

Instant XML/XSL/Java

实例解析 XML/XSL/Java 网络编程

41 个典型范例深入描述 XML/XSL/Java
网络编程的方法和技巧

2002
编程实例解析
应用丛书

4

北京希望电子出版社 总策划
李江 张威 编写



中航出版社



北京希望电子出版社

Instant XML/XSL/Java

实例解析 XML/XSL/Java 网络编程

41 个典型范例深入描述 XML/XSL/Java
网络编程的方法和技巧

2002
编程实例解析
应用丛书

北京希望电子出版社 总策划
李江 张威 编写



中航出版社



北京希望电子出版社

内 容 简 介

本书通过 41 个实例全面地介绍了 XML/XSL/Java 技术结合编程的方法和技巧。全书由七章组成, 主要内容包括: XML 简介、XML 与 CSS、DTD 与 SCHEMA、XML 与 XSL、客户端 XML 数据访问、JAVA 与 XML、XML 与 .net 的应用、基于 SOAP 的应用等。

本书组织结构合理, 从不同的应用角度分章组织; 每一章开始, 有本章相关知识的简单介绍, 提纲挈领, 然后是主要部分——应用实例, 其中的实例都是作者精心挑选的非常具有代表性的实例, 大量的实例将教会你如何将所学用于实践。

本书面向初、中级用户, 对高级用户也有参考价值, 同时也可作为社会 XML 培训教材。

本书实例程序源代码请从 www.bhp.com.cn 上下载, 或可另行购买配套光盘 (包含本套书的相关程序、素材), 定价 15 元 (含邮费), 需要购买者请直接与北京中关村大街 26 号 083 信箱 (邮编 100080) 常丽小姐联系。

需要本书的读者, 请与北京中关村 083 信箱北京希望电子出版社 (中关村大街 26 号, 邮编 100080) 联系。网址: www.bhp.com.cn, E-mail: xrl@hope.com.cn。电话: 010-62562329, 62541992, 62637101, 62633308, 62633309 (发行部); 010-62650876 (门市); 010-62629581 (编辑部)。传真: 010-62579874。

宇 航 出 版 社
北京希望电子出版社 出版发行

北京市和平里滨河路 1 号 (100013)
北京中关村大街 26 号 (100080)
发行地址: 北京阜成路 8 号 (100830)
北京中关村大街 26 号 (100080)
北京广益印刷有限公司
新华书店经销

本版本号: ISBN 7-80144-083-8

2002 年 1 月第 1 版 2002 年 1 月第 1 次印刷

开本: 787×1092 1/16 印张: 28.375 字数: 509 千字

印数: 0001-5000 册 全套定价: 320 元 (共八本) 本书定价: 40.00 元

《2002 编程实例解析丛书》序

——学习编程就是为了编程！

感谢你翻看我们这套丛书，在到收款台之前，请务必阅读如下内容，以进一步确定《2002 编程实例解析丛书》正是为你量身订做！

《2002 编程实例解析丛书》包括 8 本教程：

- 实例解析 Java 2 应用编程
- 实例解析 XML 网络编程
- 实例解析 HTML 4.0 网络编程
- 实例解析 Delphi 6.0 应用编程
- 实例解析 VC.NET 应用编程
- 实例解析 VB.NET 应用编程
- 实例解析 ASP.net 网络数据库编程
- 实例解析 AutoCAD 2002 建筑绘图

本丛书的首要目标是让读者在短期内具备现实的编程能力，同时形成自己的编程思路，而不仅仅是让读者熟悉一种或几种开发环境。所以本丛书的出发点就是实例——大量的、经典的、热门的、前卫的实例！

我们的本意是让读者在实际编程中体会一切——包括对开发界面及相关知识的掌握，但是在实际操作上，这并不容易，所以我们在丛书的合适位置添加了一些进入实战之前必须掌握的一些知识（包括界面、相关概念、相关技术等）——这是本丛书的第二个目标，“知其然，并知其所以然！”。

读者需求

开发人员从来没有像今天这样轻松过。每种技术都在变得越来越完美。几天前还需要经过复杂方式实现的功能，现在可能已经集成到某个软件中去，我们只需鼠标拖放等简单方式就可以实现！随着技术更新速度的加快，按部就班的学习方式是否也应该做出调整？

