

网页设计尖峰之旅丛书

XML

完全实例教程

卢啸龙 编著



包括书中例程、效果演示图以及与 XML 有关的共享软件的网址。



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

URL: [http:// www.phei.com.cn](http://www.phei.com.cn)

网页设计尖峰之旅丛书

XML 完全实例教程

卢啸龙 编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书根据 XML 最新标准,结合大量实例,深入浅出地介绍了 XML 及其相关知识。全书共分 14 章,包括:初识 XML、使用 XML、初识 XML 文档、XML 文档的构成、创建 XML 文档、文件类型定义、实体的定义和使用、DTD 中的属性设定、在 XML 中使用多种文字、CSS 样式单、XSL 样式单、XLink、XPointer 以及命名域。书中内容详实、全面,结构清晰,语言流畅易懂,适用于 XML 的初、中级用户。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,翻版必究。

图书在版编目(CIP)数据

XML 完全实例教程/卢啸龙编著. - 北京:电子工业出版社,2001,4
(网页设计尖峰之旅丛书)

ISBN 7-5053-6507-X

I . X… II . 卢… III . 可扩充语言,XML-程序设计-教材 IV . TP312

中国版本图书馆 CIP 数据核字(2001)第 05597 号

丛 书 名:网页设计尖峰之旅丛书

书 名:XML 完全实例教程

编 著 者:卢啸龙

策 划:王 颖 吴剑锋

责任编辑:王 颖

特约编辑:徐跃进

排版制作:电子工业出版社计算机排版室

印 刷 者:北京兴华印刷厂

出版发行:电子工业出版社 URL:<http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销:各地新华书店

开 本:787×1092 1/16 印张:18 字数:345.6 千字 附光盘:1 张

版 次:2001 年 4 月第 1 版 2001 年 4 月第 1 次印刷

书 号:ISBN 7-5053-6507-X
TP·3576

印 数:8 000 册 定价:29.00 元(含光盘)

凡购买电子工业出版社的图书,如有缺页、倒页、脱页、所附磁盘或光盘有问题者,请向购买书店调换;
若书店售缺,请与本社发行部联系调换。电话 68279077

技术先锋 尖峰之旅

——《网页设计尖峰之旅丛书》与《网站建设尖峰之旅丛书》序

我们有幸生逢技术变革不断发生的伟大时代。几年来,信息领域里最为振奋人心的事情,便是 Internet 的迅速拓展和普及应用。Internet 作为一个蕴含巨大信息资源和人类智慧的网络空间吸引了越来越多的人。今天,Internet 的应用已经与人类生活密切相关,尤其是由信息技术推动和基于 Internet 而兴起的电子商务,已经成为人们普遍接受的一种崭新的商务方式。电子商务能够提供准确、快速、高效的商务环境,能使人们更高效、更省力、更省钱地从事社会和生产活动,代表了当今世界商务模式发展的主流方向。目前,全世界很多国家和地区都在大规模地发展电子商务,以求能为传统商务活动开创新的发展机遇,并为人们提供反应迅速、成本低廉的交易模式。

在电子商务建设中,最关键、最核心的任务是技术平台的建设,其中涉及的技术主要包括三个方面,即面向客户端的网页制作技术,面向服务器端的网站建设和 Web 数据库技术,以及实现客户端与服务器端相互关系的连接和集成技术等。为了帮助广大网页设计和网站建设的技术人员快速而深入地掌握这三方面的技术知识,我们组织有关大学教师和资深技术专家编写了《网页设计尖峰之旅》与《网站建设尖峰之旅》两套丛书。

