

普通高等院校计算机类专业规划教材·精品系列

XML程序设计

XML CHENGXU SHEJI

(第二版)

王震江 编著



中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

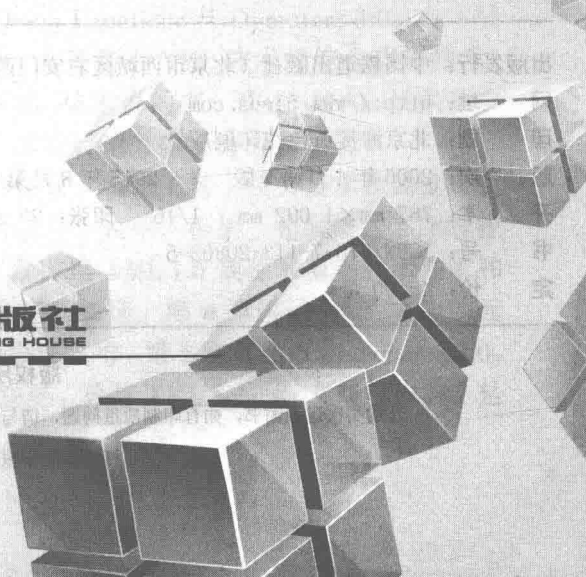
普通高等院校计算机类专业规划教材·精品系列

XML 程序设计

(第二版)

王震江 编著

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE



XML 已经在 Web 应用、新型数据库系统、移动互联网技术、网络数据交换和跨平台应用中发挥越来越重要的作用,已成为风靡全球的电子商务的核心技术。

本书全面论述了 XML 技术的主要方面:XML1.0 规范的基本技术要求和规定,XML DTD、XML Schema 2.0、XPath 3.0、XDM 3.0、XPath、XQuery Functions 与 Operators 3.0、XSLT 2.0、XML DOM Level 3、XML 的数据源技术和 XML 数据库概念,以及 HTML、JavaScript、XMLHttpRequest 等 Web 实现和 AJAX 技术,最后给出一个应用实例。

本书体系完整,采用最新的技术规范,叙述简单扼要,书中大量实例有助于读者对 XML 技术的理解和学习。本书适合作为本科院校 IT 类专业、电子商务专业的技术基础课教材,也可作为高等职业院校相关专业教材和社会培训用书。

图书在版编目(CIP)数据

XML 程序设计/王震江编著. —2 版. —北京:

中国铁道出版社, 2015. 8

普通高等院校计算机类专业规划教材. 精品系列

ISBN 978-7-113-20803-5

I. ①X… II. ①王… III. ①可扩展语言—程序设计—高等学校—教材 IV. ①TP312

中国版本图书馆 CIP 数据核字 (2015) 第 181507 号

书 名: XML 程序设计 (第二版)

作 者: 王震江 编著

策 划: 周海燕

读者热线: 400-668-0820

责任编辑: 周海燕 冯彩茹

封面设计: 穆 丽

封面制作: 白 雪

责任校对: 包 宁

责任印制: 李 佳

出版发行: 中国铁道出版社 (北京市西城区右安门西街 8 号, 邮政编码 100054)

网 址: <http://www.51eds.com>

印 刷: 北京海淀五色花印刷厂

版 次: 2006 年 4 月第 1 版 2015 年 8 月第 2 版 2015 年 8 月第 1 次印刷

开 本: 787 mm×1 092 mm 1/16 印张: 23.25 字数: 509 千

书 号: ISBN 978-7-113-20803-5

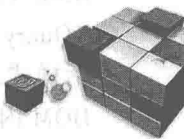
定 价: 45.00 元

版权所有 侵权必究

凡购买铁道版图书,如有印制质量问题,请与本社教材图书营销部联系调换。电话: (010) 63550836

打击盗版举报电话: (010) 51873659

前言 (第二版)



XML (eXtensible Markup Language) 1.0 技术规范从 1998 年 2 月 10 日问世以来,在 XML 应用技术方面已经发生了翻天覆地的变化,XML 作为新一代的数据交换标准和交换文本已经成为今天网络数据交换的新标准,基本取代了上一代的数据交换标准 EDI。2004 年 10 月 28 日发布的 XML Schema 2.0 技术已经在应用方面日趋成熟,现在在 W3C 发布的新的 XML 各种规范中,出现了大量以 XML Schema 数据模型为基础的描述。XML 的应用层出不穷,已经在计算机信息处理、非结构化数据表示、异种平台数据交换与传输、Web 编程、网络应用编程、跨平台编程技术等方面得到十分广泛的应用。作为一类文本文件,XML 文档和数据可以在 Internet 上不受任何阻挡,作为各种应用的数据基础承载各种应用,因而以 XML 为内核的半结构数据库技术,在 Web 和网络应用程序中具有互操作性好、数据表现灵活、数据类型丰富等特点使得支撑 XML 的各种技术规范和应用(如 AJAX)不断推进和完善,受到行业的大力支持和软件编程人员的欢迎。

