

洞悉 XML

集合多位顶级软件工程师对 XML 的开发经验，
通过精心设计的实例让您深入了解新一代因特网语言。



硕网资讯 编著
北大宏博 改编

北京大学出版社
<http://cbs.pku.edu.cn>

洞悉 XML

硕网资讯 编著
北大宏博 改编

北京大学出版社

内 容 简 介

XML 可扩展标记语言是一套规范, 允许各行业自行定义如 HTML 般的标注语, 以方便数据存取、处理、交换、转换等, 在近两年中获得迅猛发展。本书由 XML 简介揭开篇章, 深入探讨数据库设计技巧、准确剖析 XML Schema 及 DOM 等 XML 关键技术, 配合数据库应用实例与常见的 XML 应用, 以循序渐进的方式, 将读者引导至实际应用领域。本书几乎涵盖了 W3C 开发的全部 XML 规范, 每一章均提供丰富详尽的实践范例, 是了解 XML 技术的一部难得而全面的技术书籍, 书中所提供的丰富 XML 技术组织机构网址, 让您时时紧随 XML 的发展。本书适合程序员及计算机高级用户学习阅读参考。

著作权登记号: 图字 01-2001-2294

本书繁体字版名为《洞悉 XML》, 由第三波资讯股份有限公司出版, 版权归第三波资讯股份有限公司所有。本书简体字中文版由第三波资讯股份有限公司通过第三波出版国际股份有限公司授权北京大学出版社出版。专有出版版权属北京大学出版社所有, 未经本书原版出版者和本书出版者书面许可, 任何单位和个人不得以任何形式或任何手段复制或传播本书的一部分或全部。

图书在版编目(CIP)数据

洞悉 XML/硕网资讯编著; 北大宏博改编. -北京: 北京大学出版社, 2001.7

ISBN 7-301-05149-2

I.洞... II.①硕...②北... 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(2001)第 048805 号

书 名: 洞悉 XML

责任著作者: 硕网资讯

改 编 者: 北大宏博

责 任 编 辑: 王方明

标 准 书 号: ISBN 7-301-05149-2/TP · 0576

出 版 者: 北京大学出版社

地 址: 北京市海淀区中关村北京大学校内 100871

网 址: <http://cbs.pku.edu.cn>

电 话: 出版部 62752015 发行部 62754140 62765127 编辑室 62765126

电 子 信 箱: wdzh@mail.263.net.cn

排 版 者: 北京东方人华科技有限公司

印 刷 者: 河北省滦县印刷厂

发 行 者: 北京大学出版社

经 销 者: 新华书店

787 毫米×1092 毫米 16 开本 23.875 印张 570 千字

2001 年 8 月第 1 版 2001 年 8 月第 1 次印刷

定 价: 35.00 元

序

XML 是 Extensible Markup Language(可扩展标记语言)的缩写,它是一套规范,允许各行业自行定义如 HTML 般的标注语,以方便数据存取、处理、交换、转换等。XML 在近两年获得迅猛发展,对网络乃至工商界都造成了极大震撼,只要您在上网时稍加留意,XML 已经不是什么新鲜事物,针对 XML 规范及其应用的讨论已经是大众热衷且公开的事实,虽然它并未获得微软的正面响应,而且虽然此时 HTML 的应用也正如日中天。

HTML 是信息技术应用上的一个奇迹,由于其具有简单特性,加上制定了统一标准,使得 Web 技术从计算机界走向全社会,并走向千家万户。到目前为止,我们看到几乎所有 Web 页面都是用 HTML 编写的。HTML 简单易学又通用,句法简明紧凑,加上其扩充的表格、帧、脚本等功能,使它得以在 Web 主页上大显身手,由此开创的互联网信息传播新局面不能不说是信息传播技术应用上的奇迹。

但是,随着 Web 应用的越来越广泛,HTML 过于简单的弱点也越来越明显,它已经不能满足用户的需求了,这表现在 HTML 对格式的支持、信息应用的支持及数据交换等多方面的固疾——这已为越来越多程序员所垢病。正是由于这些缺点,人们开始转而研究能改进或替代 HTML 的 Web 页面制作语言。其中最有成效的,并已初步投入使用的技术有多种,其中,可扩展标记语言(Extensible Markup Language)——XML 就是一个生动的典型。

