



高等院校计算机应用技术规划教材

XML程序设计

王震江 编著

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

高等院校计算机应用技术规划教材

XML 程序设计

王震江 编著

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

内 容 简 介

今天, XML 已经在 Web 编程、新型数据库系统、计算机网络应用编程、网络数据交换和跨平台编程中发挥越来越重要的作用。XML 正在成为未来电子商务、电子政务运营和管理的核心技术。

本书全面地讨论了 XML 技术的几个主要方面, 包括 5 个部分: 第一部分是 XML 基础知识和 XML 1.0 规范; 第二部分是 XML 数据的有效性定义, 包括 DTD 技术、XML Schema 2.0; 第三部分是 XML 的支持技术, 包含 CSS 技术、XSL 技术、XPath、XPointer、XLink、XML DOM; 第四部分是 XML 数据库的相关问题, 包括与传统数据库的连接及 XML 数据库概念; 第五部分是 XML 与 Web 应用问题, 包括 Web 与 XML 如何连接和调用, 并在最后给出一个实例, 以解决基于 XML 的 Web 应用。

本书体系完整, 内容新颖, 叙述简单明了, 大量的实例有助于读者对 XML 技术的理解和学习。本书可作为本科院校和高等职业技术学院计算机专业网络应用技术课程的参考书, 也可作为电子商务专业的基础课教材, 还可供相关技术培训班使用。

图书在版编目 (CIP) 数据

XML 程序设计/王震江编著. —北京: 中国铁道出版社, 2006. 4

高等院校计算机应用技术规划教材

ISBN 7-113-07019-1

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

中国版本图书馆 CIP 数据核字 (2006) 第 033090 号

书 名: XML 程序设计

作 者: 王震江

出版发行: 中国铁道出版社 (100054, 北京市宣武区右安门西街 8 号)

策划编辑: 严晓舟 秦绪好

责任编辑: 苏 茜 谢立和 王 丹

封面设计: 薛 为

封面制作: 白 雪

责任校对: 刘 洁

印 刷: 北京市兴顺印刷厂

开 本: 787×1092 1/16 印张: 18.75 字数: 453 千

版 本: 2006 年 4 月第 1 版 2006 年 4 月第 1 次印刷

印 数: 1~5 000 册

书 号: ISBN 7-113-07019-1/TP·1779

定 价: 25.00 元

版权所有 侵权必究

凡购买铁道版的图书, 如有缺页、倒页、脱页者, 请与本社计算机图书批销部调换。

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。

本书全文由王震江编著、统稿、审核。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, 错误和疏漏之处在所难免, 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

第 1 章 XML 概述	1
1.1 XML 的历史简介	1
1.1.1 SGML	1
1.1.2 HTML	1
1.1.3 XML	1
1.2 XML 与 HTML 的比较	2
1.2.1 HTML 文档	2
1.2.2 XML 文档	3
1.3 XML 的编辑工具	4
1.4 浏览 XML 文档	5
1.5 XML 的相关技术	7
1.5.1 XML 数据类型定义	7
1.5.2 XML 文档数据显示和转换技术	8
1.6 XML 的发展	10
1.6.1 XML 的数据库技术	10
1.6.2 XML 的安全技术	10
1.6.3 XML 转换实例	11
习题一	14
第 2 章 XML 的语法基础	15
2.1 XML 的语法基础	15
2.1.1 结构良好性	15
2.1.2 XML 声明	16
2.1.3 XML 的元素	17
2.1.4 属性	19
2.1.5 实体	20
2.1.6 处理指令	24
2.1.7 CDATA 节	24
2.1.8 注释	25
2.2 XML 的文档结构	26
2.2.1 XML 的序言	26
2.2.2 XML 的数据	27
2.3 名称空间	29
2.3.1 名称空间的概念	29