这两套丛书基本上涵盖了当前电子商务建设中的主流技术与软件。其中,《网页设计尖峰之旅丛书》主要侧重于网页设计领域的流行软件和主流编程技术,如《Dreamweaver 3.0 实例教程》、《Fireworks 3.0 实例教程》、《Flash 4.0 实例教程》、《FrontPage 2000 中文版实例教程》、《CSS 完全实例教程》、《VBScript 与 JScript 实例教程》、《WAP 编程与开发实例教程》、《PhotoImpact 6.0 实例教程》、《HTML、JavaScript 与 Java 完全实战演练》、《Photoshop 6.0 实例教程》、《Dreamweaver 3、Fireworks 3 与 Flash 5 完全实战演练》、《网页制作实用素材集 I/II》、《Dreamweaver 2.0 实例教程》、《Fireworks 2.0 实例教程》等;《网站建设尖峰之旅丛书》主要侧重于网站建设、Web 数据库系统及网络系统连接与集成领域的流行软件和主流编程语言、开发技术,如《Web 数据库开发技术集成实战演练》、《ASP 实例教程》、《PHP 4 & MySQL 完全实例教程》、《Perl 5、PHP 4 与 CGI 实例教程》、《Dreamweaver 与 ASP 实战演练》、《FrontPage 2000 与 ASP 编程实战演练》、《JSP 完全实例教程》等。

两套丛书紧紧围绕“深入、实用、精炼”的创作主题,以轻松、简练、明快的行文笔调,深入浅出地讲解各书相应的内容,语言流畅、活泼、通俗易懂,举例新颖、实用并易于操作,能使学习者在比较短的时间里快速学习到丰富的软件使用方法和开发技术,收到卓有成效的学习效果,并迅速提高自己的技术开发水平。

两套丛书的特色和风格是一致的,总体上来讲,主要包括以下 5 个方面:

其一,立足技术前沿。两套丛书所选软件和编程语言、数据库系统等都是国内外著名软件公司的知名产品,也是当前国内应用最新、最广泛的产品。

其二,读者定位明确。丛书中“实例教程”类主要针对电子商务的初、中级专业技术人才,“实战演练”类主要针对电子商务的中、高级专业技术人才。

其三,实例讲解内容。两套丛书均注重通过大量的实例,以实际的技术问题和开发环境,

讲解开发案例、编程方法、软件使用技术与操作技巧等。

其四,版式灵活醒目。两套丛书在力求文字精练、脉络清晰的同时,注意通过大量的图表直观地说明问题,并在正文之外设计了一些特殊的段落,来讲解具有技巧性、提示性或重要性的内容。我们希望通过这种灵活醒目的版式,减轻读者的阅读难度,增加学习兴致,真正让学习变成一种乐趣。

其五,随书附带光盘。为了节省读者的学习和练习时间,两套丛书的每一部书都配有光盘。光盘包括3部分内容:多媒体操作指导教程、书中所有举例及例程源代码、相应的系统软件(由出品公司授权)及常用工具软件(共享或免费)。

由于技术在不断地发展,所以我们这两套丛书也采取了不断进取,不断推陈出新的创作与出版思路。也就是说,在已推出各书的基础上,一方面我们要继续创作和出版电子商务领域中其他流行软件和主流编程语言、数据库开发技术的“实例教程”或“实战演练”;另一方面,当软件或语言版本升级时,我们将及时推出更新版本的相应书籍,以保证我们丛书“立足技术前沿”的鼎力追求!

“严谨、实用、高质量、新技术”是我们两套丛书出版中孜孜以求的目标,尽管我们精心而为,刻意而志,但书中错误和不足之处难免,所以恳请广大读者不吝批评指正,多提宝贵意见。我们一定会认真听取读者心声、意见和建议,把后续工作做好,把两套丛书出好!

巅峰风景无限好。紧跟时代潮流,勇攀技术高峰,打造技术先锋,请与我们共赴尖峰之旅!

