

坚果系列



XML

技术手册

XML in a Nutshell

O'REILLY®
中国电力出版社

Elliott Rusty Harold & W. Scott Means 著

孔小玲 商艳莉 译



XML

技术手册

XML in a Nutshell

Elliote Rusty Harold &
W. Scott Means 著

孔小玲 商艳莉 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Paris • Sebastopol • Taipei • Tokyo

O'Reilly & Associates, Inc. 授权中国电力出版社出版

中国电力出版社

图书在版编目 (CIP) 数据

XML 技术手册 / (美) 哈诺尔德 (Harold, E.), (美) 明斯 (Means, W.) 编著; 孔小玲, 商艳莉译. - 北京: 中国电力出版社, 2001.11

书名原文: XML IN A NUTSHELL

ISBN 7-5083-0808-5

I .X... II . ①哈... ②明... ③孔... ④商... III . 可扩充语言, XML - 手册 IV .TP312
中国版本图书馆 CIP 数据核字 (2001) 第 074039 号

北京市版权局著作权合同登记

图字: 01-2001-4626 号

© 2001 by O'Reilly & Associates, Inc.

Simplified Chinese Edition, jointly published by O'Reilly & Associates, Inc. and China Electric Power Press, 2001. Authorized translation of the English edition, 2001 O'Reilly & Associates, 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.

英文原版由 O'Reilly & Associates, Inc. 出版 2001。

简体中文版由中国电力出版社出版 2001。英文原版的翻译得到 O'Reilly & Associates, Inc. 的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly & Associates, Inc. 的许可。

版权所有、未得书面许可, 本书的任何部分和全部不得以任何形式重制。

书 名 / XML 技术手册

书 号 / ISBN 7-5083-0808-5

责任编辑 / 关敏

封面设计 / Ellie Volckhausen, 张健

出版发行 / 中国电力出版社

地 址 / 北京三里河路 6 号 (邮政编码 100044)

经 销 / 全国新华书店

印 刷 / 北京市地矿印刷厂

开 本 / 787 毫米 × 1092 毫米 16 开本 33.5 印张 488 千字

版 次 / 2001 年 11 月第一版 2001 年 11 月第一次印刷

印 数 / 0001-5000 册

定 价 / 59.00 元 (册)

5083

O'Reilly & Associates 公司介绍

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 & Associates 公司授权中国电力出版社,翻译出版一批该公司久负盛名的英文经典技术专著。

O'Reilly & Associates 公司是世界上在 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,同时是联机出版的先锋。

从最畅销的《The Whole Internet Use's Guide & Catalog》(被纽约公共图书馆评为二十世纪最重要的 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 & Associates 一直处于 Internet 发展的最前沿。

许多书店的反馈表明,O'Reilly & Associates 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 & Associates 公司具有深厚的计算机专业背景,这使得 O'Reilly & Associates 形成了一个非常不同于其他出版商的出版方针。O'Reilly & Associates 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 & Associates 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 & Associates 依靠他们及时地推出图书。因为 O'Reilly & Associates 紧密地与计算机业界联系着,所以 O'Reilly & Associates 知道市场上真正需要什么图书。

译者序

XML (Extensible Markup Language, 可扩展标记语言) 是由 W3C (World Wide Web Consortium, 万维网联盟) 于 1998 年 2 月发布的一种标准, 与 HTML 一样是 SGML (Standard Generalized Markup Language, 标准通用标记语言) 的一个简化子集。由于它将 SGML 的丰富功能与 HTML 的易用性结合到了 Web 应用中, 因而自推出以来迅速得到软件开发商的支持和程序开发人员的喜爱, 显示出强大的生命力。

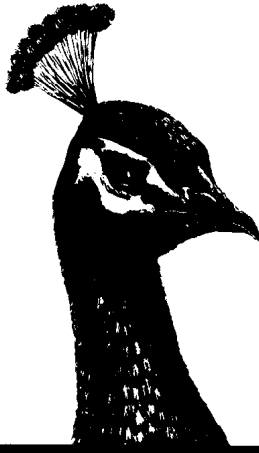
由于 XML 较好地解决了 HTML 无法表达数据内容等问题, 从而在政府、金融、证券、邮电、保险、税务、司法、出版以及电子商务等方面均得到了广泛的应用。

本书是全面介绍 XML 技术的一部权威性宏篇巨著, 内容全面详实, 并且包含大量示例应用程序和代码段, 全面阐明和展示了 XML 的使用方法。

本书共分四个部分二十三章。第一部分介绍组成内核的基本标准。第二部分讨论主要用于像网页、图书、文章、日记和剧本等叙述型 XML 文档中的技术。第三部分讨论了数据密集型文档中的 XML 应用。最后, 第四部分包括一系列组成任何“坚果系列”手册核心的速查章节。