2.3.2 声明名称空间	30
2.4 语言标识	32
2.4.1 国家/地区代码和语言代码	32
2.4.2 覆盖	34
习题二	35
第 3 章 文档类型定义	36
3.1 概述	36
3.2 元素声明	36
3.2.1 元素声明	36
3.2.2 元素类型	39
3.3 属性定义	41
3.3.1 基本概念	41
3.3.2 属性	42
3.3.3 条件节	50
3.4 内部和外部 DTD	52
3.4.1 内部 DTD	52
3.4.2 外部 DTD	53
习题三	54
第 4 章 XML Schema	55
4.1 概述	55
4.2 一个 XML Schema 文档示例	56
4.3 XML Schema 文档的结构	57
4.3.1 模式的基本概念	59
4.3.2 复杂类型和简单类型	60
4.3.3 元素声明	65
4.3.4 属性声明	68
4.3.5 其他	70
4.4 XML Schema 的数据类型	75
4.4.1 数据类型体系	75
4.4.2 数据类型二分法	80
4.4.3 内置数据类型	82
4.5 XML Schema 的名称空间	96
4.5.1 不同的名称空间	96
4.5.2 名称空间表示	97
4.5.3 在 Web 上定位模式文件	97
4.5.4 一致性	97

4.6 XML Schema 设计	98
4.6.1 定义局部和全局成分	98
4.6.2 定义名称空间	101
4.6.3 模式组装	104
习题四	107
第 5 章 CSS 转换 XML	109
5.1 概述	109
5.2 问题的引入	109
5.3 显示属性	110
5.3.1 可视区和包含块	110
5.3.2 显示属性及其值	111
5.3.3 表格属性	115
5.3.4 none 值	116
5.3.5 程序优化	116
5.4 颜色与背景	116
5.4.1 颜色	117
5.4.2 背景	119
5.5 字体与文本	120
5.5.1 尺寸设置	120
5.5.2 字体属性	122
5.5.3 文本属性	125
5.6 边界设置	127
5.7 元素筛选	131
5.8 CSS 对 XML 的转换	133
5.8.1 以数据为主的 XML 文档转换	133
5.8.2 以文本为主的 XML 文档转换	138
5.8.3 内部和外部 CSS	141
习题五	142
第 6 章 XSL 转换 XML	143
6.1 概述	143
6.2 一个 XSL 文档的讨论	143
6.3 XSL 样式表	146
6.3.1 样式表元素	146
6.3.2 样式表名称空间	148
6.3.3 样式表模块组合	149
6.4 模版规则	150
6.4.1 模版	150

6.4.2 模式	156
6.4.3 序列构造	158
6.4.4 循环	159
6.4.5 条件处理	161
6.5 样式表设计	162
6.5.1 排序	164
6.5.2 求和	164
6.5.3 彩色效果	166
6.5.4 自动编号	168
6.5.5 创建元素和属性	172
6.5.6 变量和参数使用	173
习题六	177
第 7 章 XML 查询定位链接	179
7.1 概述	179
7.2 XPath	179
7.2.1 什么是 XPath	179
7.2.2 XPath 的数据类型	181
7.2.3 XPath 的表达式	181
7.2.4 表达式的运算	186
7.2.5 函数	188
7.2.6 应用举例	193
7.3 XPointer	195
7.3.1 为什么需要 XPointer	195
7.3.2 XPointer 框架	196
7.3.3 element()	199
7.3.4 xmlns()	200
7.4 XLink	201
7.4.1 基本概念	201
7.4.2 XLink 的属性	202
7.4.3 XLink 的链接	203
习题七	207
第 8 章 XML DOM	208
8.1 概述	208
8.2 一个 DOM 示例	208
8.3 DOM 基础知识	210
8.3.1 DOM 的结构模型	210
8.3.2 核心模块基础知识	211

8.4 DOM 常用接口	213
8.4.1 Document 接口	213
8.4.2 Node 接口	219
8.4.3 Element 接口	221
8.4.4 Attr 接口	222
8.4.5 Text 接口	223
8.4.6 Comment 接口	223
8.4.7 DocumentFragment	224
8.4.8 DOMImplementation 接口	224
8.4.9 NodeList 接口	224
8.4.10 NamedNodeMap	224
8.4.11 CharacterData	225
8.4.12 NameList 接口	225
8.5 XML DOM 设计	226
8.5.1 创建和浏览 DOM 对象	226
8.5.2 修改 DOM 树结构	231
8.5.3 实现 XSLT 转换	233
8.5.4 出错判断	234
8.5.5 浏览 DOM 树	235
习题八	238
第 9 章 XML 与数据库的交换	239
9.1 概述	239
9.2 XML 数据源对象	239
9.2.1 基本概念	239
9.2.2 数据集操作	239
9.2.3 建立数据源	241
9.2.4 移动指针	242
9.2.5 表格转换	244
9.3 XML 与数据库的连接	245
9.3.1 ASP 对象	245
9.3.2 ADO	250
9.3.3 连接数据库	251
9.4 数据库产生 XML 文档	252
9.4.1 通过 ADO 连接数据库	252
9.4.2 SQL Server 2000 的 XML 功能	254
9.5 XML 数据转换为数据库	260
9.5.1 关于转换模型	260