电子工业出版社

前 言

XML 是 Web 的未来。

随着 Internet 应用的日趋复杂和多样,原来构筑 Web 的基础——HTML 也日益显得力不从心。在一些发展非常迅猛的领域,如电子商务、远程教育、信息交换等领域,HTML 甚至成为这些应用发展的桎梏。尽管人们付出了很大的努力来改进 HTML,如增加表格、框架、脚本语言等等,但均未能从根本上改变其局限性。于是人们开始全力发展新的技术,但大部分新的技术都由于种种原因没有得到业界的首肯而推广、普及开来。惟有一项技术,不仅一出台便得到各方肯定,且在短短几年内迅速崛起,并得到 Microsoft、IBM 等各大公司的鼎力支持。各大公司竞相开发相关应用,成为继 Java 之后,Internet 上最引人注目、众望所归的技术明星。各种迹象表明,它必将成为未来网络、甚至未来计算机世界中信息类型的主流。这便是本书所要介绍的 XML——可扩展标记语言。

XML 是万维网联盟(W3C)创建的一组规范,用于在 Web 上组织、发布各种信息,它不仅可以满足迅速增长的网络应用的需求,还能够确保在通过网络进行交互操作时,具有良好的可靠性与互操作性。

在国外,XML 技术已经进入快速发展的时期。各种介绍、讨论 XML 技术的网站纷纷出现,与 XML 相关的技术书籍也大量涌现。但国内对 XML 的介绍和推广工作却刚刚起步,所出版的大部分介绍 XML 的书籍都是译著,我们自己编写的、侧重于中文应用的书籍还很少。为了给广大读者提供一种内容较为全面、详实,结构由浅入深,语言流畅易懂的介绍 XML 的入门级图书,电子工业出版社委托笔者编写了此书。

本书的内容分为四大部分。

第一部分:XML 简介,由第 1 章、第 2 章构成。该部分介绍了 XML 的基础知识和其他相关的内容。

第二部分:XML 语法,包括从第 3 章到第 9 章共 7 章的内容。该部分全面介绍了编写 XML 文档所需要的语法和结构方面的知识。掌握了本部分内容,就可以创建出合法、有效的 XML 文档了。这是本书最重要的部分。

第三部分:XML 文档的显示,包括第 10 章、第 11 章共 2 章的内容。该部分介绍如何结合 CSS 和 XSL 等样式表语言,将 XML 文档在浏览器中进行显示。由于当前情况下在浏览器中进行显示几乎是 XML 最重要的应用,所以本部分内容也是本书的重点内容之一。

第四部分:XML 高级技术,包括第 12 章、第 13 章、第 14 章共 3 章的内容。该部分介绍了与 XML 相关的标准和技术,包括 XLink、XPointer 和命名域等内容。

由于本人对 XML 技术也在不断加强学习,加之本书的编写时间又较为紧张,书中疏漏之处在所难免,恳请各位读者批评指正。本人的 E-mail 地址:Lxlong72@sina.com。

在本书的编写过程中,得到了于韶平女士的大力支持。在此向她表示衷心的感谢。

卢啸龙

2001 年 1 月

· V ·

目 录

第 1 章 初识 XML	(1)
1.1 直接进入 XML	(1)
1.1.1 理解标记语言	(1)
1.1.2 理解 XML	(3)
1.2 从 HTML 到 XML	(5)
1.3 内容和显示的分离	(13)
1.4 为什么要用 XML	(15)
1.4.1 HTML 有缺陷	(15)
1.4.2 SGML 如何	(17)
1.4.3 应运而生的 XML	(19)
第 2 章 使用 XML	(21)
2.1 编辑 XML 文档	(21)
2.1.1 使用记事本编辑 XML 文档	(22)
2.1.2 使用 XML Notepad 编辑 XML 文档	(22)
2.2 检测 XML 文档	(26)
2.2.1 两种 XML 文件	(27)
2.2.2 XML 文件的解析器	(27)
2.2.3 检测 XML 文档	(28)
2.3 最终使用 XML 文档	(33)
第 3 章 初识 XML 文档	(36)
3.1 XML 文档样本	(36)
3.2 剖析 XML 文档	(37)
3.2.1 XML 文档的前言	(38)
3.2.2 XML 文档的主体	(40)
3.2.3 XML 文档的样式单	(40)
3.2.4 XML 文档的树状结构	(40)
第 4 章 XML 文档的构成	(43)
4.1 置标和字符数据	(44)
4.2 标记	(44)
4.2.1 标记的构成	(44)
4.2.2 标记的命名规则	(45)
4.2.3 标记的种类	(45)
4.3 元素	(46)

