在加入Oracle公司前，他曾在IBM公司、太平洋Bell公司、Bellcore公司、以及GE R&D公司工作过。他在加州大学Berkeley分校获得电子工程与计算机科学专业的学士学位，在斯坦福大学获得电子工程（计算机系统）专业的硕士学位。

Mark Scardina 是Oracle公司在服务器端产品上对XML的热心提倡者。他是CORE与XML开发组的小组产品经理，任务是提供XML的底层组件，所有的Oracle产品包括XML开发包在内都会使用这些组件。他同时还是Oracle公司在W3C DOM与XSL工作组和SunJavaX专家组的代表。在加入Oracle公司前，他在Socket Communications与ACE Technologies工作过。他在旧金山大学获得管理信息系统专业的学士学位。

K.Karun 是Oracle公司CORE与XML开发组的开发经理。他在印度马德拉斯理工学院获得学士学位，在密歇根州大学获得计算机科学专业的硕士学位。

Stefan Kiritzov 在Oracle公司CORE与XML开发组中是主要的技术人员。他在开发系统软件方面有着18年的经验。在加入Oracle公司前，他为AT&T、NCR、SHL Systemhouse以及ICT等公司工作过。他在保加利亚Kliment Ohridsky大学获得数学专业的学士及硕士学位。

Ian Macky 在Oracle公司CORE与XML开发组中也是主要的技术人员。在该行业中，他有17年的经验，以前曾在SRI International和麻省理工学院人工智能实验室工作过。他在加州大学Santa Cruz分校获得计算机信息系统专业的学士学位。

Angel Novoselsky 在Oracle公司工作，是CORE与XML开发组的主要技术成员。在加盟Oracle之前，他作为程序员在美国、加拿大、保加利亚等地工作过。他在保加利亚Sofia Technical大学获得计算机专业硕士学位。

Niranjan Ramakrishnan 曾是Oracle公司CORE与XML开发组的高级技术人员。他在印度Pilani的Birla Institute of Technology and Science获得电子工程方面的学士学位，在密歇根大学获得计算机科学与工程专业的硕士学位。

菲利普序

许多年以来，各个公司一直都在计算机上大量投入钱财，但其投资并没有显示出什么回报。假定Spacely Sprockets公司要从Acme Widgets公司购买零件。Spacely Sprockets有高级的计算机化购置系统。从事购置工作的雇员通过一些在线表单指定：Spacely公司需购2500个零件，Spacely公司的区域号码为W147，Acme公司的零件型号为A491，要求在7月1日将货送到。

Acme同样是个现代化的公司。他们有着集成化的订单记录、存货清单，以及结账系统，该系统以关系数据库管理系统（RDBMS）为后端。订单刚进入系统，系统就开始安排在工厂中同步进行生产。这2500件零件的订单的数据是怎样从Spacely公司传送到Acme公司的呢？每个年代都有其特定的技术：

- 在70年代，Spacely公司的雇员首先从先进的Spacely系统中打印出订单，再把订单装进信封邮寄到Acme公司。Acme公司的雇员打开信封把订单键入到Acme系统中，打字时把2500打成了25 000。
- 在80年代，Spacely公司的雇员首先从先进的Spacely系统中打印出订单，再把订单传真给Acme公司。Acme公司的雇员由传真机的输出中拿起订单，把订单键入到Acme系统中，打字时把2500打成了25 000。
- 在90年代，Spacely公司的雇员首先从先进的Spacely系统中抽出订单，再把订单通过email发给Acme公司。Acme公司的雇员从他（她）的收件箱中得到订单后，把订单重新键入到Acme系统中，打字时把2500打成了25 000。

如果所有的这些听起来与你所熟悉的现实的商务世界同样的低效，那么请记住当Acme公司为25 000件零件向Spacely公司发出账单的时候，同样的过程会再次重复。

为什么Spacely公司的计算机无法与Acme公司的计算机直接进行交流？

在过去（没有Internet）的日子里，我们会说把Acme公司的信息送到Spacely公司是不可能的。现在，我们有了Internet，世界上的任何一台计算机都可以与其他计算机进行交流。但令人沮丧的事实是：它们没有东西可以交流。