对于初、中级用户

传统的“学以致用”的学习方法似乎越来越跟不上技术发展的节奏了。初、中级用户总是有一种“完美主义”情结，在动手研究一种软件或一种技术之前将听说过的所有相关软件逐一攻克。有没有这种必要？没有！并且也不可能！原因之一：现在的软件的很多功能即使专业人士也很少使用，而且使用某种功能的时机、方式把握不好会弄巧成拙；原因之二：即使你弄懂了一个软件的所有功能，到使用时可能已经忘掉了；原因之三：软件的

MS07/03

更新速度比硬件还快，任何人都不可能跟上所有相关软件的更新速度。所以《2002 编程实例解析应用丛书》大力倡导了一种“用以致学”的全新思路！

初学者能否掌握一种技术的关键是快速入门，如果研究一种技术的时间太久可能会失去兴趣和信心（而且这期间总有些看起来更好的技术在扰乱你的注意力）。对于应用来说，其实最讲究现学现卖，没有必要将一种或几种软件的所有功能都了如指掌。其实只要到达了某一层次，学习技术将是非常容易的一件事情（很多技术都是相通的）。

对于中、高级用户

现实地说，国内程序员几乎个个都是国外程序员“须仰视才见”的“多面手”，他们的简历上往往写着精通数种程序语言。这是因为国内软件行业的现状要求程序员必须能快速掌握企业需要的技术。从这个意义上说，我们的中高级用户也在时刻充当着初学者，但是他们的要求往往只是几个典型的、实用的实例。

《2002 编程实例解析应用丛书》，也为这些用户提供了大量的这类经典实例。丛书的作者都是一线开发人员，所以实例的内容涵盖了几乎所有目前的热点内容。我们有理由相信《2002 编程实例解析应用丛书》一定有你需要的！

对于社会培训班学员、大中专院校师生

长此以来，我们的编程教材一直在坚持“理论高于一切”的思想，这保证了我们的学生具有最坚实的理论基础，但是往往在下手编写程序时才感觉捉襟见肘。这可以说是现有的绝大多数教材的弊端之一：理论性强，但可操作性差，实用性差。

《2002 编程实例解析应用丛书》完全可以作为编程类教材一个有效补充。帮助学生将所学理论融会贯通，达到中级开发人员的水平。

《2002 编程实例解析应用丛书》，共包括 8 种热点程序开发语言：Java 2, XML, HTML 4.0, Delphi 6.0, VC.net, VB.net, ASP.net, AutoCAD 2002。帮助用户彻底走出“学以致用”的误区，踏上“用以致学”的正轨。由浅入深的组织方式使你不断获得成就感；所讲述的实例都是最常用、最经典的用法，使用这些实例完全可以胜任绝大多数情况下的编程任务。所以说《2002 编程实例解析应用丛书》不仅是初学者成为高手的入门捷径，而且中高级用户也可以从中找到大量有价值的东西。

北京希望电子出版社

2002 年 1 月

前 言

欢迎来到令人惊奇和充满动感的 XML 世界！如果你从未接触过 XML，那么在这里可以享受难得的乐趣。XML 是一种非常丰富且简单易学的标记语言，XML 为程序员提供了前所未有的编程方法以解决人们看来十分复杂的任务。

什么是 XML？XML 代表 eXtensible Markup Language，意为可扩展标记语言。XML 是一套定义语义标记的规则，这些标记将文档分成许多部件并对这些部件加以标识。它也是元标记语言，即定义了用于其他与特定领域有关的、语义的、结构化的标记语言的句法语言。

本书的使用对象

如果读者对 XML 有一定了解，但是对 XML 的应用想得到提高，那么本书就是你的选择，它会告诉你 XML 应用的各个方面；如果你以前没有接触过 XML，本书可以作为你学习 XML 的向导，教你如何使用它。

本书介绍了与 XML 结合应用的各个方面，不管你是网页作者，还是程序员，都可以找到你需要的东西。

本书如何组织

本书由七章组成。在每一章的开始，都有本章内容概述，介绍本章涉及的主要内容；然后简单介绍本章中可能用到的相关知识，接下来给出了大量的有代表性的应用实例，这些实例按应用分类，结构非常清楚。

绪论部分 介绍了 XML 的诞生、它带来的好处、XML 工具及其实现，以及它的发展趋势等。

第一章 XML 与 CSS 介绍了什么是 XML、CSS。实例涉及如何用 XML 描述数据，如何编写自己的 CSS 样式单，如何将 CSS 用于 XML，如何将不同的 CSS 加于 XML 等等，是应用 XML 的基础。

第二章 XML 定义语言 介绍了两种 XML 定义语言：DTD 与 SCHEMA。XML 的使用要符合一定的语法规则，本章首先介绍 DTD 与 SCHEMA 的基本语法知识，然后给出实例教读者如何编写自己的 DTD 与 SCHEMA。

