



SHEN RU
QIAN CHU

深入浅出 XML

■ 老虎工作室 贾小珠 宋立智 赵玮 等 编著

人民邮电出版社
www.pptph.com.cn



深入浅出 XML

■ 老虎工作室 贾小珠 宋立智 赵玮 等 编著

人民邮电出版社

图书在版编目 (CIP) 数据

深入浅出 XML/贾小珠等编著. —北京:人民邮电出版社, 2001.9
ISBN 7-115-09676-7

I. 深... II. 贾... 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 064573 号

内 容 提 要

可扩展标记语言 XML 是一种新的 Web 开发语言,利用它可以通过 Internet/Intranet 创建并发布信息。本书首先介绍了标记语言和 XML 的基本概念,并通过一个简单的实例说明了 XML 文档的基本组成,然后详细介绍了 XML 的语法规则、文档类型定义、XML Schema、级联样式单、可扩展样式语言等 XML 的基本内容,最后介绍了 W3C 的 DOM 接口标准、Microsoft 的 XML DOM 以及 XML 在 ASP、VB 和 HTML 中的应用。

本书内容由浅入深,在讲解基本概念和基本知识的同时给出了大量实例。本书适合具有一定网络基础知识的读者阅读,也可作为工程设计人员的技术参考书或大中专学生的自学教材。

深入浅出 XML

◆ 编 著 老虎工作室 贾小珠 宋立智 赵玮 等
责任编辑 姚彦兵
执行编辑 李永涛

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@pptph.com.cn
网址 <http://www.pptph.com.cn>
读者热线 010-67129212 010-67129211(传真)
北京汉魂图文设计有限公司制作
北京鸿佳印刷厂印刷
新华书店总店北京发行所经销

◆ 开本:787×1092 1/16
印张:19.25
字数:460千字 2001年9月第1版
印数:1-5000册 2001年9月北京第1次印刷

ISBN 7-115-09676-7/TP·2482

定价:28.00元

本书如有印装质量问题,请与本社联系 电话:(010)67129223



老虎工作室

主 编：沈精虎

编 委：许曰滨 黄业清 姜 勇 冯 辉 宋雪岩
陈 杉 郭剑峰 蔡汉明 王 宁 宋一兵
李 仲 周 锦 王 钰 高长铎 张艳花

内容和特点

Internet 是目前全球最大的计算机网络,有着丰富的资源,已经成为科学研究、企业宣传、商务往来和人们日常生活中不可缺少的信息交流工具。WWW 是 Internet 上的一个重要应用,其信息主要通过网页发布。HTML (超文本标记语言)是最初编写网页的主要语言,对 Internet 的发展起到了巨大的推动作用。然而,HTML 的局限性使其越来越无法满足网上日益增多的信息检索和特殊显示效果的需求。为了弥补 HTML 的不足,W3C (万维网联盟)在 1998 年 2 月将 XML (可扩展标记语言)作为下一代网页发布的语言,同时发布了 XML 1.0 版。XML 是一个精简的 SGML (标准通用标记语言),它保留了 SGML 的可扩展功能,并具有 HTML 的易用性。

本书主要介绍了 XML 的基础知识和常用的应用技术,在内容安排上由浅入深,并列举了大量的实例。全书共分 10 章,主要内容如下:

- 第 1 章:介绍了标记语言和 XML 的基本概念。
- 第 2 章:详细介绍了 XML 的基本语法,包括元素、属性、XML 文档的逻辑结构和物理结构等内容,并通过对一个完整的 XML 文档的分析,说明了 XML 文档的结构。
- 第 3 章:介绍了 DTD 的基本要素、组成、实体、名称空间等内容。
- 第 4 章:介绍了 XML Schema 的基本知识和使用方法。
- 第 5、6 章:详细介绍了 CSS 及其使用方法。其中,第 5 章介绍 CSS1,主要内容有继承性、字体属性、颜色属性、文本属性等;第 6 章介绍 CSS2,包括 CSS2 的新增特性、有声样式单和可视格式化等。
- 第 7 章:介绍了 XSL 的使用以及 XSL 模式语言。
- 第 8 章:简单介绍了 XML 链接语言的基本概念和寻址方式。
- 第 9 章:介绍了 W3C 的 DOM 模型和 Microsoft 的 XML DOM。
- 第 10 章:介绍了 XML 的一些常用应用技术,包括数据岛技术和在 ASP、VB 中的应用技术。