XML是对结构化数据的具有一致性的书面表示，它是朝着解决上述问题的方向迈出的重要一步，同时它也保证社会在IT业上的投资可以得到回报。类似于XML的新标准，都会有一群软件生产商，追逐并渴望通过出售新软件而致富。买软件的人看起来从来不会自问一下诸如此类的问题：“为什么我为了使用该标准需要更多的软件？”或“全天候地维护该附加软件系统，成本会是多大？”以及“如果我对于已采用的软件还不能完全理解，那为什么还要增加新的呢？”。

许多大的组织都基于RDBMS进行经营活动。而Oracle是最常用的RDBMS。因此，现在所建立的大部分信息系统都需要由Oracle RDBMS服务器来提供全天候的服务。Oracle XML

VIII

Handbook 之所以很有价值，是因为它可以教会人们怎样利用标准的Oracle系统来建立支持XML的灵活的信息系统，而对于前者，大多数组织都已经掌握了怎样进行安装和维护。

用户有权利要求信息系统是可靠的，而最可靠的系统是那些变化部分最少的系统。本书将说明在不向贵组织的计算机增加软硬件的情况下，如何建立世界级的XML应用程序。

Philip Greenspun

2000年5月于马萨诸塞州剑桥

前 言

1999年3月，Philip Greenspun在Oracle公司总部举行了一次讲座，正是这次讲座给了作者写作本书的灵感。他的讲座涵盖了Internet、XML以及怎样生成以数据库为后端的Web站点。为什么Oracle数据库的使用在业界如此流行，以及为什么他认为Oracle开发组成员应该多为Oracle出版社写一些书等。

作者由图书提纲开始，该提纲是关于他们在Oracle公司所做的XML方面的工作；然后编写各个章节。他们尽可能地听从了Philip的建议，把本书按问题域组织起来，而不是按技术或部门体系。

时间很快过去了一年，到了2000年中期，他们最终完成了本书。在此期间，他们在对Oracle的工作投入大量时间的同时尽可能保持着本书的进度，经历了一段真正的紧张生活。真的，任何一位作者，他们的家庭，以及Oracle出版社的编辑们，都不会忘记这段经历的。XML万岁！

注意 本书的一部分销售收入将资助ArsDigita基金会和America's Promise两组织，以支持他们尝试推动教育进步的可敬事业。他们的Web站点分别是<http://arsdigita.org>和<http://www.americaspromise.org>。

应用中的XML

你之所以阅读本书，目的可能是计划在现实的应用中使用XML。这种想法很好，因为XML并不是应用，它只是一种使能性的技术。为帮助你建立具有XML功能的应用，我们在合适之处介绍了大量不同的应用背景、特定的代码实例以及XML和XSL实例文件。你可以自由地使用这些代码，我们希望它可以引导你有效地使用XML来增强你的应用。

注意 如果你对本书或CD要发表评论或在其中发现了错误，请不要犹豫，务请发e-mail到 oraclexmlhandbook@yahoo.com。

本书面向的读者

本书是Oracle XML组件的用户指南。它并不是完全参考手册，因为所有的组件API文档都可以通过电子版得到，而且，如果把这些文档加进来，会使本书的规模加倍。参考文档可以在Oracle技术网络（Oracle Technology Network）上得到。本书的读者对象主要有两类：一类是使用Oracle进行应用开发并且需要开始使用XML的程序员；另一类是有兴趣学习怎样在Oracle产品中使用XML的读者。书中许多例子使用了Java，通过使用组件的命令行界面和简单的可以编辑XML、XSL、XSQL文件的文本编辑器，就可以试验XML的大量功能。

怎样使用本书

不必按顺序阅读本书。本书的第1、2章是XML和Oracle的XML技术的介绍性材料，如果你熟悉这些，可以跳过这两章。那些对建立具有XML功能的Oracle8i数据库应用感兴趣的读者，可以阅读第3、7章，以及第8章中对OTN应用的深入解释。如果你打算使用Oracle应用服务器、Internet文件系统(Internet File System)、以及interMedia Text（一种文本搜索引擎），那么请阅读第4章、第5章、第6章。第9章展望了XML的未来。附录A给出了一些基本的W3C定义。各章的概要如下：