第三章 XML 与 XSL 介绍 XSL 的基本以及高级应用。XSL 是 XML 中的一个重要组成部分，与 XML 密不可分。CSS 是 XSL 的过渡手段，XSL 的应用非常广泛。本章开始介绍了 XSL 的基本知识，然后给出了大量的基础和高级应用实例，循序渐进的教会读者如何将 XSL 有效地用于自己的应用。

- 第四章 客户端 XML 数据访问** 介绍了客户端 XML 数据使用领域的知识。如何将服务器端减肥，将应用转移到客户端是目前研究的一个重要领域，特别是在分布式应用领域，客户端的 XML 数据应用都很有价值。本章首先介绍了数据岛（DSO）的概念，然后用实例向读者展示了在客户端访问 XML 数据的几种基本方法。
- 第五章 JAVA 与 XML** 介绍了 XML 与 JAVA 的结合应用领域。从一开始，XML 与 JAVA 的结合就被业界看好，称其为“完美的组合”。本章首先介绍了 JAVA 对 XML 应用提供的 API 支持，然后用综合实例介绍了 JAVA 与 XML 结合的分布式应用、建立消息系统的应用、平衡遗留系统中的应用等领域。
- 第六章 XML 与 .NET 的应用** 介绍了 XML 与 ASP 的结合应用。因为 ASP 的应用比较广泛，我们给出一些 ASP 与 XML 应用的实例。
- 第七章 基于 SOAP 的应用** 介绍了 SOAP 基本知识及其应用实例。SOAP（Simple Object Access Protocol）是简单对象访问协议的简称，这种技术有助于实现大量异构程序和平台之间的互操作性，从而使存在的应用能够被广泛的用户所访问。SOAP 把基于 HTTP 的 WEB 技术与 XML 的灵活性和可扩展性组合在一起，具有广阔的发展空间。本章的综合实例向读者介绍了如何开发 SOAP 应用，通过实例，读者能够掌握 SOAP 的基本开发技巧。
- 附录一 XML 中使用的实体** 总结了在 XML 中可能用到的实体、实体表示、实体数字表示及其简单描述。
- 附录二 XML 专业技术词汇表** 总结了 XML 中经常使用的专业词汇，供读者查阅。
- 附录三 XML 资源** 列出了读者可以参考的许多有价值的资源。

本书选配光盘

本书中所有源代码都放在选配光盘中，我们将源程序按章按实例建立目录存放代码，例如，第三章实例 16 的源代码放在目录“\清单\第三章 XML 与 XSL\实例 16”目录下。

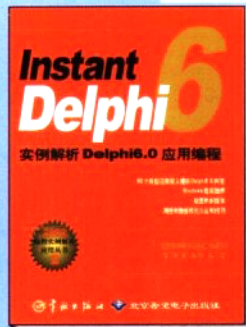
本书主要作者：李江、张威。参加本书编写工作的有王壮、李义、沈鹏、张侃、刘路、易非、何进、王朗、黄剑波、龚雪、骆玲、于鑫、蔡凡、朱海东、蒋珊、肖刚、张洁、赵辉、蔡杰、李林丽、吴敏、刘匡、李静、周元、徐浩、冯军、李楠、许雯、李明、吕芳、孙梅、徐振英、曹丽君、钟惠、张燕等。由于作者水平有限，时间仓促，错误和疏漏之处在所难免，恳请广大读者批评指正。

作者

掌握编程方法与技巧快捷之路



CX-3567
定价: 40.00 元



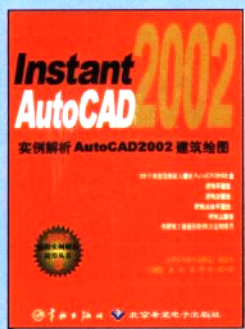
CX-3568
定价: 40.00 元



CX-3571
定价: 40.00 元



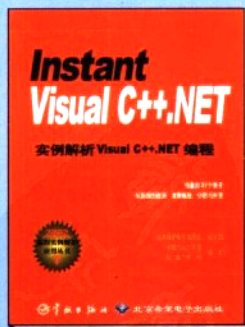
CX-3572
定价: 40.00 元



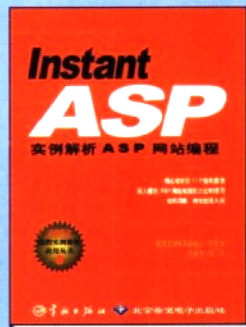
CX-3569
定价: 40.00 元



CX-3581
定价: 40.00 元



CX-3583
定价: 40.00 元



CX-3591
定价: 40.00 元

北京希望电子出版社
Beijing Hope Electronic Press
www.bhpe.com.cn

社址: 北京中关村大街 26 号 (黄庄路口东)
电话: (010) 62562329 62541992
通讯: 北京中关村 083 信箱 (100080)
传真: (010) 62579874 62633308