本书从第一版出版至今已经 9 年。这 9 年中 XML 从 1.0 第一版连续更新到第五版。伴随 XML 的发展,一些原来的技术被弃之不用,一些当时不成熟的技术陆续更新。此次修订,基本上按照 W3C 最新的技术规范对第一版的内容进行了全面修改,对于原来的内容进行了更新,对原书没有的内容进行了增补。

因为本书的重点放在 XML 文档的 Web 应用方面,不是注重于太多的理论,但几乎所有关于 XML 的应用都离不开 HTML。因此,为了那些没有网页设计知识和基础的读者能够较好地掌握 XML 的应用,本书增加了 HTML 的内容,在文字、段落、标题、表格、表单、框架、列表、动画与声音等方面简要介绍了 HTML 的基本内容。

XML 文档的查询定位链接的内容有了很大的变化,其中 XPath 有了很大的发展,而 XLink 的内容没有太大进展,特别是 XPointer 基本没有发展,所以本书删除了 XLink 和 XPointer 的内容,采用了 XPath 3.0,增加了一些新的知识点。另外,还增加了 XQuery 和 XPath Data Model (XDM) 3.0、XPath 和 XQuery Functions 与 Operators 3.0。

从核心技术上讲,AJAX 不是什么新的技术,而是基于 XML 技术的综合应用技术,目前应用已经十分普遍。作为一本专门讲述 XML 技术群的书,引入 AJAX 技术是应该的,也是必须的。但限于篇幅,本书没有太多地讨论 AJAX 的应用问题。

本书共分 11 章。第 1 章概述 XML 的基础知识、编辑工具、相关技术及其发展,以及 Ajax 技术。第 2 章简要介绍 HTML 4.0 的制表、表单、框架、超级链接、字符控制与多媒体等,为后续的章节提供基础。第 3 章讲述 XML 1.0 规范的基本技术要求和规定,对 XML 的元素、属性、实体等重要概念进行描述。第 4 章讲述 XML 文档的元素、元素类型及其定义、XML 1.0 提供的内容模型定义等。第 5 章讲述 XML Schema 2.0,包括定义完整的 XML 数据类型,XML Schema 结构,设计 XML Schema 文档等内容,是



XML 数据库的基础部分。第 6 章讲述用于 XML 文档的查询规范 XPath 3.0、XPath 和 XQuery Functions 与 Operators 3.0。第 7 章讲述专门用于 XML 转换的技术 XSLT 2.0, 包括样式表、模板规则、样式表设计技术和实现等。第 8 章讲述 XML DOM Level 3, 包括 DOM 内核, 以及 XML DOM 应用和编程。第 9 章讲述用 ASP、ADO 实现 XML 与传统关系数据库连接, XML 数据库技术简介。第 10 章介绍 AJAX 编程基础, 讨论 JavaScript 的数据类型, 程序控制(分支, 循环), 函数等基础, 讨论 XMLHttpRequest 对象, 如何建立 Web 连接, 通过 XMLHttpRequest 访问 XML 文档。第 11 章给出一个简单实例。

为了方便读者学习, 每一章都附有习题, 通过习题和实验, 可以帮助读者加深理解 XML 技术及其应用的各个方面。本书的实例全部通过上机调试。

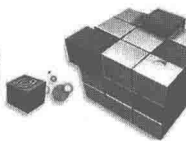
本书由王震江编著。参与本书研究工作和资料整理的人员有彭嘉凤、方刚、俞锐刚、马宏、王玉见、欧晓明、李燕, 在此向他们表示感谢。

由于编者水平有限, 书中难免存在疏漏和不足之处, 衷心希望广大读者给予批评指正。

编 者

2015 年 6 月

◀ 前言 (第一版)



随着 Web 技术和 Internet 的广泛应用, HTML 的可扩展性、结构化和灵活性已经不能满足应用的需要, 其局限性已经影响到 Internet 应用的发展。1998 年 2 月, W3C 提出了 XML 技术的第一个规范 XML 1.0, 目标是创建一种标记语言, 并同时具备定义严格、语法明确、表示方便、结构良好、适用于所有行业的新的标记定义等特点, 用来彻底解决在 Internet 应用中存在的问题。XML 技术源自 SGML, 它既具备 SGML 的核心特征, 又有 HTML 的简单性。目前, XML 技术已经开始在 Web、新型数据库系统中被广泛应用, 在未来的计算机网络应用、网络编程、跨平台编程中将发挥越来越重要的作用。