读者对象

本书适合具有一定网络基础知识的读者阅读,也可以作为工程设计人员的技术参考书和大中专院校学生的自学教材。

本书实例的原始程序存放在老虎工作室主页 (<http://www.laohu.net>) 与本书相关的主题中,有兴趣的读者可以去下载。

本书由贾小珠、宋立智、赵玮、赵煜筠、侯东昌和蔺德军编写，在编写过程中得到了潘振宽、许曰滨同志的帮助和支持，在此表示感谢。

感谢您选择了本书，也希望您把对本书的意见和建议告诉我们。
老虎工作室主页 <http://www.laohu.net>，电子函件 xyzy@263.net。

老虎工作室

2001年8月

第 1 章 XML 概述	1
1.1 标记语言	2
1.1.1 标记语言的含义	2
1.1.2 标准通用标记语言	3
1.1.3 HTML 简介	3
1.2 XML 简介	4
1.2.1 XML 的含义	4
1.2.2 XML 的设计目标	5
1.2.3 XML 的一些常用术语和基本概念	6
1.2.4 XML 的未来	6
1.3 XML 文档编辑工具	7
1.3.1 通用编辑工具	8
1.3.2 专用编辑工具	8
1.4 简单的 XML 文档	10
第 2 章 XML 语言基础	15
2.1 语法基础	16
2.1.1 大小写	16
2.1.2 数据	16
2.1.3 CDATA 段	17
2.1.4 注释	18
2.2 序码 (Prolog)	19
2.2.1 XML 声明	19
2.2.2 文档类型声明	19
2.3 处理指令 (PI)	21
2.4 元素及其属性	22
2.5 元素的层次结构	24
2.6 实体简介	25
2.7 XML 文档的结构	26
2.7.1 逻辑结构	26
2.7.2 物理结构	27
2.8 有效的 XML 文档	28

2.9 结构完整的 XML 文档实例.....	29
第 3 章 创建 DTD	37
3.1 基本概念.....	38
3.2 元素.....	38
3.2.1 元素声明.....	39
3.2.2 空元素.....	39
3.2.3 ANY 元素.....	39
3.2.4 #PCDATA.....	41
3.2.5 子元素定义.....	42
3.2.6 混合内容的元素.....	47
3.3 属性.....	47
3.3.1 属性的声明.....	47
3.3.2 属性的类型.....	48
3.3.3 属性的缺省值.....	50
3.4 DTD 中的实体.....	51
3.4.1 实体的概念.....	52
3.4.2 内部通用实体.....	52
3.4.3 外部通用实体.....	55
3.4.4 内部参数实体.....	56
3.4.5 外部参数实体.....	57
3.5 名称空间.....	57
第 4 章 XML Schema	61
4.1 XML Schema 概述.....	62
4.2 Schema 的结构.....	63
4.3 XML Schema 的元素.....	63
4.3.1 ElementType 元素.....	64
4.3.2 element 元素.....	66
4.3.3 group 元素.....	69
4.3.4 description 元素.....	73
4.4 属性.....	73
4.5 Schema 对名称空间的支持.....	76
第 5 章 级联样式单 CSS1	79
5.1 什么是 CSS.....	80
5.2 CSS 样式单与 XML 文档的链接.....	82
5.3 选择元素.....	83
5.3.1 选择符 (Selector).....	83

5.3.2	伪元素 (pseudo-element)	84
5.3.3	伪类 (Pseudo-class)	87
5.3.4	上下文关联选择符和 CSS 注释	89
5.4	继承性	91
5.5	CSS 中的单位	92
5.5.1	长度单位	92
5.5.2	URL 单位	93
5.5.3	颜色单位	94
5.6	块、内联或列表项元素	95
5.7	字体属性	99
5.8	font 简略属性	106
5.9	颜色和背景属性	107
5.10	文本属性	115
5.11	框属性	123
5.12	滤镜和渐变	133
5.12.1	滤镜和渐变	134
5.12.2	滤镜和渐变属性	134
第 6 章	级联样式单 CSS2	145
6.1	扩展定位与可视格式化	146
6.1.1	CSS2 定位	146
6.1.2	扩展定位与可视属性	147
6.2	页面格式与打印支持	154
6.3	音频属性	156
第 7 章	扩展样式语言 XSL	159
7.1	XSL 简介	160
7.2	XSL 变换	160
7.2.1	树型结构	160
7.2.2	XSL 样式单文档	162
7.2.3	XSL 样式单和 XML 文档的链接	163
7.3	XSL 模板	164
7.3.1	xsl:apply-templates 元素	165
7.3.2	模板的 select 属性	167
7.4	在 XSL 中对节点值的简单操作	170
7.4.1	使用 xsl:value-of 来计算节点值	170
7.4.2	使用 xsl:for-each 元素处理多个元素	173
7.4.3	使用 order-by 模式对输出节点排序	175
7.5	节点的匹配模式	179