本书第一部分、第二部分由商艳莉翻译, 第三部分、第四部分由孔小玲翻译。由于译者水平有限, 翻译中的不妥之处在所难免, 敬请读者批评指正。

译者



XML

技术手册

XML in a Nutshell

目录

前言	1
----------	---

第一部分 XML 概念

第一章 XML 概述	11
XML 提供了什么	11
可移植的数据	14
XML 如何工作	14
XML 的发展史	16

第二章 XML 基础	19
XML 文档和 XML 文件	19
元素、标签和字符数据	20
属性	24
XML 名称	25
实体引用	26
CDATA 部分	28

注释	28
处理指令	29
XML 声明	30
文档的合式性检查	32
第三章 DTD	35
验证	35
元素声明	43
属性声明	49
通用实体声明	57
外部已析通用实体	58
外部未析实体和记法	60
参数实体	62
条件嵌入	64
两个 DTD 例子	65
查找标准 DTD	68
第四章 名称空间	69
名称空间的必要性	69
名称空间的语法	72
解析器如何处理名称空间	78
名称空间和 DTD	78
第五章 国际化	81
编码声明	82
文本声明	82
XML 定义的字符集	83
Unicode	84
ISO 字符集	86
依赖于平台的字符集	88

字符集之间的转换	89
XML 文档缺省字符集	90
字符引用	91
xml:lang	94

第二部分 叙述型文档

第六章 作为文档格式的 XML 99

SGML 的遗产	99
叙述型文档结构	100
TEI	102
DocBook	105
文档永久性	108
转换和表示	110

第七章 Web 中的 XML 113

XHTML	114
在浏览器中直接显示 XML	121
使用模块化 XHTML 编写复合文档	126
改善的网页搜索方法前景展望	140

第八章 XSLT 145

一个输入文档举例	145
xsl:stylesheet 和 xsl:transform	146
样式表处理器	148
模板	149
使用 xsl:value-of 计算元素值	150
使用 xsl:apply-templates 应用模板	151
内置模板规则	154

模式	158
属性值模板	160
XSLT 和名称空间	160
其他 XSLT 元素	162
第九章 XPath	164
XML 文档的树结构	164
定位路径	166
复合定位路径	172
谓词	174
非简写定位路径	175
通用 XPath 表达式	178
XPath 函数	181
第十章 XLink	187
简单链接	188
链接行为	189
链接语义	192
扩展链接	192
链接基准	200
XLink 的 DTD	201
第十一章 XPointer	202
URL 中的 XPointer	202
链接中的 XPointer	204
裸名称	205
子序列	206
点	206
范围	209

第十二章 CSS..... 212

CSS 的三个级别	214
CSS 语法	214
使样式表与 XML 文档相关联	216
选择符	218
显示属性	221
像素、磅、pica 及其他长度单位	223
字体属性	224
文本属性	226
颜色	227

第十三章 XSL-FO..... 229

XSL 格式化对象	231
XSL-FO 文档的结构	232
主页面	233
XSL-FO 属性	239
在 CSS 和 XSL-FO 之间做选择	244

第三部分 数据型文档

第十四章 作为数据格式的 XML..... 249

XML 的编程应用	249
描述数据	251
对程序员的支持	253

第十五章 编程模型..... 255

事件驱动和对象驱动模型的对比	255
编程语言支持	256

非标准扩展	257
转换	258
处理指令	258
链接和引用	259
记法	259
所得并非所见	260
第十六章 DOM	261
DOM Core	262
DOM 的优点和缺点	262
用 DOM 解析文档	263
Node 接口	263
具体的节点类型	264
DOMImplementation 接口	271
一个简单的 DOM 应用程序	271
第十七章 SAX	276
ContentHandler 接口	278
SAX 特性和属性	286
第四部分 参考	
第十八章 XML 1.0 参考	291
如何使用本参考	291
带注释的样本文档	292
XML 语法的关键	292
合式性	294

有效性	299
全局语法结构	307
DTD	313
文档主体	323
XML 文档语法	324
 第十九章 XPath 参考	328
XPath 数据模型	328
数据类型	329
定位路径	330
谓词	334
XPath 函数	335
 第二十章 XSLT 参考	346
XSLT 名称空间	346
XSLT 元素	346
XSLT 函数	375
 第二十一章 DOM 参考	381
对象层次	382
对象参考	383
 第二十二章 SAX 参考	447
org.xml.sax 包	447
org.xml.sax.helpers 包	454
SAX 特性和属性	461
org.xml.sax.ext 包	463

第二十三章 字符集	465
字符表	468
HTML4 实体集	472
其他 Unicode 块	484
 词汇表	 511

前言

在信息处理技术的发展历史中，XML 是有关文档语法最重要的发展之一。近年来，XML 已经广泛应用于各个领域，如，法律、航空、金融、保险、自动机、多媒体、旅店、旅游、艺术、建筑、电信、软件设计、农业、物理、新闻、宗教、零售和中世纪文学等。XML 已经成为几乎跨越所有应用的新文档格式的精选语法。可用于 Linux、Windows、Macintosh 以及其他计算机平台上。华尔街上的计算机之间通过交换 XML 文档进行股票交易。在家用 PC 机上玩游戏的孩子们用 XML 保存自己的文档。体育爱好者利用 XML 在移动电话上实时接收比赛信息。XML 是有史以来最健壮、最可靠、最灵活的文档语法。