本书共分 10 章。第 1 章概述 XML 的基础知识、编辑工具、相关技术及发展。第 2 章讲述 XML 1.0 规范的基本技术要求和规定, 对 XML 的元素、属性、实体等重要概念进行介绍。第 3 章讲述 XML 文档的元素、元素类型及其定义、XML 1.0 提供的内容模型定义等。第 4 章讲述 XML Schema 2.0, 包括定义完整的 XML 数据类型、XML Schema 结构、设计 XML Schema 文档等内容, 是 XML 数据库的基础部分。第 5 章讲述 CSS 2.0 技术规范和要求, 用 CSS 转换 XML 文档的方法和技巧。第 6 章讲述专门用于 XML 转换的技术 XSL 2.0, 包括样式表、模板规则、样式表设计技术和实现等。第 7 章讲述用于 XML 文档的查询链接和检索的规范 XPath、XLink、XPointer。第 8 章讲述 XML DOM Level 3, 包括 DOM 内核, 以及 XML DOM 应用和编程。第 9 章讲述用 ASP 和 ADO 实现 XML 与数据库的连接、SQL Server 2000 提供的 XML 技术支持、XML 数据转换成 SQL Server 2000 数据库的相关技术以及介绍 XML 数据库技术。第 10 章讨论建立 Web 连接和用 ASP 技术访问 XML 文档, 并在最后给出一个简单的应用实例。

为了方便读者学习, 每一章都附有习题, 通过练习和实验, 可以帮助读者理解 XML 技术及其应用的各个方面内容。

本书由王震江编著。参与本书研究工作和资料整理的人员有彭嘉凤、王武、梁振碧、方刚、俞锐刚、马宏、王玉见、欧晓明、李燕。在本书的编写工作中, 得到了彭嘉凤同志的大力支持, 在此表示感谢。由于编者水平有限, 书中难免存在疏漏和不足之处, 衷心希望广大读者给予批评指正。

在此, 对大力支持本书出版的中国铁道出版社的各位编辑表示感谢!

编 者

2006 年 3 月



目 录



第1章 概述.....1	
1.1 XML 技术简介.....1	
1.1.1 XML 的历史.....1	
1.1.2 XML 与 HTML 的 比较.....3	
1.1.3 XML 的编辑工具.....4	
1.1.4 浏览 XML 文档.....6	
1.1.5 XML 的相关技术 及其发展.....6	
1.2 AJAX 技术简介.....11	
1.2.1 什么是 AJAX.....11	
1.2.2 AJAX 运行模式.....12	
1.2.3 AJAX 所使用的技术...13	
1.2.4 AJAX 的用处.....14	
习题 1.....14	
第2章 HTML 技术.....15	
2.1 概述.....15	
2.2 提出问题.....15	
2.3 HTML 语法.....17	
2.3.1 HTML 语法基础.....17	
2.3.2 简单网页设计.....18	
2.4 HTML 的制表.....27	
2.4.1 简单表格.....27	
2.4.2 制表元素.....28	
2.4.3 表格的高级技巧.....30	
2.5 表单.....36	
2.5.1 表单的定义.....36	
2.5.2 表单中常用的元素...38	
2.6 框架和超链接.....41	
2.6.1 框架的概念.....41	
2.6.2 框架的元素.....43	
2.6.3 超链接.....45	
2.7 字符控制.....47	

2.7.1 字体.....47	
2.7.2 列表控制.....48	
2.8 多媒体.....50	
2.8.1 多媒体技术.....50	
2.8.2 文本的滚动.....51	
2.9 XHTML 简介.....52	
2.9.1 关于元素的规定.....53	
2.9.2 关于属性.....54	
习题 2.....54	
第3章 XML 语法.....55	
3.1 XML 语法基础.....55	
3.1.1 结构良好性.....56	
3.1.2 XML 声明.....56	
3.1.3 XML 元素.....58	
3.1.4 XML 元素的属性....60	
3.1.5 标记与字符数据.....61	
3.1.6 实体.....62	
3.1.7 处理指令.....67	
3.1.8 CDATA 节.....67	
3.1.9 注释.....69	
3.2 XML 文档结构.....70	
3.2.1 XML 序言.....70	
3.2.2 XML 数据.....71	
3.3 名称空间.....73	
3.3.1 名称空间的概念.....73	
3.3.2 名称空间的声明.....74	
3.4 语言标识.....76	
3.4.1 国家代码和语言代码...76	
3.4.2 覆盖.....78	
习题 3.....79	
第4章 文档类型定义.....80	
4.1 概述.....80	
4.2 元素.....81	