高级程序员成长之路



CX-83391
定价: 55.00 元



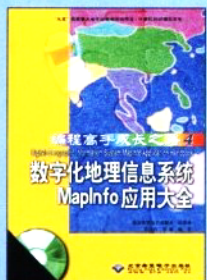
CX-83407
定价: 46.00 元



CX-83442
定价: 50.00 元



CX-83549
定价: 35.00 元



CX-83480
定价: 66.00 元



CX-83533
定价: 46.00 元



CX-83512
定价: 42.00 元



CX-83513
定价: 55.00 元



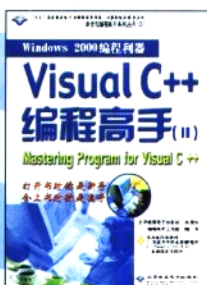
CX-83558
定价: 58.00 元



CX-83275
定价: 48.00 元



CX-83307
定价: 35.00 元



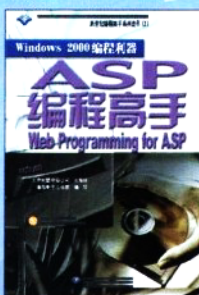
CX-83308
定价: 39.00 元



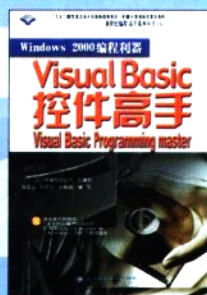
CX-83214
定价: 39.00 元



CX-83220
定价: 42.00 元



CX-83229
定价: 39.00 元



CX-83253
定价: 50.00 元

北京希望电子出版社
Beijing Hope Electronic Press
www.bhpe.com.cn

社址: 北京中关村大街 26 号 (黄庄路口东)
电话: (010) 62562329 62541992
通讯: 北京中关村 083 信箱 (100080)
传真: (010) 62579874 62633308



目 录

绪论 XML 简介	1	schema 文件	90
XML 的诞生	1	实例 12 网上书店图书管理	94
XML 入门篇, 具体实现 XML 的		2.4 本章小结	99
三种方式	10	第三章 XML 与 XSL	100
第一章 XML 与 CSS	11	3.1 XSL 的两种应用模式	101
1.1 XML 语法	11	3.2 XSL 的两种构成方式	101
1.1.1 逻辑结构	11	3.3 XSL 的基本元素	101
1.1.2 物理结构	13	3.4 XSL 的应用实例	104
1.2 CSS 简述	14	实例 13 自己动手做	
1.2.1 选择符	14	菜单 (menu)	104
1.2.2 声明	15	实例 14 咖啡屋早餐菜单	110
1.3 应用实例	16	实例 15 图书列表	114
实例 1 HELLO XML (I)	16	实例 16 排序功能	118
实例 2 HELLO XML (II)	19	实例 17 股票分类排序机	121
实例 3 客户关系管理	21	实例 18 综合转换实例: XML/XSL	
实例 4 用 XML 描述公司		阅读器演示	133
部门组成	26	3.5 本章小结	152
实例 5 植被管理	30	第四章 客户端 XML 数据访问	154
实例 6 开心 CD 目录	38	4.1 DSO 简介	154
实例 7 用 XML 描述		4.1.1 XML 的数据绑定	155
莎士比亚剧本	45	4.1.2 DSO 技术的应用	157
1.4 本章小结	62	4.2 DSO 应用实例	158
第二章 DTD 与 SCHEMA	63	实例 19 冲浪雇员列表	158
2.1 XML 定义语言: DTD, XDR		实例 20 电话号码查询	166
和 XSD	63	实例 21 在 VBScript 中	
2.2 DTD 语法简介	65	使用 DOM	180
实例 8 DTD 基础实例	67	实例 22 客户端生成并使用 XML ..	185
实例 9 帐簿管理	71	实例 23 XML 文档树查看器	191
实例 10 家谱树	76	实例 24 综合实例: 从 XML	
2.3 SCHEMA 语法简介	83	到 DHTML	205
实例 11 订货单格式的		4.3 本章小结	211