XML 作为新出现的一种 Web 创作语言,驱动了 Internet 应用程序编写方式的动态改变,现在已有越来越多的软件向它提供支持,也许在不久的将来它会成为 Internet 上唯一的语言。

XML 首先是一种元标记语言,所谓“元标记”就是开发者可以根据自己的需要定义自己的标记,任何满足 XML 命名规则的名称都可以标记,这就为不同的应用程序打开了的大门。同时 XML 是一种语义/结构化语言,它描述了文档的结构和语义。XML 可以定义信息的内容,却没有定义信息该如何表达,这实际上就是 XML 的长处,它把内容和形式分离了,同时,从一些成功的电子商务网站应用的例子我们看到,XML 简化了网络中企业到企业(B2B)的交易。因此,我们有理由相信,今后几年中我们将看到 XML 逐渐被大量地应用在各式各样的数据交换的环境中,包括(但不仅限于)电子商务和医疗保健等领域,因为 XML 非常适合用来做订单、病历乃至政府公文等的格式简化、一致化等。尤其令人兴奋的是,XML 有效结合了 SGML 和 HTML 的优点并消除其缺点,由 Sun

Microsystem 的 John Bosak(Sun 公司的在线信息技术设计师, 它说服了 W3C, 由他组织了一个关于在 Web 上使用 SGML 的委员会。在该年的 11 月, 他们完成了 SGML 简化版本的初始设计。这个简化版本保留了 SGML 主要的实用功能, 特别是与 Web 信息发布有关的功能, 同时大大缩减了 SGML 的复杂性, 这就是 XML。)始发的 W3C SGML 工作组(现在称为 XML 工作组)所要攻克的目标现在已经基本达到。

全书分 5 篇共 15 章, 由 XML 简介为首章揭开本书的序幕, 紧接着探讨数据库设计技巧、深入剖析 XML 技术, 例如: XML Schema、DOM、SAX、XSLT、XML Query、XLink、XPointer、XPath 及 XBase、Schema Repository 等关键技术, 再配合数据库应用实例与常见的 XML 应用, 以循序渐进的方式, 将读者引导至实际应用的领域。本书几乎涵盖了 W3C 开发的全部 XML 规范, 每一章均提供了丰富详尽的实践范例, 可以说是了解 XML 技术的一部难得而全面的技术书籍。由于 XML 的进取性, 这些规范也正在发展中, 当您读到这些内容的时候, W3C 可能已经提出了新版的标准, 用户如需更为详实的分析, 也可造访书中所提供的丰富网址, 它将让您的思想与发展中的 XML 一样年轻。

尤为令人注意的是, 本书集合了台湾数位软件工程师对 XML 的长年开发经验, 他们均受过良好教育, 各精所长, 当时同为供职于台湾地区 XML 引领者“硕网资讯”旗下, 因此, 书中涉及的实例无不是经过精心设计, 分外精彩, 这也许正是本书的最大特色所在吧!

目 录

第 1 篇 XML 简介

第 1 章 认识 XML	3
1.1 XML 的重要性.....	3
1.2 XML 的特点.....	4
1.3 XML 的优势.....	5
1.3.1 数据检索	5
1.3.2 XML 的检索功能	6
1.3.3 数据的显示	7
1.3.4 数据的交换	7
1.4 XML 的应用.....	7
1.4.1 内容管理	7
1.4.2 电子邮件的收发与管理.....	9
1.4.3 智能型日历	9
1.4.4 个人化信息服务	10
1.4.5 电子商务	10
1.5 XML 的限制.....	13
1.6 小结	14
第 2 章 W3C 与重要 XML 规范介绍	15
2.1 W3C 简介	15
2.1.1 W3C 组织结构	15
2.1.2 开发行动(Activity).....	16
2.1.3 建议规范(Recommendation)制定流程	17
2.1.4 常见名词一览表	19

2.2 重要 XML 相关规范.....	21
2.2.1 XHTML.....	21
2.2.2 XSLT.....	23
2.2.3 Xlink、Xpointer、Xpath、XML Base.....	24
2.2.4 MathML.....	24
2.2.5 SMIL.....	25
2.2.6 SVG.....	26
2.2.7 XML Schema.....	28
2.2.8 XML Query.....	28
2.2.9 XForms.....	28
2.2.10 DOM.....	29
2.3 小结.....	29