4.3.1	元素的构成	(46)
4.3.2	元素的嵌套	(47)
4.4	属性	(49)
4.4.1	属性的构成	(49)
4.4.2	属性名	(50)
4.4.3	属性值	(50)
4.5	元素和属性	(51)
4.6	预定义实体引用	(52)
4.7	注释	(53)
4.8	CDATA	(55)
第 5 章	创建结构完整的 XML 文档	(58)
5.1	文档的首行必须是 XML 声明	(59)
5.2	文档只有一个根元素	(60)
5.3	对称的控制标记	(61)
5.4	空元素的表示	(63)
5.5	元素必须正确嵌套	(64)
5.6	属性值必须用引号括起	(65)
5.7	关键字符的处理	(66)
5.8	置标中的英文大小写敏感	(67)
第 6 章	文件类型定义	(69)
6.1	什么是 DTD	(69)
6.2	DTD 的基本结构	(70)
6.3	在文档中引用 DTD	(72)
6.3.1	内部 DTD 引用	(72)
6.3.2	外部 DTD 引用	(73)
6.3.3	内部 DTD 和外部 DTD 的混合引用	(76)
6.4	DTD 中的元素声明	(78)
6.4.1	元素声明的语法	(78)
6.4.2	精确控制元素的内容	(80)
6.5	XML 文档的有效性	(89)
第 7 章	实体的定义和使用	(91)
7.1	什么是实体	(91)
7.2	内部实体的定义和使用	(92)
7.2.1	内部实体的定义和使用	(92)
7.2.2	在 DTD 中引用内部实体	(94)
7.3	外部实体的定义和使用	(94)
7.4	内部参数实体的定义和使用	(97)
7.5	外部参数实体的定义和使用	(100)

第 8 章 DTD 中的属性设定	(104)
8.1 什么是属性	(104)
8.2 属性的设定	(105)
8.3 属性缺省值的设定	(106)
8.3.1 使用#REQUIRED 关键字	(106)
8.3.2 使用#IMPLIED 关键字	(109)
8.3.3 使用#FIXED 关键字	(110)
8.4 属性的类型	(112)
8.4.1 CDATA 类型	(112)
8.4.2 Enumerated 类型	(113)
8.4.3 NMTOKEN 类型	(115)
8.4.4 NMTOKENS 类型	(115)
8.4.5 ID 类型	(116)
8.4.6 IDREF 类型	(118)
8.4.7 IDREFS 类型	(120)
8.4.8 ENTITY 类型	(121)
8.4.9 ENTITIES 类型	(122)
8.4.10 NOTATION 类型	(123)
8.5 XML 中的预定义属性	(124)
8.5.1 xml:space 属性	(124)
8.5.2 xml:lang 属性	(127)
第 9 章 在 XML 中使用多种文字	(129)
9.1 计算机中的语言和文字	(129)
9.1.1 字符集	(130)
9.1.2 操作系统和软件	(130)
9.1.3 字体	(131)
9.1.4 输入法	(131)
9.2 Unicode 字符集	(131)
9.3 UTF-8 字符集	(132)
9.3.1 一个字节的编码	(132)
9.3.2 两个字节的编码	(132)
9.3.3 三个字节的编码	(133)
9.3.4 UTF-8 字符集的优缺点	(133)
9.4 在 XML 中使用多种文字	(134)
9.4.1 使用英文外的其他文字	(134)
9.4.2 同时使用多种文字	(135)
第 10 章 CSS 样式单	(138)
10.1 什么是 CSS 样式单	(138)
10.1.1 样式单	(138)