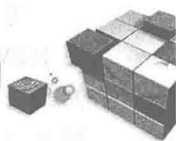


4.2.1 元素声明	81	6.2.1 XPath 应用举例	168
4.2.2 元素类型	84	6.2.2 XPath 数据类型	170
4.2.3 元素取值	85	6.2.3 表达式上下文	171
4.3 属性	86	6.2.4 路径表达式	173
4.3.1 基本概念	86	6.2.5 序列表达式	178
4.3.2 属性的声明及类型 ...	87	6.2.6 过滤表达式	178
4.3.3 条件节	98	6.3 XPath Data Model	178
4.4 内部和外部 DTD	100	6.3.1 数据模型构建	179
4.4.1 内部 DTD	100	6.3.2 访问程序	179
4.4.2 外部 DTD	101	6.3.3 结点类型	181
习题 4	102	6.4 XPath 和 XQuery Functions 与 Operators	187
第 5 章 XML Schema	103	6.4.1 数值运算和函数 ...	188
5.1 概述	103	6.4.2 字符串函数	191
5.2 一个 XML Schema 文档示例	104	6.4.3 逻辑运算和函数 ...	193
5.3 XML Schema 文档结构	106	6.4.4 日期时间运算 和函数	194
5.3.1 模式的基本概念 ...	109	6.4.5 上下文函数	198
5.3.2 复杂类型和简单类型 ...	110	6.5 应用举例	199
5.3.3 元素声明	116	习题 6	200
5.3.4 属性声明	120	第 7 章 XSL 转换 XML	202
5.3.5 其他	122	7.1 概述	202
5.4 XML Schema 数据类型	128	7.2 一个 XSL 文档的讨论	203
5.4.1 数据类型体系	128	7.3 XSL 样式表	206
5.4.2 数据类型二分法 ...	134	7.3.1 样式表元素	206
5.4.3 内置数据类型	136	7.3.2 样式表名称空间 ...	209
5.5 XML Schema 名称空间	151	7.3.3 样式表模块组合 ...	210
5.5.1 不同的名称空间 ...	152	7.4 模板规则	211
5.5.2 名称空间表示	152	7.4.1 模板	211
5.5.3 在 Web 上定位模式 文件	153	7.4.2 模式	220
5.5.4 一致性	153	7.4.3 序列构造器	221
5.6 XML Schema 设计	154	7.4.4 条件	224
5.6.1 定义局部和全局成分 ...	154	7.4.5 循环	226
5.6.2 定义名称空间	158	7.5 样式表设计	228
5.6.3 模式组装	160	7.5.1 排序	230
习题 5	164	7.5.2 求和	230
第 6 章 XPath	166	7.5.3 彩色效果	233
6.1 概述	166	7.5.4 自动编号	237
6.2 XPath 基础	166	7.5.5 创建元素和属性 ...	242

7.5.6 变量和参数使用 ...	243
习题 7	249
第 8 章 XML DOM 技术	250
8.1 概述	250
8.2 一个 DOM 示例	250
8.3 DOM 基础	252
8.3.1 DOM 结构模型	252
8.3.2 核心模块基础	254
8.4 DOM 常用接口	256
8.4.1 Document 接口	256
8.4.2 Node 接口	262
8.4.3 Element 接口	263
8.4.4 Attr 接口	264
8.4.5 Text 接口	265
8.4.6 Comment 接口	266
8.4.7 DocumentFragment 接口	266
8.4.8 DOMImplementation 接口	266
8.4.9 NodeList 接口	267
8.4.10 NamedNodeMap 接口	267
8.4.11 CharacterData 接口	268
8.4.12 NameList 接口	268
8.5 XML DOM 设计	268
8.5.1 创建和浏览 DOM 对象	268
8.5.2 修改 DOM 树结构 ...	275
8.5.3 实现 XSLT 转换 ...	280
8.5.4 出错判断	281
8.6 浏览 DOM 树	282
习题 8	287
第 9 章 XML 数据库技术	288
9.1 概述	288
9.2 数据岛	288
9.2.1 基本概念	288
9.2.2 简单 XML 文档的 处理	288
9.2.3 多级 XML 文档的 处理	291
9.3 XML 数据源对象	294
9.3.1 基本概念	294
9.3.2 处理数据源对象 ...	294
9.4 数据集操作	296
9.4.1 数据集的 概念与操作	296
9.4.2 移动指针	298
9.4.3 ASP 对象	301
9.4.4 ADO	306
9.4.5 连接数据库	307
9.4.6 通过 ADO 实现转换	308
9.5 XML 数据库	311
9.5.1 XML 数据库概念 ...	311
9.5.2 访问 XML 数据库 ...	311
习题 9	312
第 10 章 AJAX 编程基础	313
10.1 概述	313
10.2 JavaScript 基础	313
10.2.1 构建 JavaScript 编程环境	313
10.2.2 入门示例	314
10.2.3 JavaScript 的 交互方法	316
10.2.4 变量、数据类型和 表达式	318
10.3 控制语句	321
10.3.1 条件控制	321
10.3.2 循环	323
10.4 函数与对象	324
10.4.1 函数定义	324
10.4.2 函数调用和 参数传递	324
10.4.3 对象	326
10.5 其他	329
10.5.1 错误捕获和处理 ...	329
10.5.2 内部 JavaScript 和 外部 JavaScript ...	331