第五章 JAVA 与 XML	213	和 DOM 树	330
5.1 JAVA 对 XML 的技术扩展	213	5.5 本章小结	347
5.1.1 针对 XML 的 Java 技术标准扩展	213	第六章 XML 与 .NET 的应用	348
5.1.2 企业平台支持	214	6.1 综合应用实例	348
5.1.3 XML 与 JAVA: 完美组合	214	实例 33 用 ASP 和 XML 创建 音乐站点	348
5.2 JAVA 对 XML 的两种编程接口	215	实例 34 ASP+XML 实现留言板	357
5.2.1 SAX 接口简介	215	实例 35 ASP+XML 实现新闻 更新系统	365
5.2.2 DOM 接口简介	219	实例 36 用 XML 构建 ASP+config.web 配置	370
5.3 JAVA 与 XML 综合应用	220	6.2 本章小结	393
实例 25 在 JAVA 中利用 DOM 遍历树	220	第七章 基于 SOAP 的应用	394
实例 26 处理 XML 定单 (order)	229	7.1 SOAP 简介	394
实例 27 XML 信息交换系统	238	7.2 SOAP 核心	396
实例 28 用 XML、XSL、JAVA 技术平衡遗留系统	246	7.3 SOAP 数据类型	398
实例 29 用 XML 和 JAVA 建立 分布式应用	259	7.4 SOAP 综合应用	402
实例 30 XML/DATABASE 与 JAVA	290	实例 37 一个简单 SOAP 应用	402
实例 31 用 XML 简化 AWT 中 菜单制作	300	实例 38 JAVA 实现 SOAP Client	407
5.4 Servlet 与 XML 结合使用	328	实例 39 远程图书索引	411
5.4.1 Servlet 简介	328	实例 40 MS SOAP Toolkit 1.0 的 C++SOAP 客户端	420
实例 32 用 Servlet 生成 XML		实例 41 SOAP 实现 XML 消息机制	428
		7.5 本章小结	439

绪论 XML 简介

XML 的诞生

XML (eXtensible Markup Language, 可扩展标记语言), 被称为“第二代 Web 语言”、“下一代网络应用的基石”。自它被提出以来, 几乎得到了业界大公司的支持, 丝毫不逊于当年 HTML 被提出时的热度。

■ HTML 的缺点

谈 XML, 不得不先说说“第一代的 Web 语言”——HTML。HTML 老矣, 已经快被淘汰了(夸张了一点, 不过 HTML 4.0 后就没有了, 取而代之的是依据 XML 定义的 XHTML)。为什么呢? HTML 对带动 WWW 的蓬勃发展, 可谓功不可没。想要在 Internet 上作点事情, 几乎无 HTML 不行。但是 HTML 有一个致命的缺点, 就是: 只适合于人与计算机的交流, 不适合计算机与计算机的交流。

大家都知道, HTML 是通过一大堆的标记来定义文档内容以什么样的形式显现, 也即, HTML 是一种显示描述语言, 它仅仅描述了 Web 浏览器应该如何在页面上布置文字、图形等, 并没有对 Internet 上最重要的东西——信息的本身含义进行描述。这些通过 HTML 表现出来的文字、图形内容很容易被人理解, 而要计算机去理解这些标记内的文字的含义, 就很困难了。

举个例子来说, 我们设计一个程序, 它可以自动地到各大网上商场去将最新的价目抓回来。然而问题是, 每个网上商场可能在网站中写商品名称和价格时, 都有它们自己的一套写法, 如: 甲用价格, 而乙用<H>价格</H>, 还有更为复杂的表格。那么我们的程序怎样才能知道, 哪种标记里面的东西才是要抓的价格信息呢? 再比如: 在 HTML 里, Apple只代表了 Apple 这个单词在 Web 浏览器里用粗体来表现, 并没有指出 Apple 代表什么, 是苹果? 是苹果计算机公司? 还是其它什么? 这造成了 HTML 不能揭示文件中信息的含义。

HTML 的另一个问题就是它的标记的集合是固定的, 用户不能增加自己的有意义的标记。而各大浏览器的规格不尽相同, 要使我们用 HTML 做的网页能够被所有浏览器正常显示, 我们只能使用 W3C 给我们定义好了的标记来创建网页。

在当今的网络世界里, 随着电子商务的蓬勃发展和基于 Web 的应用日益广泛, 大量的信息需要被快速地处理。实际上, 在 Internet 上的大部分信息, 在最初都是被存放在结构良