10.1.2	CSS 样式单	(139)
10.1.3	CSS 的现状	(140)
10.2	XML 文档与样式单的链接	(140)
10.3	CSS 的构成实例	(143)
10.4	在 CSS 中使用中文	(144)
10.5	显示方式的设置	(151)
10.5.1	块级元素和内联元素	(151)
10.5.2	列表项元素	(153)
10.5.3	不可见元素	(154)
10.5.4	white-space 属性	(155)
10.6	字体属性的设置	(158)
10.6.1	font-family 属性	(158)
10.6.2	font-style 属性	(161)
10.6.3	font-size 属性	(162)
10.6.4	font-weight 属性	(168)
10.6.5	font-variant 属性	(170)
10.7	颜色属性的设置	(171)
10.8	文本属性的设置	(173)
10.8.1	text-align 属性	(173)
10.8.2	text-indent 属性	(174)
10.8.3	line-height 属性	(177)
10.8.4	vertical-align 属性	(178)
10.8.5	text-decoration 属性	(181)
10.8.6	text-transform 属性	(183)
10.8.7	word-spacing 属性	(184)
10.8.8	letter-spacing 属性	(185)
10.9	背景属性的设置	(187)
10.9.1	background-color 属性	(187)
10.9.2	background-image 属性	(188)
10.9.3	background-repeat 属性	(190)
10.9.4	background-attachment 属性	(192)
10.9.5	background-position 属性	(192)
10.10	框属性的设置	(194)
10.10.1	设置外边距	(195)
10.10.2	设置边框	(197)
10.10.3	设置内边距	(201)
10.10.4	设置框的大小	(203)
10.10.5	设置框的位置	(203)

第 11 章 XSL 样式单 (209)

11.1 关于 XSL	(209)
11.2 XSL 中的树形结构	(210)
11.3 XSL 样式单的构成	(212)
11.4 XSL 样式单的引用	(214)
11.5 XSL 样式单的工作过程	(214)
11.6 XSL 的模式	(217)
11.6.1 匹配根节点	(218)
11.6.2 匹配元素名	(218)
11.6.3 匹配属性	(219)
11.6.4 匹配指令	(219)
11.6.5 匹配注释	(220)
11.6.6 匹配文本	(222)
11.6.7 组合匹配	(222)
11.6.8 使用[]进行测试	(223)
11.7 XSL 的模板	(226)
11.7.1 xsl:apply-templates 元素	(227)
11.7.2 xsl:value-of 元素	(228)
11.7.3 xsl:for-each 元素	(229)
11.7.4 xsl:copy 元素	(232)
11.7.5 xsl:sort 元素	(232)
11.7.6 xsl:if 元素	(233)
11.7.7 xsl:choose 元素	(235)
11.7.8 xsl:pi 元素	(235)
11.7.9 xsl:comment 元素	(236)
11.7.10 xsl:element 元素	(236)
11.7.11 xsl:attribute 元素	(237)

第 12 章 XLink (238)

12.1 关于 XLink	(238)
12.2 定义链接元素	(239)
12.2.1 类型属性	(239)
12.2.2 链接地址属性	(240)
12.2.3 语义属性	(240)
12.2.4 行为属性	(240)
12.2.5 身份属性	(241)
12.2.6 方向属性	(241)
12.2.7 元素类型和属性	(241)
12.3 简单链接	(242)
12.4 扩展链接	(243)

第 13 章 XPointer (247)