第 2 篇 XML 数据库设计技巧

第 3 章 XML 与 database.....	33
3.1 XML 与 Database 之间的关系.....	33
3.2 Data-Centric 和 Document-Centric Document.....	33
3.2.1 Data-Centric Document.....	34
3.2.2 Document-Centric Document.....	34
3.3 Storing and Retrieving Data from Data-Centric Document.....	36
3.3.1 Template-Driven Mappings.....	36
3.3.2 Model-Driven Mappings.....	37
3.3.3 选择 middleware.....	41
3.4 Storing and Retrieving Documents.....	43
3.5 Middleware 与 Native XML Database.....	43
3.5.1 Data-Transfer Middleware.....	43
3.5.2 Native XML Database.....	44
3.6 小结.....	46

第 3 篇 XML 应用技术深入解析

第 4 章 XML Schema.....	49
4.1 何谓 XML Schema.....	49
4.1.1 粗略检验与细致检验.....	50
4.1.2 数据与麻烦相随而生.....	50
4.1.3 XML Schema 与 DTD 的差异.....	51
4.2 使用 Schema.....	52
4.2.1 一个简单的 XML Schema.....	52
4.2.2 简化 XML Schema.....	56
4.2.3 定义型态.....	57
4.2.4 群组、组合和衍生.....	60
4.3 Schema 高级应用.....	64
4.3.1 内容型态.....	64
4.3.2 定义限制.....	66
4.3.3 创建有再利用性的 Schema 文件.....	68
4.3.4 命名域.....	70
4.3.5 实体文件.....	72
4.4 相关工具介绍.....	73
4.4.1 Xerces-J.....	73
4.4.2 Oracle Schema Processor.....	78
4.4.3 其他工具.....	81
4.5 小结.....	85
第 5 章 DOM 应用技术介绍.....	87
5.1 认识 DOM.....	87
5.1.1 以 DOM 表示文件.....	87
5.1.2 DOM 的 Level.....	88
5.2 DOM 的应用.....	89

5.2.1 JDOM	89
5.2.2 建立 DOM	90
5.2.3 存取 DOM	94
5.2.4 高级技巧	100
5.3 其他 DOM	103
5.4 小结	108
第 6 章 SAX	109
6.1 认识 SAX	109
6.1.1 SAX 的由来	109
6.1.2 什么是 Event-Based Interface	110
6.2 SAX 实际操作与应用	112
6.2.1 SAX 解析器的操作	113
6.2.2 SAX package 的应用	127
6.3 SAX 与 DOM	133
6.3.1 SAX 与 DOM 的真实关系	134
6.3.2 使用 SAX 转换成 DOM	135
6.4 小结	139
第 7 章 XSLT	141
7.1 XSLT 介绍	141
7.2 XSLT 文件的根元素及命名域	145
7.3 XSLT 文件的操作	145
7.3.1 内容元素	146
7.3.2 回路元素	146
7.3.3 条件元素	148
7.3.4 属性元素	151
7.4 XSLT 样版	151
7.4.1 样版套用	152
7.4.2 样版调用	155

7.5 小结	158
第 8 章 XML Query	161
8.1 什么是 XML Query	161
8.2 XML Query 语法	161
8.2.1 XML 样版	161
8.2.2 高级语法说明	174
8.3 相关工具介绍	178
8.3.1 GMD-IPSI XQL Engine	178
8.3.2 Xtract	185
8.4 小结	193
第 9 章 Xlink、Xpointer、XPath 及 XBase	195
9.1 XLink、XPointer、XPath 及 XBase	195
9.1.1 XLink(XML Link Language)	195
9.1.2 XBase	197
9.1.3 XPointer	198
9.1.4 XPath	198
9.2 了解 XLink	199
9.2.1 扩展链接(Extended Links)	199
9.2.2 简单链接(Simple Links)	209
9.3 了解 XBase	214
9.4 了解 XPath	215
9.4.1 XPath 中的轴	217
9.4.2 XPath 的节点测试	219
9.4.3 XPath 的预测(predicate)	219
9.5 了解 XPointer	228
9.5.1 XPointer 中的轴	228
9.5.2 XPointer 的节点测试	229
9.5.3 XPointer 的预测	230