好的数据库里面，信息按照它的意义被存放在相应的字段里，比如：“员工档案”数据库，它至少包括名称、性别、部门等字段。对于“张三”这个数据，计算机能够根据它所存放的位置，知道它代表一个员工的姓名。但是，一旦这些数据被调出来，经过 CGI、ASP、JSP、PHP 等转换成 HTML 后，原本有意义的数据就变成了些无特定含义的 HTML 标记的组合。用户必须通过自己的“头脑”才能解析这些数据，进而“手动”将它们进行记录、处理，显然处理信息的速度会很慢。如果我们能够将最初保存在数据库中的原始结构的数据在计算机之间传递，那么肯定会加快信息处理的速度。显然利用 HTML 办不到，并且，由于计算机体系、操作系统以及所使用的数据库不同，不同的计算机之间要想互相理解对方的数据库格式是相当困难且非常麻烦的，为了使各种不同的计算机之间能够互相交换信息，似乎 HTML 又必不可少。怎样来解决这个难题呢？

迎接 XML

使用 XML 可以解决上述的难题。W3C 对 XML 作了如下描述：“XML 描述了一类被称为 XML 文档的数据对象，并部分描述了处理它们的计算机程序的行为。XML 是 SGML 的一个应用实例或一种受限形式。从结构上说，XML 文档遵从 SGML 文档标准。”同 HTML 一样，XML 也是一种基于文本的标记语言，都是从 SGML(Standard Generalize Markup Language, 标准通用标记语言，一种老的标记语言，最初用于出版行业，非常复杂，依据 SGML 开发的应用非常昂贵，只在少数大公司和政府部门有应用)发展而来的，XML 保留了 SGML 80% 的功能，降低了 20% 的复杂程度，这样使得开发有关 XML 的应用变得很便宜，使 XML 能够进入“寻常百姓家”。

XML 与 HTML 的不同在于：XML 可以根据我们要表现的文档，自由地定义标记来表现具有实际意义的文档内容，比如：我们可以定义<文档名称></文档名称>这样具有实际意义的标记（可以用中文）。在 XML 中，我们只需要注意文档的内容，而文档的表现形式则交给 CSS（层叠样式表）和 XSL（可扩展样式语言）完成。如果 XML 文件只用于计算机与计算机之间交流信息，仅仅需要一个 XML 文件即可；如果要将 XML 文件中的信息以某种形式显示出来，如通过浏览器显示，则可引用一个样式表文件来定义浏览器怎样来显示 XML 文件中信息。而且 XML 不像 HTML 那样具有固定的标记集合，它实际上是一种定义标记的语言，也就是说使用 XML 的用户可以定义无穷的标记来描述文档中的任何数据元素，将文档的内容组织成丰富复杂的完整的信息体系。XML 主要有三个要素：Schema（模式）、XSL(eXtensible Stylesheet Language 可扩展样式语言)和 XLL(eXtensible Link Language 可扩展链接语言)。Schema 规定了 XML 文档的逻辑结构，定义了 XML 文档中的元素、元素的属性，以及元素和元素的属性之间的关系，它能够帮助 XML 的解析器校验 XML 文档标记是否合法；XSL 是用来规定 XML 文档表现形式的语言，同 CSS 类似；XLL 则进一步地扩展了当前 Web 上已有的简单链接。

XML 是一种定义标记的语言，现在已经有几个以 XML 规范为主所创建的标记语言，如：Chemical Markup Language (CML：定义怎样描述化学分子式的结构，将它显示在网页上)，Mathematical Markup Language (MathML：将复杂的数学公式以网页的形式显示在浏览器中)、Synchronized Multimedia Integration Language (SMIL：如何将多媒体信息展现在 WWW 中)。

XML 文件的解析器（一种检查 XML 文件是否有结构上的错误，将 XML 文件中的标记剥离，读出正确信息的工具）大多数是使用 Java 语言写成的，这样，只要计算机支持 Java 虚拟机，都可以支持 XML（几乎所有的计算机都支持 Java 虚拟机）。所以，即使是异构系统，也不用担心读不懂对方的资料，大家都用 XML 文件作为传送资料的介质即可，因为只要对方计算机上有合适的 XML 解析器，就可以正确地读取信息。现在，几大数据库厂商的数据库产品，如：Oracle 8i, Informix, IBM DB2 等都开始支持 XML，输入数据库的数据可以轻易地转化为 XML，甚至可以直接以 XML 的形式输入数据。有预言称：将来的电子文档必将是 XML 的天下。