- 13.1 关于 XPointer (247)
- 13.2 实例例程 (248)
- 13.3 绝对位置寻址 (252)
 - 13.3.1 root() (252)
 - 13.3.2 id() (252)
 - 13.3.3 html() (253)
- 13.4 相对位置寻址 (253)
 - 13.4.1 child() (254)
 - 13.4.2 descendant() (255)
 - 13.4.3 ancestor() (255)
 - 13.4.4 following(n) (255)
 - 13.4.5 preceding(n) (256)
 - 13.4.6 psibling() (257)
 - 13.4.7 fsibling() (257)
 - 13.4.8 根据编号选择目标 (258)
 - 13.4.9 根据节点类型选择目标 (259)
- 13.5 属性位置寻址 (262)
- 13.6 字符串位置寻址 (263)

第 14 章 命名域 (265)

- 14.1 命名域的由来 (265)
- 14.2 定义命名域 (268)
- 14.3 多命名域和缺省命名域 (270)
- 14.4 命名域的作用范围 (272)
- 14.5 使用命名域引用 HTML 标记 (273)

第1章 初识 XML

本章面向初学者，介绍 XML 的一些最基本的概念。主要包括以下内容：

- 直接进入 XML
- 从 HTML 到 XML
- 内容和显示的分离
- 为什么要用 XML

1.1 直接进入 XML

XML 是 Extensible Markup Language 的缩写，中文含义为“可扩展标记语言”。顾名思义，XML 首先是一种标记语言，其次，它是一种可扩展的标记语言。要准确地理解 XML，首先要理解什么是“标记语言”。

1.1.1 理解标记语言

标记语言 (Markup Language) 是指在普通文本中加入一些具有特定含义的标记 (Tag)，以对文本的内容进行标识和说明的一种文件表示方法。

在文档中加入标记或者说使用标记语言的目的，是为了对文档的内容进行整理、细化，一般来讲，加入的标记越多越细致，文档提供给人或计算机的信息就越多越丰富，人或计算机对文档的理解就越准确越深刻。

以下面的文本为例：

“《XML 完全实例教程》，卢啸龙，男，1972 年 11 月，山东莱阳，电子工业出版社，2001 年 4 月，由浅入深地讲述了 XML 的相关知识。”

对于这段文字的意义，可能有各种各样的理解。但如果在文本中加入一些标记，对它

的理解将会容易而准确得多。

【例程 1.1】

<书名>《XML 完全实例教程》</书名>

<作者信息>卢啸龙, 男, 1972 年 11 月生, 山东莱阳</作者信息>

<出版社>电子工业出版社</出版社>

<出版日期>2001 年 4 月</出版日期>

<内容提要>由浅入深地讲述了 XML 的相关知识</内容提要>。

进一步增加标记, 以上文本的内容将更加清晰而明确, 参见例程 1.2。

【例程 1.2】

<书>

<书名>《XML 完全实例教程》</书名>

<作者>

<姓名>卢啸龙</姓名>

<性别>男</性别>

<出生日期>1972 年 11 月</出生日期>

<祖籍>山东莱阳</祖籍>

</作者>

<出版社>电子工业出版社</出版社>

<出版日期>2001 年 4 月</出版日期>

<内容简介>由浅入深地讲述了 XML 的相关知识</内容简介>

</书>

上面例子中标记的作用, 主要还不是为便于读者阅读而是帮助读者理解。一段意义表达并不十分清楚、容易引起误解的文字在添加标记后, 其意义变得清晰明确, 易于被人准确理解。但在实际中, 使用标记语言进行描述的文档, 其主要读者并不是人, 而是计算机。对于计算机而言, 它并不能理解任何纯粹的文本, 无论该文本对于人来讲其意义表达得是如何清楚。如下述文本:

“书名 《XML 完全实例教程》, 作者姓名 卢啸龙, 性别 男, 出生日期 1972 年 11 月, 祖籍 山东莱阳, 电子工业出版社 2001 年 4 月出版, 主要内容 由浅入深地讲述了 XML 的相关知识。”