10.5.3 注释.....	331	11.2 XML 应用举例.....	348
10.6 XMLHttpRequest 对象.....	332	11.2.1 XML 数据.....	348
10.6.1 操作		11.2.2 界面设计.....	348
XMLHttpRequest		11.2.3 数据输入.....	351
对象.....	333	11.2.4 数据查询.....	353
10.6.2 XMLHttpRequest		11.2.5 数据修改.....	356
请求检测.....	340	11.2.6 数据删除.....	359
习题 10.....	343	习题 11.....	361
第 11 章 XML 与 Web 应用.....	344	参考文献.....	362
11.1 通过 ASP 访问 XML.....	344		



1.1 XML 技术简介

XML 技术是在 Internet 广泛应用之后, 传统的 Web 技术 HTML 的可扩展性、结构化和灵活性已经不能满足应用需要, 且已经影响到 Internet 应用发展的背景下提出的。1998 年 2 月, W3C (World Wide Web Consortium) 提出了 XML 技术的第一个规范 XML 1.0, 目标是创建一种标记语言, 并同时具备定义严格、语法明确、表示方便、结构良好、适用于所有行业的新标记定义等, 用来彻底解决 Internet 应用中存在的问题。XML 技术源自 SGML, 它既具备 SGML 的核心特征, 又有 HTML 的简单性。目前, XML 技术已经开始在 Web、新型数据库系统中广泛应用, 在计算机网络应用、网络编程、跨平台编程、移动互联网、物联网技术中发挥越来越重要的作用。

1.1.1 XML 的历史

1. SGML

1979 年国际标准化组织 (ISO) 设立了一个文本处理小组, 开始开发一种基于 GML (Generalized Markup Language) 的文本描述语言。1980 年, 第一个 GML 工作草案发布。1983 年, 这个文本描述语言最终演化成 SGML (Standard GML), 这是一个标准化的信息结构化技术, 后来 SGML 扩展和修改成为一种全面适应工业范围的信息标准。1986 年, ISO 采纳了 SGML。

SGML 语言庞大, 功能强, 体系严密, 同时技术比较复杂, 价格昂贵, 需要大量的软件来支持它, 导致运营成本较高。20 世纪 80 年代主要用于电子产品交易、科技文献分类等方面。

2. HTML

HTML (HyperText Markup Language, 超文本标记语言) 于 1989 年在欧洲核子物理实验室问世, 这个技术采用超文本传输协议 (HyperText Transfer Protocol)。

HTML 的出现对 Internet 的爆炸性发展起到了积极的作用, WWW 成为人类了解信息、了解世界的一种全新的概念和模式。HTML 的巨大成功, 使 HTML 迅速从 1.0 版本发展成 4.0 版本。在发展的过程中, 给 HTML 赋予了比最初设想要复杂得多的功能, 目的是使 HTML 完成来自于商业应用、科学研究、信息发布的所有任务, 使得 HTML 的语言失去了最初的简单性, 并且在使用 HTML 时出现了随意性、不规范和不



严格等问题。

另一方面,HTML 的专用词表有限,用户无法自由增加新标记并进行有效性验证,用 HTML 来完成不同行业内的数据定义、数据表示,以及行业之间的数据交换很不方便,而随着 Internet 的广泛应用,这种交换又是必须的、大量的、十分广泛的,这使得 HTML 捉襟见肘,无法满足这种要求。

3. XML

为了解决 HTML 在 Internet 应用中的局限性,1996 年,W3C 开始寻找在 Web 中使用 SGML 的方法。因为,SGML 具备 HTML 所没有的三种优势:可扩展性、结构化和灵活性。他们的目标是创建一种标记语言,这种语言既要具备 SGML 的核心特征,又要有 HTML 的简单性。同时具备许多新的特征,如定义严格、语法明确、表示方便、结构良好、适用于所有行业的新的标记定义等。1998 年 2 月 10 日,W3C 发布了 XML 1.0 规范,这就是 XML (Extensible Markup Language),可扩展标记语言。

XML 是 Web 发展到一定阶段的必然产物。W3C 在 XML 1.0 规范中是这样定义 XML 的:“可扩展标记语言(XML)是用来描述一种称为 XML 的文件的数据对象,同时也部分地描述了处理这些数据对象的计算机程序的行为。XML 是 SGML 应用的一个子集,或为 SGML 的某种限制形式。根据指定规格的定义,XML 文件是符合规范的 SGML 文件。”

根据 XML 1.0 (第 5 版)规范^①,XML 的设计目标是:

- ① XML 应能直接用于 Internet。
- ② XML 应支持广泛的应用。
- ③ XML 应与 SGML 兼容。
- ④ 处理 XML 文档的程序应该容易编写。
- ⑤ XML 的可选择性特征保持绝对小,理想情况下为零。
- ⑥ XML 文档应该是人易读的且合理清晰的。
- ⑦ XML 设计应该是可很快准备好的。
- ⑧ XML 文档的设计应该正规和简洁。
- ⑨ XML 文档应容易创建。
- ⑩ 在 XML 标记中简洁性可以忽略不计。

XML 1.0 规范包含三个主要部分,它们是 XML 文档内容、文档的逻辑结构、文档的物理结构。

XML 文档是由被称为实体的存储单元组成的。实体或者是解析的,或者是不可解析的。解析的数据由字符组成,其中一部分形成字符数据,一部分形成标记。标记对描述 XML 文档的存储布局 and 逻辑结构进行编码。XML 提供了把各种约束强加在该存储布局 and 逻辑结构上的机制。

XML 文档内容中,包括结构良好性、字符集、通用语法结构、字符数据与标记、注释、处理指令、CDATA 节、序言及文档类型定义、文档的独立性声明、空白符处理、行结束控制和语言标识等十余项内容,对书写 XML 文档的数据给出了详细的规定。

^① Tim Bray, et al. Extensible Markup Language (XML) 1.0 (Fifth Edition) [EB/OL]. <http://www.w3.org/TR/2008/REC-xml-20081126/>.

在 XML 文档的逻辑结构部分，定义了 XML 文档的标记书写的规则、元素类型声明、属性表声明和条件节等内容，从而规定了 XML 元素和属性的定义规范。

在 XML 文档的物理结构中，对字符和实体参考、实体声明、实体解析、XML 处理器对实体和参考的处理、构造内部实体替代文本、预定义实体、表示法声明和文档实体等内容，对 XML 文档中实体的定义、引用、内部和外部实体进行了统一规定。

1.1.2 XML 与 HTML 的比较

对于 XML 的语法，将在后续章节中详细讨论。为了便于理解 XML 的语法、标记、文件格式，对 HTML 与 XML 进行简单的比较讨论是一件有意义的事情，可以帮助我们理解 XML。

1. HTML 文档

HTML 文件是一系列用 “<” 和 “>” 符号，把具有特定含义的英文字符串括起来，构造成称为标记的元素来描述语法的。

例 1.1 一个简单的网页文件。

```
<html>
  <head><title>我的第一个网页</title></head>
  <body bgcolor="#c0c0c0">
    <h1 align="center">学习 HTML，设计自己的网页！</h1>
    <p align="center">这是用 HTML 语言编写的第一个主页。</p>
    <p align="center">
      
    </p>
  </body>
</html>
```

在 HTML 文件中出现的标记都是由 HTML 规范规定的，如<html>、<head>、<body>、<p>、<h1>等，每个标记应该有一个结束标记，结束标记由 “</” 和 “>” 包括标记名构成，如</body>、</p>。一个 HTML 文件必须在第一行写上<html>，最后一行写上</html>来包含所有的内容。浏览器判断一个文件是否为 HTML 文件，就是根据文件是否包含在<html>和</html>这一对标记中。但是也有几个元素没有结束标记，如、
、<hr>等。目前使用的是 W3C 在 1998 年 4 月 24 日推出的 HTML 4.0 规范。

2. XML 文档

一个 XML 文档是由用户自行定义的标记组成的，这些标记与 HMTL 一样，都使用 “<>” 来包含标记名。下面来考查一组数据的表示，如表 1.1 所示。

表 1.1 学生成绩表

学 号	姓 名	性 别	专 业	成 绩		
				高等数学	程序设计	电路基础
200811010201	于丹	女	软件工程	89	73	92
...	...	...	...	...	...	...



例 1.2 根据表 1.1 建立 XML 文档。

```
<?xml version="1.0" encoding="GB2312"?>
<studentlist>
<student id="200811010201">
  <name>于丹</name>
  <sex>女</sex>
  <major>软件工程</major>
  <score>
    <mathematics>89</mathematics>
    <programming>73</ programming>
    <circuit>92</circuit>
  </score>
</student>
<!--more students information here-->
</studentlist>
```

与 HTML 文件一个最重要的差别是,在每一个 XML 文档的第一行必须写上 `<?xml version="1.0"?>` 命令,用于标识该文件是一个 XML 文档,也方便 XML 处理器识别哪些是 XML 文档,哪些不是 XML 文档。接下来的内容是用户自定义的根元素标记,根元素必须有一个结束标记,所有用户定义的其他元素都被包含在根元素中,如 `<studentlist></studentlist>` 之间的内容是其他元素标记的定义。每一个标记必须有一个结束标记与之对应。一对标记之间包含的对象称为一个元素。标记的表示与 HTML 的标记表示一样,都是把标记名放在“<”和“>”中间构成,结束标记是把标记名放在“</”和“>”之间。如 `<name></name>`、`<sex></sex>` 等。但是,XML 的标记与 HTML 的标记名不一样。XML 的标记名定义的含义、合法性和有效性,由用户专门在 DTD (文档类型定义) 或者 XML Schema (XML 模式) 中定义。而 HTML 的标记名由 HTML 规范确定。