学习 XML 并不困难，因为 XML 的规范很简洁，整个标准打印出来也只有几十页，而且 XML 的写法和 HTML 类似，都是把标记用<>符号括起来。更加方便的是，我们能够使用中文创建 XML 标记，比如，创建<价格>...</价格>这样的标记，在此标记内的内容，它的含义就是某件东西的价格。想象一下，如果各大网上商场都用这样的 XML 标记来说明他们网页中文字的含义，那么我们就可以用一个自动化的软件去抓取我们感兴趣的东西，比如：我们想了解一下有关 XML 书籍的信息，这个软件就自动去抓取各个网页中<XML 书籍>...</XML 书籍>标记内的字段，这该是多么方便。

XML 具有卓越的性能，它具有四大特点：优良的数据存储格式、可扩展性、高度结构化以及方便的网络传输。因为 XML 能针对特定用户的应用定义自己的标记，这就使 XML 能够在多种行业的信息交换中一显身手，根据不同行业来提供具有各自特色的解决方案。

■ XML 带来的好处

XML 给基于 Web 的应用软件赋予了强大的功能和灵活性，因此它给开发者和用户带来了许多好处。

更有意义的搜索

数据可被 XML 唯一的标识。没有 XML，搜索软件必须了解每个数据库是如何构建的。这实际上是不可能的，因为每个数据库描述数据都是不同的。有了 XML，书就可以很容易以标准的方式按照作者、标题、ISBN 序号或其他的标准分类。搜索书就变得十分方便。

开发灵活的 Web 应用软件

数据一旦建立，XML 能被发送到其他应用软件、对象或者中间层服务器做进一步地处理。

理。或者它可以发送到桌面用浏览器浏览。XML 和 HTML、脚本、公共对象模式一起为灵活的三层 Web 应用软件的开发提供了所需的技术。

不同来源数据的集成

现在搜索多样的不兼容的数据库实际上是不可能的。XML 能够使不同来源的结构化的数据很容易的结合在一起。软件代理商可以在中间层的服务器上对从后端数据库和其他应用处来的数据进行集成。然后,数据就能被发送到客户或其他服务器做进一步的集合、处理和分发。

多种应用得到的数据

XML 的扩展性和灵活性允许它描述不同种类应用软件中的数据,从描述搜集的 Web 页到数据记录。同时,由于基于 XML 的数据是自我描述的,数据不需要有内部描述就能被交换和处理。

本地计算和处理

XML 格式的数据发送给客户后,客户可以用应用软件解析数据并对数据进行编辑和处理。使用者可以用不同的方法处理数据,而不仅仅是显示它。XML 文档对象模式(DOM)允许用脚本或其他编程语言处理数据。数据计算不需要回到服务器就能进行。分离使用者观看数据的界面,使用简单灵活开放的格式,可以给 Web 创建功能强大的应用软件,这些软件原来只能建立在高端数据库上。

数据的多样显示

数据发到桌面后,能够用多种方式显示。通过以简单开放扩展的方式描述结果化的数据,XML 补充了 HTML,被广泛地用来描述使用者界面。HTML 描述数据的外观,而 XML 描述数据本身。由于数据显示与内容分开,XML 定义的数据允许指定不同的显示方式,使数据更合理地表现出来。本地的数据能够以客户配置、使用者选择或其他标准决定的方式动态地表现出来。CSS 和 XSL 为数据的显示提供了公布的机制。

粒状的更新

通过 XML,数据可以粒状的更新。每当一部分数据变化后,不需要重发整个结构化的数据。变化的元素必须从服务器发送给客户,变化的数据不需要刷新整个使用者的界面就能够显示出来。目前,只要一条数据变化了,一整页都必须重建。这严重限制了服务器的升级性能。XML 也允许加进其他数据,比如预测的温度。加入的信息能够输出到 XML 页面,不需要再向浏览器发一个新的页面。

在 Web 上发布数据

由于 XML 是一个开放的基于文本的格式,它可以和 HTML 一样使用 HTTP 进行传送,

不需要对现存的网络进行变化。

升级性

由于 XML 彻底把标识的概念同显示分开, 处理者能够在结构化的数据中嵌套程序化的描述以表明如何显示数据。这是令人难以相信的强大的机制, 使得客户计算机同使用者间的交互作用尽可能的减少了, 同时减少了服务器的数据交换量和浏览器的响应时间。另外, XML 使个人的数据只能通过更新的布告发生变化, 减少了服务器的工作量, 大大增强了服务器的升级性能。

压缩性

XML 压缩性能很好, 因为用于描述数据结构的标签可以重复使用。XML 数据是否要压缩要根据应用来定, 还取决于服务器与客户间数据的传递量。XML 能够使用 HTTP1.1 中的压缩标准。