第1章介绍了Oracle在XML工业技术方面做出的努力，解释了XML的基本概念与术语、XML存在的原因以及Oracle在业界的XML战略。同时本章也给出了Oracle技术网络及其XML链接、Oracle的具有XML功能的产品和XML组件的一览表，最后是一个具有XML功能的关于书籍出售的应用实例。

第2章介绍了Oracle的XML开发包（XML Developer's Kit,XDK）组件，回顾了Java/C/C++/PLSQL等语言的XML解析器及XSLT处理程序、Java/C++语言的类生成器以及Java TransViewer组件。用于访问及处理XML/XSL文件的可工作代码也包括在本章中。

第3章涵盖了Oracle8i内建的Java虚拟机和Oracle的Java XML组件。同时也介绍了怎样利用XDK的XSQL Servlet存储XML数据到数据库中或从数据库中获取XML数据。最后，介绍了具有XML功能的售书应用的设计与实现。

第4章涵盖了Oracle应用服务器的体系结构及其与XML和XDK组件的联系。介绍了把书店应用作为QAS Servlet进行生成的实现。

第5章介绍了Oracle的下一代文件系统——iFS。这一章涵盖了iFS的体系结构，它作为XML文件系统的行为特性，以及它怎样在内部使用XML来把文件数据映射到关系模式。本章中还包括了关于通用文件操作的几个代码实例。

第6章涵盖了Oracle8i的文本搜索引擎——interMedia Text，回顾了其体系结构及搜索已存储的XML文档的能力。本章中的实例代码用于扩展书店应用，如产生索引并进行基于XML的搜索。

第7章涉及到怎样把XML组件组装到端到端解决方案中。本章介绍了一个具有XML功能的数据库应用的设计，其中包括用XML增强的Web站点和由商家到客户（business-to-client,B2C）的消息传递系统。上述XML和XSL实例文件以及实例代码都包含在本章中。

第8章介绍了Oracle技术网络（OTN）的关于XML的Web站点及其提供的XML示范。本章涉及怎样安装和运行这些示例，其中包括每个例子所示范的特定XML特性和能力方面的细节。

第9章回顾了W3C和OASIS在XML标准上所做的努力，并基于这些组织已有的工作推测了XML未来的发展方向。同时本章注意到了业界的几家主要公司使用XML及其相关技术的方式。

附录涵盖了主要的XML规范。它对W3C关于XML文档接口的两个主要标准——OM和SAX进行了讨论，同时也涉及到其他相关的XML技术标准如名字空间、XPath以及XSL变换等。

随书光盘与Web站点

你可以在随书光盘上找到Oracle XDK的产品版，它是相应于Java组件的，其中包括XML解

析器、XSL处理程序、TransView Java beans以及XSQL Servlet等，还有许多富于说明性的示范。你还可以找到另外一些产品的试用版，如Oracle8i Release 2、JDeveloper 3.1以及Oracle的具有XML功能的Java开发环境的最早版本等。最后，我们把相应于Java和Web-to-Go Web服务器的业务组件也加入在内，来协助你开发端到端的商务应用。

源程序的代码也可以通过从Oracle出版社的Web站点下载而得到，网址是<http://www.osborne.com/oracle>。你也可以从Oracle技术网络的Web站点下载最新的XML组件，网址是<http://technet.oracle.com/tech/xml>。