9.5.2 数据类型	261
9.5.3 创建数据库	262
9.6 XML 数据库	264
9.6.1 XML 数据库概念	264
9.6.2 访问 XML 数据库	264
习题九	265
第 10 章 XML 与 Web 应用	266
10.1 建立 Web 连接	266
10.1.1 XML HttpRequest 对象	266
10.1.2 操作 HttpRequest 对象	267
10.1.3 HttpRequest 请求检测	271
10.2 ASP 访问 XML 文档	273
10.2.1 通过 DOM 访问 XML	273
10.2.2 通过 HttpRequest 处理 XML	275
10.3 XML 应用举例	277
10.3.1 XML 数据	277
10.3.2 界面设计	277
10.3.3 数据输入	280
10.3.4 数据查询	282
10.3.5 数据修改	284
10.3.6 数据删除	287
习题十	289
参考文献	290

1.1 XML 的历史简介

1.1.1 SGML

20 世纪 60 年代,人们想把文件结构化成标准的格式,用来促进数据的交换和操作。IBM 公司研究了一种描述文本及其格式的通用标记语言 GML (Generalized Markup Language)。GML 不是使用简单的文档标记方式,而是实现了一个正式的文档类型定义 (DTD) 的使用,这个正式的 DTD 明确地说明了元素标记及其类型,还为它们出现的顺序和它们与其他元素的关系定义了规则,把文档结构与内容本身分离开来(这种方法描述了 SGML 的工作原理)具有十分重要的意义。1980 年底, GML 在出版界得到了广泛的应用。

1979 年美国国家标准化组织设立了一个文本处理小组,开始开发一种基于 GML 的文本描述语言。1980 年,这项工作导致了第一个 SGML 工作草案的出版。1983 年,这个文本描述语言最终演化成 SGML (Standard Generalized Markup Language),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,国际标准化组织 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,它主要用于电子产品交易、科技文献分类等方面。

1.1.2 HTML

HTML (HyperText Markup Language) 为超文本标记语言。1989 年,该语言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。

HTML 的出现给 Internet 的爆炸性发展产生了积极的作用,WWW 成了人类了解信息、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 发展成 4.0。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 完成所有的来自于商业应用、科学研究、信息发布任务,使得 HTML 语言失去了最初的简单性,并且 HTML 的使用上出现随意性、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,用它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,随着 Internet 的广泛应用,这种交换又是必需的、大量的、十分广泛的,这使得 HTML 捉襟见肘,无法满足要求。

1.1.3 XML

为了解决 HTML 在 Internet 应用中的局限性,1996 年,W3C(World Wide Web Consortium)开始寻找在 Web 中使用 SGML 的方法。因为 SGML 具备 HTML 所没有的 3 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（缩写为 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 在应用上的一个子集，或为 SGML 的某种限制形式。根据指定规格的定义，XML 文件是符合规格的 SGML 文件。”

在 XML 1.0 规范中，包含 3 个主要的部分，分别是 XML 文档内容、文档的逻辑结构、文档的物理结构。

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。

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，包括字符和实体参考、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。

1.2 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，把 HTML 与 XML 进行简单比较讨论是一件有意义的事情，可以帮助我们理解 XML。

1.2.1 HTML 文档

HTML 文件是用一系列“<”和“>”符号，把具有特定含义的英文字符串括起来，构造成为标记的元素来描述语法的。一个简单网页的 HTML 程序如下。

【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>等。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。浏览器根据文件是否包含在<html>和</html>这一对标记中判断一个文件是否是 HTML 文件。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。

1.2.2 XML 文档

一个 XML 文档是由用户自行定义的标记组成，这些标记与 HTML 一样，都使用“<>”来包含标记名。下面来考察如表 1-1 所示的一组数据。

表 1-1 学生基本情况表

学 号	姓 名	性 别	电 话	地 址		
				省	市	街 道
2003110101	刘艳	女	0871-3350356	云南	昆明	人民中路 258 号
...	...	...	...	...	...	...