一般说来,在设计一个标记名时,应该使标记名具有特定的含义,以表示元素的意义。虽然,XML 规范没有对哪种语言可以用作标记名做特别的限定,但为了使用方便,建议使用英文字符作为标记名,最好使用与元素名具有相同或相近含义的英文单词作为标记名,如学生名单 `<studentlist>`、学生 `<student>`、姓名 `<name>`、性别 `<sex>` 等。另外,使用中文字符作为元素名称,也是可行的,如 `<成绩>`、`<高等数学>`、`<程序设计>`、`<街道>` 等。

在例 1.2 中,有一个根元素 `<studentlist>`,根元素下有若干个 `<student>` 元素,每个 `<student>` 元素下都有 `<name>`、`<sex>`、`<major>` 和 `<score>`,`<score>` 标记中还嵌套了一级子元素 `<mathematics>`、`<programming>`、`<circuit>`,这样的元素嵌套,根据需要可以有許多层次。

1.1.3 XML 的编辑工具

XML 文档是文本文件,任何一种纯文本文件编辑工具都可以用于 XML 文档的编辑。如记事本、XML Writer、XML SPY 等。下面是 XML 文档编辑器的简介。

1. 记事本

Windows 中的记事本,是一个文本文件编辑器,也是 XML 文档最常见的编辑工



具。当读者手中一时没有更好的编辑工具时，可以使用记事本来编辑 XML 文档。Windows 98 的记事本没有引入处理 Unicode 字符集的功能。Windows 2000 及其以后版本的记事本提供了处理 Unicode 字符集的功能。

2. XML Notepad

XML Notepad 是微软公司为 XML 文档编辑特意设计的编辑工具。其界面如图 1.1 所示，这是一个简单的 XML 文档编辑器。



图 1.1 XML Notepad 界面

3. Amaya

Amaya 是 W3C 专门为 XML 设计的开发工具，这是一个比较全面完整的 XML 开发工具，其界面如图 1.2 所示。读者可以在 <http://www.w3.org/> 中下载该软件。

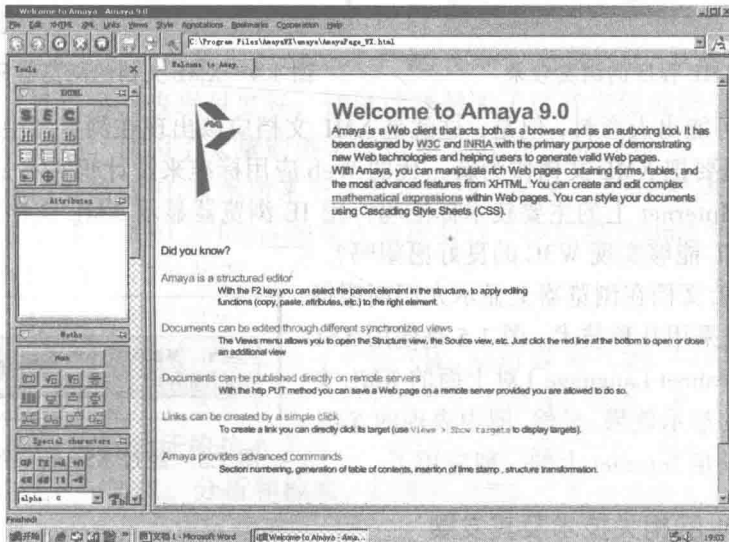


图 1.2 Amaya 界面



4. XML SPY

XML SPY 是 Altova 公司开发的专用 XML 编辑管理工具, 一共有三个版本: XMLSpy Enterprise Edition/Professional Edition/Home Edition。用户可以获得三个月免费使用的试用版, 读者可以在 <http://www.altova.com/> 中下载。

1.1.4 浏览 XML 文档

作为 Internet 上的主要浏览器, 一个是微软的 Internet Explorer, 另一个是网景公司的 Navigator (该公司已破产)。微软的浏览器 IE 5.5 及其以后的版本均支持 XML。支持 XML 的还有 W3C 推荐的 Awaya 和 Opera。

HTML 文件在浏览器上显示, 这就是众所周知的网页。HTML 文件在 IE 浏览器上显示效果如图 1.3 所示, 这是一个包含了标题、段落和图形的简单网页。

对于 XML 文档, 在 IE 浏览器上直接显示的效果如图 1.4 所示。从图 1.4 可看出, XML 文档在 IE 浏览器上显示出元素的树状结构, 这类似于微软 Windows 资源管理器中的文件夹的形态。每个元素前的“-”表示该元素已经展开, 若某个元素包含下一级子元素, 展开后将显示该元素的下级元素, 若某个元素前出现“+”, 表示这个元素及其下一级子元素被封闭起来。