本书译者包括：高波、王琰、郭旭、张君慧、王洁、杨燕、王蔚宏。

目 录

贺辞	
序	
作者简介	
菲利普序	
前言	
第1章 Oracle与XML	1
1.1 XML基本概念和术语	1
1.1.1 序言	2
1.1.2 文档类型定义	3
1.1.3 文档的主体	3
1.1.4 文档对象模型API	4
1.1.5 XML简化编程接口API	7
1.1.6 名字空间API	11
1.1.7 解析器API	14
1.1.8 扩展格式页语言变换API	14
1.2 为何使用XML	14
1.3 Oracle的XML战略	15
1.4 Oracle在XML工业技术上的努力	16
1.4.1 Oracle在W3C工作组委员会中	16
1.4.2 Oracle的XML开发包	17
1.5 Oracle技术网络与XML链接	17
1.6 具有XML功能的Oracle产品概览	21
1.6.1 提供XML API的Oracle产品	21
1.6.2 使用XML用于数据交换的Oracle产品	21
1.6.3 使用XML进行配置的Oracle产品	22
1.6.4 使用XML进行内容管理和出版的Oracle产品	22
1.7 Oracle XML组件的使用概览	23
1.8 实例与应用	24
第2章 Oracle XML核心技术	27
2.1 Java XML解析器版本2	27
2.1.1 对SAX的支持	27
2.1.2 对DOM的支持	30
2.1.3 对XSLT的支持	33
2.2 Java类生成器	38
2.2.1 输入的DTD	38
2.2.2 处理DTD来生成Java类	39
2.2.3 由Java类生成合法的XML文档	39
2.2.4 由Java应用生成的XML文档	41
2.3 用Java浏览和转换XML	41
2.3.1 DOM建立器组件	42
2.3.2 XSL转换器组件	44
2.3.3 XML源程序浏览器组件	44
2.3.4 XML树浏览器组件	48
2.3.5 XSL TransViewer组件	48
2.4 PL/SQL版本的XML解析器	51
2.5 C版本的XML解析器	54
2.5.1 独立的解析器	54
2.5.2 解析器库	55
2.5.3 文档对象模型API	58
2.5.4 XML简易编程接口	59
2.6 C++版本的XML解析器	63
2.7 C++版本的类生成器	64
第3章 为Oracle8i数据库开发应用	66
3.1 Oracle8i——具有XML功能的数据库	66
3.2 Oracle JServer与Java XML组件	67
3.2.1 JServer基本知识	67
3.2.2 Java XML组件	69
3.2.3 发布和调用Java XML组件	70
3.3 数据库模式与XML文档	72
3.3.1 把XML文档映射到数据库模式	73

3.3.2 把数据库模式映射到虚拟XML文档	74	5.9.1 类型定义示例	107
3.3.3 存储和获取XML数据	76	5.9.2 标准类型属性	108
3.4 XSQL——XSLT/SQL服务器页面	76	5.9.3 定制类型属性	109
3.4.1 XSQL页面的体系结构	78	5.9.4 属性特性	109
3.4.2 XSQL Servlet的安装	79	5.10 文件扩展	112
3.4.3 SQL查询生成的动态XML文档	79	5.11 应用iFS	113
3.4.4 XSQL中对XSLT标记的支持	81	5.11.1 例1——创建并保存Hello World	113
3.5 售书的例子	81	5.11.2 例2——创建一个更好的 Hello World	114
3.5.1 设计数据库模式	82	5.11.3 例3——用文件来工作	115
3.5.2 使用XSQL来设计Web站点	83	5.11.4 例4——搜索文件	115
第4章 Oracle应用服务器开发	88	第6章 用interMedia Text搜索XML文档	117
4.1 Oracle应用服务器体系结构	89	6.1 Oracle interMedia Text——下一代文本 搜索引擎	117
4.1.1 HTTP接收器	89	6.2 Oracle interMedia Text索引结构	118
4.1.2 OAS组件	90	6.2.1 数据存储	118
4.1.3 应用程序组件	90	6.2.2 过滤器	119
4.2 用作服务器端小应用程序的OAS的 Bookstore应用程序	91	6.2.3 分段器	119
4.2.1 服务器端Bookstore应用程序	93	6.2.4 词法分析器	121
4.2.2 注册书库小服务程序和组件	96	6.3 用interMedia Text工作	121
4.2.3 调用服务器端Bookstore应用程序	96	6.3.1 数据存储	122
4.2.4 数据库访问	97	6.3.2 字段和区段	124
4.2.5 使用事务服务	100	6.3.3 终止段	127
4.2.6 调用OAS组件	100	6.3.4 属性段	128
第5章 Oracle Internet文件系统	101	6.3.5 动态添加段	129
5.1 功能	101	第7章 Oracle与操作中的XML	130
5.2 优点	102	7.1 Oracle XML SQL Utility	130
5.3 组件	102	7.1.1 对XML格式的数据进行检索	130
5.4 XML	103	7.1.2 保存XML格式的数据	132
5.5 媒介文本	104	7.1.3 利用XML SQL Utility实现更新 功能	133
5.6 文档模型	104	7.1.4 利用XML SQL Utility实现删除 功能	134
5.7 文档属性	105	7.1.5 安装XML SQL Utility	135
5.7.1 无版本属性	105	7.1.6 拓展XML SQL Utility	135
5.7.2 标准属性	105	7.2 Oracle XSQL Servlet	136
5.7.3 链接属性	106	7.2.1 XSQL页面	136
5.7.4 用户属性	106		
5.8 文档处理	106		
5.9 定义文档类型	107		