7.5.1	匹配根模式.....	179
7.5.2	匹配元素名.....	179
7.5.3	使用 “/” 字符匹配子节点.....	182
7.5.4	使用 “//” 符号匹配子元素.....	184
7.5.5	使用 “@” 符号匹配属性.....	186
7.5.6	使用 comment() 匹配注释.....	187
7.5.7	使用 text() 匹配文本节点.....	188
7.5.8	“或” 操作符(l).....	189
7.6	条件模板.....	191
7.6.1	条件表达式.....	191
7.6.2	XSL 中的常用节点方法.....	195
7.6.3	在模板输出中使用条件.....	196
7.6.4	在模板属性中使用条件.....	199
7.7	XSL 格式化对象.....	201
7.7.1	XSL 格式化语言概述.....	201
7.7.2	格式化对象及其属性.....	201
7.7.3	fo 名称空间.....	202
7.7.4	格式化对象实例.....	203
第 8 章	XML 链接.....	205
8.1	XML 链接简介.....	206
8.2	XLink.....	206
8.2.1	简单链接.....	206
8.2.2	扩展链接.....	208
8.3	XPointer.....	210
8.3.1	绝对路径.....	210
8.3.2	相对路径.....	211
第 9 章	文档对象模型.....	213
9.1	DOM 概述.....	214
9.1.1	节点.....	215
9.1.2	DOM 中的数据类型.....	216
9.2	VBScript 简介.....	216
9.2.1	变量与常量.....	216
9.2.2	运算符.....	219
9.2.3	流程控制.....	220
9.2.4	VBScript 中的过程和函数.....	224
9.3	JavaScript 简介.....	225
9.3.1	变量和常量.....	225

9.3.2	运算符	226
9.3.3	程序流程控制	228
9.3.4	JavaScript 中的函数	231
9.4	W3C DOM 接口	231
9.4.1	Document 接口	232
9.4.2	Node 接口	233
9.4.3	NodeList 接口	239
9.4.4	NamedNodeMap 接口	239
9.4.5	DocumentFragment 接口	241
9.4.6	CharacterData 接口	241
9.4.7	Attr 接口	242
9.4.8	Element 接口	242
9.4.9	其他接口介绍	243
9.5	Microsoft XML DOM	244
9.5.1	使用 Microsoft XMLDOM ActiveX 数据对象	244
9.5.2	XMLDOMDocument 对象	244
9.5.3	XMLDOMElement 对象	246
9.5.4	XMLDOMNode 对象	247
9.5.5	XMLDOMNodeList 对象	249
9.5.6	XMLDOMNamedNodeMap 对象	250
9.5.7	XMLDOMAttribute 对象	251
9.5.8	XMLDOMText 对象	253
9.5.9	XMLDOMParseError 对象	253
9.5.10	其他对象介绍	254
第 10 章	XML 的应用	255
10.1	ASP 对象	256
10.1.1	ASP 内置对象概述	258
10.1.2	Response 对象	258
10.1.3	Request 对象	260
10.1.4	Server 对象	261
10.2	ADO 对象	262
10.2.1	Connection 对象	264
10.2.2	Recordset 对象	266
10.2.3	Fields 集合和 Field 对象	269
10.3	数据岛	271
10.3.1	数据岛的加载	271
10.3.2	数据岛的数据绑定	276
10.3.3	数据岛中的记录集	277

10.4 XML 在 ASP 中的应用.....	278
10.4.1 利用 ASP 生成动态 XML 页面	278
10.4.2 利用 FSO 存取 XML.....	280
10.4.3 利用 DOM 存取 XML.....	283
10.5 XML 在 VB 中的应用	285
10.5.1 加载 XML 文档.....	285
10.5.2 遍历 XML 文档元素	287
10.5.3 操作 XML 文档元素.....	289
10.5.4 下载和上载 XML 文档.....	291



第1章 XML 概述

主要内容

- 标记语言
- 标准通用标记语言 (SGML)
- 可扩展标记语言 (XML)
- XML 的设计目标
- XML 中的常用术语
- XML 文档编辑工具



万维网 (WWW, World Wide Web) 又称环球信息网络, 是基于超文本 (Hyper Text) 方式发布和获取信息的网络, 由欧洲核子物理研究中心 (CERN) 研制, 并得到了迅速而广泛的应用。现在大多数 Web 页面都使用 HTML (Hyper Text Markup Language) 语言编写, 但随着计算机技术的发展, HTML 的局限性越来越明显, 例如不能满足人们对信息检索的要求。为了弥补 HTML 的不足, 1996 年, 万维网联盟 (W3C, The World Wide Web Consortium) 开始对 HTML 的后续语言进行研究, 并于 1998 年 2 月正式推出了 XML (可扩展标记语言, eXtensible Markup Language)。为了全面了解 XML, 本章首先介绍标记语言、标准通用标记语言和 HTML 存在的缺陷, 然后介绍 XML 及其特点、优势和将来的发展趋势, 并在最后给出一个简单的 XML 应用实例。