本书可为你走向飞速发展的 XML 世界提供综合性指导。包括 XML 的各个方面：从最基本的语法规则到 DTD 的创建细节，以及在多种编程语言中可以用来读写 XML 文档的 API 函数。

本书包含哪些内容

W3C (World Wide Web Consortium, 万维网联盟) 和其他标准组织，如 OASIS 和 OMG (Object Management Group, 对象管理小组)，正式开发了数百种 XML 应用。还有一些个人和公司也开发了非正式、非标准的应用，例如微软的 CDF (Channel Definition Format, 通道定义格式) 和 John Guajardo 的 Mind Reading Markup Language (Mind 读取标记语言)。对此本书无法一一列举，正如一本关于 Java 的书籍无法对每一种已经或者可以用 Java 写的程序都予以讨论一样。本书主要是针对

XML 本身，内容包括所有 XML 文档和作者必须遵循的基本规则，无论是网页设计者使用 SMIL 为网页添加动画，还是 C++ 程序员使用 SOAP 将对象串行化到远程数据库中。

本书还包括跨越 XML 并且广泛用于各种 XML 应用中的通用支持技术。包括：

XLink

一种基于属性的用于 XML 和非 XML 文档之间超链接的语法，可以提供与 HTML 相似的简单、单向的链接，多文档之间的多向链接，以及没有写入权限的文档间的链接。

XSLT

一种 XML 文档，用于描述具有相同或不同 XML 词汇表的两个文档之间的转换。

XPointer

标识 XML 文档中由 URI 指定的特殊部分；通常与 Xlink 结合使用。

XPath

XPointer 和 XSLT 使用的一种非 XML 语法，用于标识 XML 文档的特殊部分。例如，XPath 能够定位文档中的第三个 address 元素，或者 email 属性值为 elharo@metalab.unc.edu 的所有元素。

Namespace (名称空间)

区分来自不同 XML 词汇表但名称相同的元素和属性的一种方式；例如，书名和关于书籍的网页的标题等。

SAX

用于 XML 的简单 API (Simple API for XML)，许多 XML 解析器都已实现的一个基于事件的 Java 应用程序编程接口。

DOM

文档对象模型 (Document Object Model)，一个面向树的 API，它将 XML 文档当做具有多种属性的嵌套对象的集合。

所有这些技术，无论是用 XML (XLink、XSLT 和名称空间) 还是用其他语法 (XPointer、XPath、SAX 和 DOM) 定义的，都可用于各种不同的 XML 应用中。

本书不涉及只与某些 XML 用户有关的 XML 应用。包括：

SVG

可缩放矢量图形 (Scalable Vector Graphic)，是 W3C 推荐的 XML 中矢量图的编码标准。

MathML

数学标记语言 (Mathematical Markup Language), 是 W3C 推荐的标准 XML 应用, 用来在网页或其他文档中嵌入公式。

CML

化学标记语言 (Chemical Markup Language), 是最早的 XML 应用之一。描述化学、固体物理学、分子生物学和其他分子科学。

RDF

资源描述框架 (Resource Description Framework), 是 W3C 推荐的 XML 应用, 用来描述资源, 特别是图书馆分类卡上的元数据。

CDF

通道定义格式 (Channel Definition Format), 是微软公司定义的一种非标准 XML 应用, 用来向 Internet Explorer 发布可离线浏览的网页。

有时我们会在例子中使用一个或多个上述应用, 但是不会对所有相关词汇表的各个方面都做深入探讨。虽然这些应用 (类似的有数百种) 都很有趣并且很重要, 但是它们主要用于那些了解这些应用具体格式的特殊软件中。例如, 图形设计者不直接使用 SVG; 相反, 他们使用像 Adobe Illustrator 这样的传统工具创建 SVG 文档, 他们甚至根本不知道自己在使用 XML。

本书着重介绍与几乎所有使用 XML 的开发者都相关的标准。我们所研究的 XML 技术涉及广泛的 XML 应用, 而不只局限于少数领域。

本书的组织

第一部分, XML 概念, 介绍基本核心标准的组成部分, 所有 XML 应用和软件都必须遵循这些标准。此部分可使你尽快掌握合式的 XML、DTD、名称空间和 Unicode。

第二部分, 叙述型文档, 讨论主要用于网页、图书、文章、日记和剧本等叙述型 XML 文档中的技术。我们将学习到 XSLT、CSS、XSL-FO、XLink、XPointer 和 XPath。

XML 发展中最为出人意料的一点是它在诸如电子表格、财务统计表、数学表格和软件文件格式等数据密集型的结构化文档中的广泛应用, 第三部分, 数据型文档, 将讨论数据密集型文档中的 XML 应用。此部分着重介绍如何编写处理 XML 软件所必需的工具和 API 函数, 包括 SAX 和 W3C 的 DOM。

最后, 第四部分, 参考, 包括一系列组成任何“坚果系列 (in a Nutshell)”手册核心的速查章节。各章将详细介绍核心 XML 技术的语法规则。这些技术包括 XML、