9.5.4	XPointer 的位置组函数	230
9.5.5	使用 XPointer 的点功能	231
9.5.6	使用 XPointer 的范围功能	233
9.6	相关工具介绍	234
9.6.1	Fujitsu XLink Processor	235
9.6.2	XLink2Html	236
9.6.3	4suite	236
9.7	小结	237
第 10 章	其他 XML 规范介绍	239
10.1	XML Base	239
10.1.1	什么是 XML Base	239
10.1.2	语法	240
10.2	Xinclude	242
10.2.1	什么是 Xinclude	242
10.2.2	Xinclude 的语法	243
10.2.3	parse 属性	244
10.2.4	使用 XPointer	245
10.2.5	命名域(namespace)与基地址(base URL)	247
10.2.6	可能发生的错误	247
10.3	XForms	248
10.3.1	什么是 Xforms	248
10.3.2	数据模型(data model)与用户接口(user interface)	249
10.3.3	XForms 的结构	250
10.3.4	XForms Model	254
10.4	MathML	264
10.4.1	MathML 简介	264
10.4.2	MathML 语法范例	265
10.4.3	MathML 的应用	268
10.5	RDF	268
10.5.1	RDF 简介	268

10.5.2	RDF 范例	269
10.5.3	RDF 应用	270
10.6	RSS	270
10.6.1	RSS 简介	270
10.6.2	RSS 范例	271
10.6.3	RSS 应用	272
10.7	小结	273
第 11 章	Schema Repositories	275
11.1	Schema Repositories 介绍	275
11.2	XML.org 的 repository	275
11.3	Biztalk 的 Repository	277
11.4	小结	279

第 4 篇 数据库应用范例

第 12 章	Microsoft SQL Server 2000 与 XML 的支持	283
12.1	由 IIS(Internet Information Server)存取 Microsoft SQL Server 2000	283
12.1.1	设置 SQL Server 2000 的虚拟目录	284
12.1.2	通过 URLs 存取 SQL Server 2000	289
12.2	从 SQL Server 2000 撷取 XML 数据	294
12.2.1	FOR XML 子句	294
12.2.2	使用 FOR XML 子句时 QUERY 语法的限制	297
12.3	小结	297
第 13 章	Oracle 支持 XML 相关技术	299
13.1	前言	299
13.1.1	未来 Oracle 对于 XML 技术方面的期望	300
13.1.2	为何 Oracle 这么注重 XML 技术	300

13.1.3 简化 XML 应用程序的开发.....	301
13.2 Oracle XML Developer's Kit	302
13.2.1 Oracle XML Developer's Kit for Java	302
13.2.2 Oracle XML Developer's Kit for C/C++.....	306
13.2.3 Oracle Internet Developer Suite.....	308
13.3 小结	312