开放的标准

XML 基于的标准是为 Web 进行过优化的。微软和其他一些公司以及 W3C 中的工作组正致力于确保 XML 的互用性, 以及为开发人员、处理人员和不同系统和浏览器的使用者提供支持, 并进一步发展 XML 的标准。

XML 包括一套相关的标准:

可扩展标识语言(XML)标准, 这是 W3C 正式批准的。这意味着这个标准是稳定的, 完全可用于 Web 和工具的开发。

XML 名域标准, 这用来描述名域的句法, 支持能识别名域的 XML 解析器。

文档对象模式(DOM)标准, 这为给结构化的数据编写脚本提供了标准, 这样开发人员就能够同计算机在基于 XML 的数据上进行交互作用。

可扩展类型语言(XSL)标准, 这是一个工作草案。XSL 有两个模块: XSL 转换语言和 XSL 格式化对象。转换语言可用来转换 XML 以满足显示要求。由于 XSL 的两部分是模块, 转换语言能够独立地用来进行多用途的转换, 包括把 XML 转换成结构完整的 HTML。CSS 可应用于结构简单的 XML 数据, 但不能根据不同的信息以不同的传递方式来显示信息。

可扩展链接语言(XLL)标准和 XML 指针语言(Xpointer)标准是当前的工作草案。XLL 提供类似与 HTML 的链接, 但功能更强大。例如, 链接可以是多方向的, 可以存在于对象上而不仅仅是页面上。但 IE5 不支持 XLL。

微软产品的支持

微软 IE4 支持通用的 XML 解析器、XML 对象模式和 XML 数据源对象。IE5 和 Office 2000 支持许多添加的 XML 特性和技术。

新的机会

作为表示结构化数据的一个工业标准,XML 为组织、软件开发者、Web 站点和终端使用者提供了许多有利条件。更多的纵向市场数据格式建立起来,被应用于关键市场,诸如高级的数据库搜索、网上银行、医疗、法律事务、电子商务和其他领域,这使得发展机会更进一步地扩大。当站点更多地进行分发数据,而不仅仅是提供数据浏览时,特别的机会就产生了。

顾客服务正从电话和地理位置转移到 Web 站点上来,而且将会由于 XML 的强大功能受益更多。并且,由于大多数商业应用软件包括数据的处理和转移,如购买单、发货单、顾客信息、合同、图纸等等,XML 将会改革终端用户在 Internet 上的行为,许多商业应用将能实现。另外,使用基于 XML 的面向企业内部互连网的词汇库,Web 站点上的信息,无论是储存在文档中还是数据库中,可以被标识。这些词汇也能够对那些需要在顾客和供应商之间交换信息的中小型企业提供帮助。

一个重要的未开发的市场是开发使终端用户很容易建立自己的 Web 站点的工具,包括用来从数据库信息和存在的使用者界面中产生 XML 数据的工具。另外,标准模式可以开发用来描述数据,可以使用规划、图表、Excel 或其他电子数据表的功能。开发公布的用来描述从数据库中产生的 XML 的可视化工具是个很好的机会。观看 XML 数据的工具可以用 Visual Basic, Java 和 C++ 编写。

XML 需要强大的新工具用来在文档中显示丰富和复杂的 XML 数据,可以在分层的动态变化的数据上映射用户友好的显示层来实现这一目的。XML 数据的布局图包括数据透视表等。

Web 站点可以提供股票报价、新文章或实时的交易数据。通过制定信息老化的规则,信息超载可以避免。开发用户用来制定规则和服务器、客户软件用来实现规则的基于 XML 的工具是个巨大的机会。可以用脚本编写一个标准对象模式用来过滤进来的信息,检查储存的信息,创建输出的信息,进入数据库等等。

■ 实用工具概览

伴随着 XML 技术本身的成熟和应用领域的不断扩大,相关的工具也如雨后春笋不断地被开发出来。下面将介绍现有的、较有代表性的 XML 的实用工具。

XML 浏览工具

(1) Microsoft Internet Explorer 是 Microsoft 公司开发的 Web 浏览器,是当今两大主流浏览器之一,最先开创了 XML+CSS、XML+XSL 的 Web 浏览方式。但 IE5 对 CSS 的支持并不完全,尚不完全支持 CSS1,对 CSS2 只提供部分支持。IE5 对 XSLT 有所支持,能实时地将一个 XML 文档根据 XSL 样式单转换为 HTML 文档来显示,但支持的是早期草案。