7.2.2 安装XSQL Servlet	137	8.3.2 Employee Page程序	163
7.2.3 向XSQL Servlet呈交请求	138	8.3.3 Insurance Claim应用程序	165
7.2.4 用样式表对XSQL 输出进行转换	139	8.3.4 Invalid Classes应用程序	168
7.2.5 在XSQL Servlet中插入XML文档	141	8.3.5 XSQL范例索引	169
7.2.6 在XSQL Servlet中更新数据	142	8.3.6 “Do You XML?” 站点	171
7.3 具有XML功能的站点地址	144	8.3.7 Employec/Department Object View 应用程序	171
7.3.1 XML解决方案	144	8.3.8 Airport Code Validation应用程序	173
7.3.2 设计要求	144	8.3.9 Airport Code Display应用程序	177
7.3.3 结构	144	8.3.10 Ad hoc Query Visualization 应用程序	179
7.3.4 一个实现例子	145	8.3.11 XML 文档演示程序	179
7.3.5 拓展例子	147	8.3.12 XML插入请求演示程序	181
7.3.6 Oracle Portal-to-Go	147	8.4 安装和运行XML应用程序	183
7.4 电子商务的XML功能信息	148	第9章 未来趋势	187
7.4.1 XML解决方案	148	9.1 标准化组织的职责	187
7.4.2 设计要求	148	9.1.1 W3C的任务	188
7.4.3 结构	149	9.1.2 OASIS的作用	193
7.4.4 一个实现例子	149	9.2 工业界认可的XML模式	194
7.4.5 拓展例子	155	9.2.1 XML模式/DTD领域的主要使用者	195
7.4.6 Oracle Integration Server	155	9.2.2 一个特定工业部门DTD/XML模式 成就的例子	197
第8章 OTN提供的基于XML的应用程序	157	9.3 XML在互联网上的影响力	198
8.1 访问XML应用程序	157	附录 各种规范	201
8.2 XML应用程序说明了什么	158		
8.3 XML应用程序	160		
8.3.1 Hello World程序	160		

第1章 Oracle与XML

扩展标记语言 (XML) 是一种元标记语言。就像在广域网协会 (World Wide Web Consortium, W3C) 的XML 1.0规范中所说明的, XML可以让用户定义自己的标记语言, 从而可以在XML文件中描述并封装数据。这些文件可以在类似于Netscape Navigator或Microsoft Internet Explorer的浏览器中显示, 并通过因特网在应用之间或业务之间交换, 存储到数据库中或从数据库中取出。XML具有简单性, 它是开放标准的一部分, 加入了用户自定义的标记。

XML起源于通用标记语言 (Standard Generalized Markup Language, SGML), 后者于1986年由国际标准化组织 (International Standards Organization, ISO) 批准, 基于通用标记语言的超文本标记语言 (Hypertext Markup Language, HTML) 诞生于1990年。虽然在文档世界中SGML仍然是广为使用的标准, 而HTML作为广域网上数以百万计的Web页面的基础依然用途广泛, 但XML正在获得广泛的接受, 因为它与现有的标记语言相比, 在数据的交换、存储、描述等方面都具有优点。自从W3C在1998年2月出版XML 1.0版的规范以来, 在相当广泛的范围内, 许多人都认为它将成为电子商务在语言和数据交换方面的选择。