第 5 篇 常见的 XML 应用

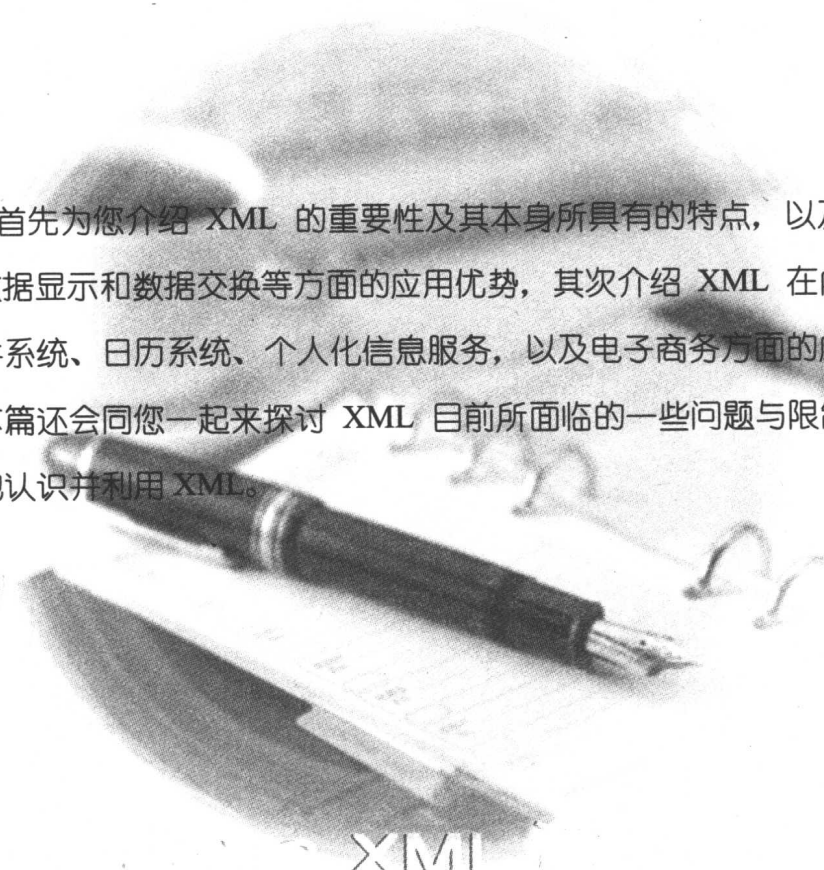
第 14 章 XML Messaging	315
14.1 XML Messaging 概要	315
14.2 SOAP 概要	316
14.3 SOAP 和 RPC.....	318
14.4 在 Tomcat 中设置 SOAP 的环境.....	319
14.5 编写一个简单的 SOAP 程序.....	320
14.5.1 简单的 SOAP 程序(server 端).....	320
14.5.2 简单的 SOAP 程序(client 端).....	320
14.5.3 设置 XML-SOAP Admin	323
14.5.4 执行结果	324
14.6 ebXML 概要.....	324
14.7 Sun 所推出的 JAXM.....	325
14.8 小结	327
第 15 章 XML Data Binding	329
15.1 何谓 XML Data Binding	329
15.2 marshal 和 unmarshal	330
15.3 XML Schema	331

目 录

15.4 Breeze XML Data Binding	335
15.5 小结	336
附录 A Element Syntax Summary	337
附录 B 目前各产业所有的 Schema.....	345

第 1 篇 XML 简介

本篇首先为您介绍 XML 的重要性及其本身所具有的特点，以及它在数据检索、数据显示和数据交换等方面的应用优势，其次介绍 XML 在内容管理、电子邮件系统、日历系统、个人化信息服务，以及电子商务方面的应用特性。最后，本篇还会同您一起来探讨 XML 目前所面临的一些问题与限制，以帮助您更好地认识并利用 XML。



XML

第 1 章 认识 XML

本章将简单说明 XML 的重要性、XML 的特点，以及这些特色应用到内容管理、信息检索、电子邮件系统、智能型日历系统以及电子商务上所具有的优势。最后同您一起来探讨 XML 目前应用上受到的一些限制。

1.1 XML 的重要性

XML 的全名为 Extensible Markup Language，中文译义为“可扩展性标记语言”，它是继 HTML 之后，另一种新兴的因特网标准。与 HTML 相似，它是一种显示数据的标记语言，它能使数据通过网络无障碍地进行传输，并显示在用户的浏览器上。然而，由于 XML 具备的某些优越性(下一节详述)，使得许多大型企业对 XML 的应用远景充满期待，并纷纷投入对 XML 的应用开发当中。比尔·盖兹就曾预言 XML 将成为因特网的主要语言，并宣称要带领微软全力发展 XML 的应用软件；英国《经济学家》杂志也曾大力推崇 XML，认为 XML 有潜力成为电子商务的主要媒介。

XML 的结构图标

XML 除了本身所具有的技术外，还拥有一系列辅助技术，例如 RDF、XLink、XPointer、DTD 及 XML Schemas 等，这些都可以使 XML 的功能更加完善，XML 的技术及功能结构图，如图 1.1 所示。

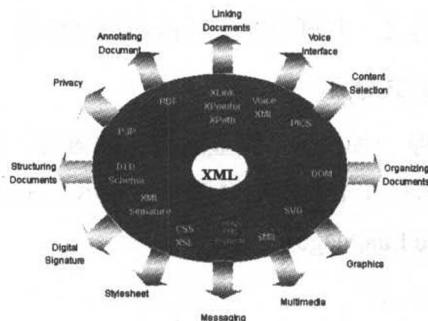


图 1.1 XML 的技术及功能结构图