对人来讲, 该段文字的意义表达与例程 1.2 所表达的是一样的。但对于计算机来讲, 对这段文字的认识, 恐怕仍旧停留在“这是一个字符串”的水平上。至于字符串的内容, 计算机是一无所知的。如果在该段文字中添加标记, 改写成例程 1.2 的格式, 再经简单处理, 则计算机虽然仍旧不能理解“性别”为何物, “祖籍”是什么意思, 但至少当我们问

计算机“性别”、“祖籍”的时候，计算机能够毫不犹豫地回答出“男”、“山东莱阳”。

所以说，标记语言对文档的意义进行了分析和细化，使得相应的计算机应用软件对文本的意义也就具有了一定程度的“理解”，这使得计算机所能够处理的信息又得到了极大地增长。相应地，计算机的“智能”似乎也提高了。

1.1.2 理解 XML

XML 的英文拼写为 Extensible Markup Language，是一种可扩展的标记语言。这里的“语言”与编程中“计算机语言”的概念有一定的差别。“计算机语言”与日常生活中的“语言”的概念更为接近。不同的语言，有不同的词汇集和语法。使用某种语言进行编程，即使用这种语言的特定的词汇，按照这种语言的语法来书写计算机指令，所书写的指令中不含有这种语言词汇集以外的词语，除非作为指令的参数。

作为“可扩展标记语言”的 XML，虽然也称为“语言”，但在使用的时候，它并没有特定的词汇集，也不用书写计算机指令。它只是按照一定的规则，在文本文档中添加标记，有时候为了其他应用软件能够对添加过标记的文档进行识别和处理，在文档之内或之外，按照一定的规则和语法，对所使用的标记和标记之间的关系进行定义和说明；有时候根据需要，还要在文档之外，编写规定文档的显示方案的文件。

以上的关于“使用 XML”的陈述，表达且蕴涵了如下意义：

(1) 作为一种标记语言，XML 最基本、最主要的“功能”就是在文档中添加标记，而且，标记的添加必须遵循一定的规则。如例程 1.3 所示，在<>和</>里面的文本，就是一些“标记”。很显然，这些标记的使用遵循了某些规则，最直观的规则就是标记必须成对出现，如<book>和</book>、<name>和</name>、<country>和</country>等。

【例程 1.3】

```
<?xml version="1.0" standalone="yes"?>
<book>
  <name>XML:Step by Step</name>
  <author>
    <name>Kiem</name>
    <sex>male</sex>
    <age>45</age>
    <country>Australia</country>
  </author>
```

</book>

(2) 所使用的标记都是非预定义的, 即用户自己定义的。只要遵守 XML 的标记命名规则, 可以在文档中添加任何标记。如在例程 1.3 中, 可以将<book>、</book>标记改为<BookInfomation>和</BookInfomation>, 当然也可以改为和。用户可以自定义标记, 这就是 XML 称为“可扩展”标记语言的由来。

(3) 对于自定义的标记, 用户可在文档内, 也可以在文档之外进行说明。当然也可以不进行说明, 如例程 1.3。无论在文档内部, 还是在其外部, XML 对所使用的标记进行说明的部分, 称为 DTD (Document Type Definition), 即文档类型定义。DTD 定义了用户所使用的所有标记以及标记之间的逻辑关系, 同时也就定义了文档的逻辑结构。一个 XML 文档若包含了 DTD, 应用程序就可以根据 DTD 的定义来检查文档的完整性和正确性。

(4) XML 文档的标记只描述文档的内容, 具体说, 只描述文档的结构和意义, 不描述其如何显示、输出等格式化信息。若要格式化文档的输出, 需要另外编写控制其输出的文件, 即样式表文件。如例程 1.3 所描述的那样, 它在 IE5.5 中的显示结果如图 1.1 所示。