1.1 标记语言

在学习 XML 之前, 首先介绍一下标记语言, 因为正确地认识和理解标记语言是学习 XML 的基础。

1.1.1 标记语言的含义

HTML 就是一种典型的标记语言。下面我们以一个简单的 HTML 文档为例来讲解标记语言的语法结构。

【例 1.1】 一个简单的 HTML 文档。

```
<html>
  <head>
    <title>标记语言示例</title>
  </head>
  <body>
    <font size=10>这是一个 HTML 文档</font>
  </body>
</html>
```

该例子的显示结果如图 1-1 所示。

从例 1.1 的文档中可以看出, 这个 HTML 文档由两个基本部分组成, 即由一对尖括号“<>”括起来的特殊符号和由两对尖括号包围的文本。由尖括号括起来的部分称为“标记 (Tag)”, 它们并不在浏览器中显示, 如图 1-1 所示, 但表述了一种功用, 即提示浏览器该怎样处理后面的数据。由两对尖括号包围的文本称为数据, 它们可以在浏览器中显示, 但受到标记的约束。标记语言就是这些标记的集合。

许多标记通常都成对出现, 即由开始标记和结束标记构成, 例如上面的“title”。成对标记通常用来描述内容, 而单个标记通常用来表示向文档中插入某些信息。所有标记都包围在一对尖括号中, 结束标记是在“<”的后面插入一个斜杠“/”。



图1-1 例 1.1 文档在 IE 5.0 中的显示结果

目前的标记语言通常可以分为特殊用途的标记语言和一般用途的标记语言两大类。特殊用途的标记语言主要应用于某种特定的应用软件，而且通常带有描述文档格式的信息。例如，RTF 就是为 WordPad 软件设计的标记语言，而 HTML 是为 Web 页面设计的标记语言，并以浏览器为应用软件。一般来说，用特殊用途的标记语言书写的文档只能由某个特定的应用软件读取和解释，很难由其他的应用软件解释，因此，用这种标记语言描述的文档难于移植。一般用途的标记语言的典型代表是 SGML，它实际上也是所有标记语言的鼻祖，开发它的初衷就是为了解决文档的移植问题。一般用途的标记语言主要用来描述文档内容，而至于如何显示和处理文档则由应用程序来完成。同时一般用途的标记语言也被限制成文本方式，因此，它解决了跨平台问题，因为大多数不同平台下的计算机都可以阅读文本文件。

1.1.2 标准通用标记语言

标准通用标记语言（SGML，Standard Generalized Markup Language）是最原始的标记语言，出现于 20 世纪 60 年代，由 IBM 公司的 Charles Goldfarb、Ed Mosher 和 Ray Lorie 创建。当时他们是为了解决在某个平台下建立的法律文档向另一个平台移植的问题，由于这些法律文档是在不同系统中使用专用格式创建的，因此不可能在整个公司间移植。他们研究的结果揭示了任何可互操作系统的 3 个主要特征：

- 文件处理程序需要支持一个公共的文档格式；
- 这个文档格式对它们的领域应该是专有的；
- 为了达到高度可靠性，文档格式必须遵守特殊规则。

他们创建的第一个语言被称为通用标记语言（GML，Generalized Markup Language），且只在 IBM 内部使用。在此后的若干年中，他们继续研究 GML，并于 20 世纪 70 年代后期开发出了 SGML。国际标准化组织（ISO）于 1986 年批准和采用了 SGML，使之成为一个官方标准——ISO 8879。

1.1.3 HTML 简介

1989 年，设在瑞士日内瓦的欧洲核子物理实验室（CERN）的英国研究人员 Tim



Berners-Lee 对该研究中心的大量研究资料不可移植和不兼容深感失望。他希望能开发出一种工作在某一个系统上,而且和平台无关、容易使用的语言。在 SGML 思想的启发下, Tim Berners-Lee 在 1991 年建立了 HTML。

HTML 建立在超文本和标记语言的概念之上。超文本是一种将文档链接在一起的方法,通过点击文档中的文本块可以在文档之间移动。HTML 是一种与平台无关的语言,它使用标记描述文档和超文本之间的链接,达到链接和在文档之间移动的目的。