1.1 XML基本概念和术语

类似于所有的标准, XML也有数不清的概念和技术术语需要解释。由于开发XML是用来传送数据的, 因此, 介绍一个例子, 看一下标准的数据库中书籍列表的数据记录。一个复杂的SQL查询典型情况下将会返回下述格式的数据:

```
History of Interviews, Juan, Smith, 99999-99999, Oracle Press, 2000.
```

如果XML用作输出表单, 则该记录的每个数据项都将具有附加的上下文环境, 如下所示:

```
<book>
  <title>History of Interviews</title>
  <author>
    <first name>Juan</first name>
    <last name>Smith</last name>
  </author>
  <ISBN>99999-99999</ISBN>
  <publisher>Oracle Press</publisher>
  <publish year>2000</publish year>
  <price type="US">1.00</price>
</book>
```

本例中某些值得注意的项将在稍后进行探讨。可以注意到文件具有对称性, 且每个数据块都被其上下文所包围, 形式类似于<context>...</context>。尖括弧及其内部的文字称为标记 (tag), 每组标记及其包围的数据称为元素。这种关系可以认为类似于数据库中表的一列, 其中标记的文字相当于列标题, 标记之间的文字相当于该列中某一行的数据。在前一例中, title应该是列名,

而History of Interviews应当是某一行中的数据。

还可以注意到，有几个标记包含的并非数据而是标记。这是XML一个很重要的特性，它允许数据的嵌套以便更好的定义关系。返回到数据库的意义上，<author>标记可以用表来模拟，而该表的列则是<first name>和<last name>。在XML的术语中，这些列标记被称为表名标记，即父标记Author的子标记。

现在注意<price>标记，你可以看到它以“名字=‘值’”的形式包含了文字。这些“名——值”对叫做属性，开始标记中可以包括一个或多个属性。属性出现在结束标记中是非法的（例如，</tag name=“foo”>）。

下面是最后一个值得注意的术语，显然整个XML实例是被<book>...</book>所包围的。这种标记定义为文档的根，在整个文档中只能有一个。若XML文档遵循这样的规则：只有一个根标记并且所有打开的标记都正确地关闭，则称该文档是构造良好（well formed）的，也称良构。

XML的基本概念和术语是简单的，且已经形式化为开放的因特网标准。像W3C的XML 1.0规范中所说的，“XML文档由称为实体的存储单元组成，实体中包含分析过或未分析的数据。分析过的数据由字符组成，其中一些形成字符数据而另一些形成标记。标记对文档的存储布局和逻辑结构进行描述。”XML文档同时具有物理结构和逻辑结构。XML文档的物理结构简单地指向XML文件本身及可能引入的其他文件，而XML文档的逻辑结构则指向文档的序言（prolog）和主体（body）。

图书列表的XML实例显示了XML文档的主体，但它缺失了有助于识别其性质的重要信息。该信息在文档的序言中，将在下一节定义。

1.1.1 序言

序言由XML的声明组成，包括：版本号、可能的语言编码、其他属性（形如“名字=‘值’”的对）、可选的文档类型定义（Document Type Definition, DTD）等，DTD可以是内部的，即包含在XML文档中，也可以是外部的，即指向另一个文件。例如：

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<!DOCTYPE book SYSTEM "book.dtd">
```

请注意，形如<?...?>的行是XML中的处理指令（processing instruction, PI），将在本章中稍后讨论。在这个例子里，XML处理指令的名字就是xml。另外，本例所支持的字符集编码是Unicode的压缩版本，称为UTF-8。最后，standalone属性是指处理程序是否需要包含或引入其他的外部文件。

序言的第二行使用了DOCTYPE。该文档在这里声明数据模型。它为什么如此重要？请记住，XML文件同时具有物理和逻辑描述。在某些应用中，即使不知道信息是否缺失，也可能足以处理XML。但在大多数情况下，应用需要验证其收到的XML文档的合法性。要这样做，应该需要知道哪些元素是必需的，哪些可以有子标记，哪些可以有属性，等等。用XML的术语来说，数据模型称为文档类型定义。文档类型定义可以包含在XML文件中，也可以简单地顺次引用，这样处理程序就可以对其进行定位。