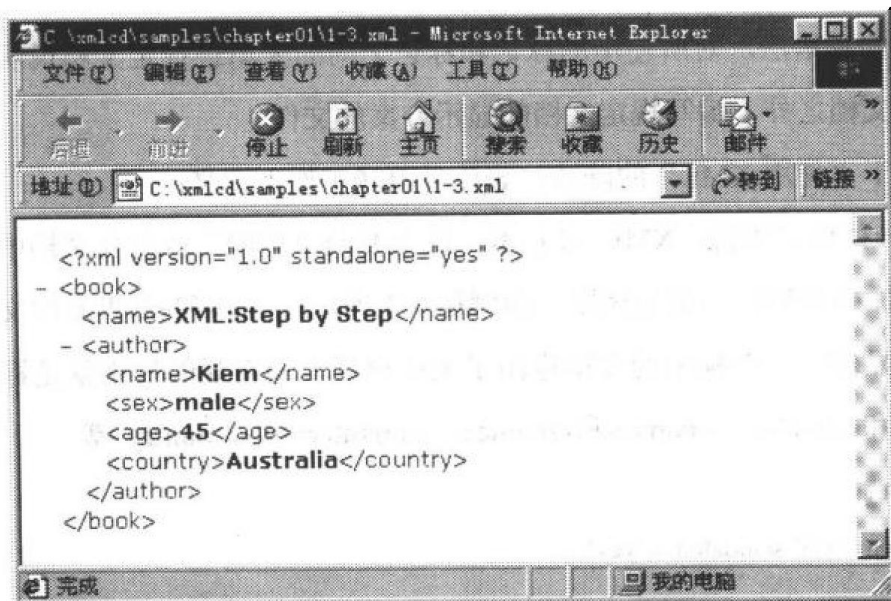


图 1.1 例程 1.3 在 IE5.5 中的显示

在本例中, 文档的原始内容连同 XML 指令和标记都被一同显示出来。虽然浏览器对文档进行了语法分析, 文档内容、指令和标记分别被显示成不同的颜色, 但这并不是我们想要的结果。一般来讲, 我们需要显示的只是文档的原始内容, 指令和标记作为附加的信息在实际显示时应该被隐藏起来, 并且书名和作者信息等不同级别的信息要使用不同的字

体和字号。要达到这个目的，就要为文档编写样式表。

至此，我们对于 XML，除了“可扩展标记语言”之外，应该有更深一步的理解：XML 实质上是一种信息的表示方法，更具体地说，XML 确定了一种文件格式，就像 TXT 文件和 RTF 文件格式一样，它规定了使用这种文件格式所应该遵守的规则。

1.2 从 HTML 到 XML

HTML 的全称是“HyperText Markup Language”，即“超文本标记语言”，它主要用来进行基于 Internet 的网页设计，也就是在 Internet 上发布信息。HTML 也是一种标记语言，实质上也是一种文件格式，Internet 上最大量、最精彩、最吸引人的内容（也可以说是信息，实质上就是数据）都是使用这种文件格式进行组织和表达的。可以说，易学易用的 HTML 是当今 Internet 能够飞速发展的重要原因之一。

请见例程 1.4 所描述的 HTML 源文档。

【例程 1.4】

```
<html>

<head>
<meta http-equiv="Content-Type" content="text/html; charset=gb2312">
<style type="text/css">
<!-- A:link {text-decoration: none; color:#006677; font-family: 宋体}
      A:visited {text-decoration: none; color: #000000; font-family: 宋体}
      A:active {text-decoration: none; color: #FF0000; font-family: 宋体}
      A:hover {text-decoration: underline; color: FF0000}
      body {font-size=9pt}
      TH{FONT-SIZE: 9pt}
      TD{ FONT-SIZE: 9pt}-->
</style>
<title>奥运快讯</title>
</head>

<body topmargin="0" leftmargin="0" bgcolor="#FFFFFF" background="beijing.gif">

<table border="0" width="773" cellspacing="0" cellpadding="0">
  <tr>
    <td width="298" valign="top"><p align="center"><br>
```