HTML 以其简洁、易学、易用的特点,在 Web 应用中获得了巨大成功。然而,随着电子商务、电子图书和远程教育等全新领域的出现和迅速发展,HTML 显得越来越力不从心,这主要是由其本身的缺陷所造成的。HTML 的缺陷主要体现在以下几个方面:

1. 难以扩展

HTML 是一种固定格式的超文本标记语言,它的标记集合是固定的,用户不能新增有意义的并能供其他人使用的标记,故难以扩展。

2. 交互性差

当用户与应用程序进行交互,并想以不同的角度观察同一数据时,HTML 不能对现有的数据进行本地处理,只能请求服务器重传另一个页面,这无疑将加重服务器的负担。

3. 单向的超级链接

HTML 的超级链接是单向的,只能在它的某一端(源端)定义,并确定了它的另一端(目的端),用户只能从源端走到目的端。

4. 缺乏对双字节或多国文字的支持

HTML 缺乏对双字节或多国文字的支持,或者说支持不够。例如,中文信息页面在不同的平台上会出现格式不同等问题。

由于 HTML 存在以上不足,人们才开始研究能改进或取代 HTML 的 Web 页面制作语言。其中最有效,并已初步投入使用的有:可扩展标记语言(eXtensible Markup Language),简称 XML;级联样式单(Cascading Style Sheet),简称 CSS;动态 HTML(Dynamic HTML),即 DHTML。这些语言从不同的角度弥补了 HTML 的不足。XML 有利于信息的表达和结构化组织,从而使数据搜索更有效;CSS 解决了 Web 页面的继承和显示问题;DHTML 则主要用于 Web 页面的动态显示。

1.2 XML 简介

本节介绍 XML 的基本概念、设计目标和一些常用术语。

1.2.1 XML 的含义

1998 年 2 月发表的 XML 1.0 规范中有如下描述:

“eXtensible Markup Language (XML) 是 SGML 的一个子集……其目标是能够以目前 HTML 可能实现的方式在 Web 上使用、接受和处理通用 SGML。XML 的设计目标是实现



简便并且能与 SGML 和 HTML 共同操作。”

也就是说，XML 是一个精简的 SGML，它保留了 SGML 的可扩展功能，将 SGML 的丰富功能和 HTML 的易用性结合到 Web 应用中。简单地讲，XML 是一种定义标记语言的元标记语言，是一套定义语义标记的规则，即定义了用于定义其他与特定领域有关的、语义的、结构化的标记语言的语言。

XML 不是 HTML 的扩展，而是一种用于定义标记语言的类似于 SGML 的语言。XML 不再使用固定标记，而允许定义数量无限的标记来描述文档中的资料。HTML 是 Web 中显示数据的通用方法，而 XML 提供了一个直接处理 Web 数据的通用方法。HTML 着重描述 Web 页面的显示格式，而 XML 着重描述 Web 页面的内容。

为了使编写的 Web 页面成为一个有效的 XML 文档，文档中每一个添加的标记必须记入 DTD (Document Type Definitions, 文档类型定义) 中。XML 主要用于描述 Web 页面的内容，因此它必须与多种相关技术相互作用才能达到正确的显示效果和链接。这些技术主要包括 CSS (Cascading Style Sheet, 级联样式单)、XSL (eXtensible Style Language, 可扩展样式语言) 和 XLL (eXtensible Linking Language, 可扩展链接语言)。CSS 可以使 XML 页面具有更复杂的表现形式。XSL 可以将 XML 数据翻译为 HTML 风格或其他风格，并提供一种级联样式单的功能，使开发者能构造出具有表达层次结构的 Web 页面。XLL 是 XML 的链接语言，它与 HTML 的链接相似，但功能更强大，同时支持可扩展的链接和多方向的链接。

1.2.2 XML 的设计目标

XML 来源于 SGML，由 XML 工作组开发，该工作组的前身是“SGML 编辑审查委员会”，于 1996 年由万维网联盟 W3C 主持成立。

XML 的设计目标如下：

- 可在 Internet 上直接使用；
- 支持各种各样的应用程序；
- 与 SGML 兼容；
- 处理 XML 文档的程序应该容易编写；
- XML 中的可选项应尽可能少，理想情况下应为零；
- XML 文档清晰易读；
- 应该易于设计；
- XML 的设计应该井井有条并且简明扼要；
- XML 文档应易于创建；
- XML 标记的简洁性较为次要。

这些目标实现了 XML 保留 SGML 精致而可扩展的特性，保证了软件开发人员能够在合理的时间限度内创建主要的 XML 工具，同时也保证了 Web 页面的设计不复杂。