图 1.3 HTML 程序的浏览效果



图 1.4 XML 文档的浏览器显示效果

这个效果可能出人意料, 但是, 这就是 XML 文档应该出现在浏览器中的样子。读者不禁要问: 不是要把 XML 文档作为新一代的 Web 应用标准来设计吗? 不是要让 XML 代替 HTML 成为 Internet 上的主要技术标准吗? 用 IE 浏览器显示 XML 文档怎么是这个样子? 这样的 XML 能够实现 W3C 的良好愿望吗?

为了使 XML 文档在浏览器上显示人们期望得到的效果, 可以采用几种技术。图 1.5 是使用 XSL (eXtensible Stylesheet Language) 对上面的 XML 文档进行转换后的显示效果。显然, 图中表现的 XML 文档已经可以满足 Internet 上的一般应用了。

学生成绩表					
学号	姓名	性别	专业	成绩	
				高等数学	程序设计 电路基础
20081101201	于丹	女	软件工程	89	73 92

图 1.5 经过 XSL 转换的文档的显示

1.1.5 XML 的相关技术及其发展

从 1998 年到现在, XML 技术规范从 1.0 第一版本发展到 1.0 第五版本。XML 不



仅在 Internet 领域,还在各种行业的数据描述和表示方面展现出良好的行为和品质,尤其在电子商务领域的表现十分出色,因而 XML 得到了广泛应用。现在,在 W3C 和业界的大力支持下,XML 逐渐向广泛的应用领域迅速发展,开始成为 Web 技术中的核心。同时 XML 的相关技术也快速出现,基于 XML 技术核心的软件体系正在逐渐形成。

1. XML 数据类型定义

XML 文档数据的类型和有效性定义,可以保证 XML 文档数据的正确性和完整性。在 XML 技术规范中,通过 DTD 和 XML Schema 来定义数据类型。

(1) DTD

DTD (Documents Type Definition, 文档类型定义)是 W3C 建议的 XML 有效性检验的标准,其使用先定义根元素,然后分别定义各级子元素以及元素属性的方式对 XML 元素 (Element) 和属性表 (Attlist) 进行定义,提供十分有限的元素类型来定义元素和属性。

(2) XML Schema

XML Schema 最开始是微软公司推荐并且在微软产品中通用的 XML 文档有效性标准,在 2001 年 5 月成为 W3C 的建议规范,W3C 一直在致力于适合描述 XML 内容模型的规范的开发工作。2001 年 5 月 3 日,W3C 发布了 XML Schema 1.0 规范的推荐版本,并声明这个规范是稳定的,有助于 Web 互操作性,并且被 W3C 的成员修订,这些成员来自学术界、业界和研究团体,他们赞成采用这个规范。XML Schema 1.0 定义了公用标记词表、使用这些词表的 XML 文档的结构,以及提供了与这些词表和结构相关的联系。W3C 认为:“XML Schema 1.0 为 XML 发挥其全部潜力提供了一个实质构件”。在开发过程中,参与标准制定的各方都一致认为 XML Schema 1.0 标准的制定将是 XML 发展历史上意义重大、影响深远的事件。

XML Schema 1.0 包含三个模块:结构、数据类型和初级读物 (Primer)。结构描述了 XML 文档的结构和内容模型的约束和规则来操纵文档元素的有效性;数据类型为确定 XML 的元素和属性的数据类型提供了 40 多种,允许 XML 软件在操作数据、数字、时间、串等信息时表现得更好;初级读物是为了读者方便理解 XML Schema 而写的一个非标准文档,使用了大量的例子来描述 XML Schema 1.0。

XML Schema 1.0 给 XML 技术带来更大的灵活性,加速了 XML 技术在重要领域的应用,因此,也加速了 XML Schema 的深入发展和进一步成熟。2004 年 10 月 28 日,W3C 发布了 XML Schema 2.0 规范。

(3) XML 文档的有效性验证

微软对 XML 的技术支持主要反映在它的 XML 解析器 MSXML.dll 中。这个解析器伴随着 XML 的发展而发展,从最初的 2.0 发展到今天的 4.0 版本,在 MS XML 4.0 中,引入对 XML Schema 进行验证的技术支持——SOM (Schema Object Model)。该对象对 XML Schema 文档进行加载、分析和检验,以确定 XML 元素定义的有效性。

2. XML 文档数据显示和转换技术

XML 文档直接在 Web 浏览器上显示时的形态,上面已经讨论过,这一点与 HTML 文档的显示不一样。为了解决 XML 文档的显示问题,需要借助其他技术,下面对这