表 1-1 中数据的 XML 文档表示如下。

【例 1.2】根据表 1-1 建立 XML 文档（文件名为 ch1-2.html）。

```
<?xml version="1.0" encoding="GB2312"?>
<studentlist>
  <student id="2003110101">
    <name>刘艳</name>
    <sex>女</sex>
    <phone>0871-3350356</phone>
    <address>
      <province>云南</province>
      <city>昆明</city>
      <street>人民中路 258 号</street>
    </address>
  </student>
  <!--more students information here-->
</studentlist>
```

与 HTML 文件的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，用以标识该文件是一个 XML 文档，也方便 XML 处理器识别哪些是 XML 文档，哪些不是。接下来的内容是用户自己定义的根元素标记，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。XML 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 的标记名与 HTML 的标记名不一样。XML 标记名的定义的含义、合法性和有效性，由用户自己专门在 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做标记名作特别的限定, 但为了使用方便, 建议使用英文字符作为标记名为宜, 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<街道>、<城市>、<商品名>、<生产厂>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>、<phone>和<address>, 在< address>标记中还嵌套了下一级子元素<province>、<city>、<street>, 这样的元素嵌套根据需要可以有許多层次。

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 Notepad、XMLSPY 等。下面是 XML 文档编辑器的一个简介。

1. 记事本

Windows 98/2000/XP/2003 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 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, 界面如图 1-1 所示, 它是一个简单的 XML 文档编辑器。

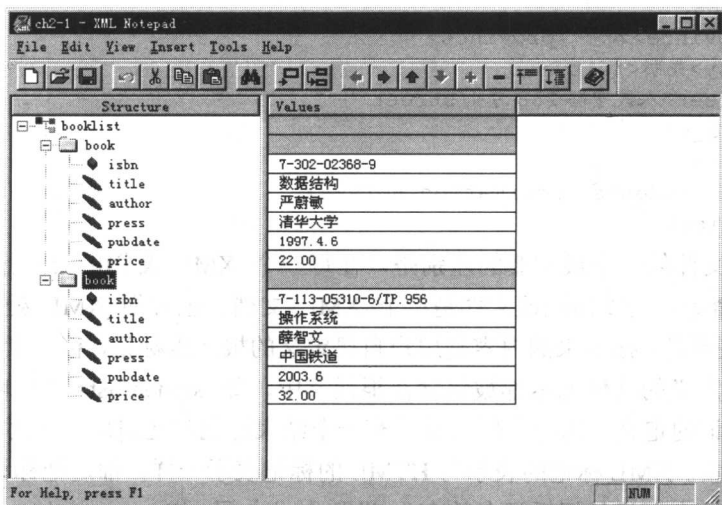


图 1-1 XML Notepad

3. Amaya

Amaya 是 W3C 专门为 XML 设计的开发工具, 这是一个比较全面的 XML 开发工具。Amaya 9.0 的界面如图 1-2 所示, 读者可以在 <http://www.w3.org> 中下载该软件。

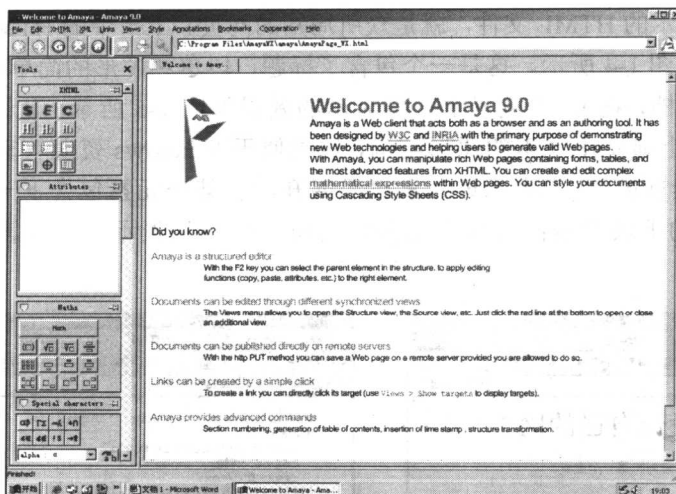


图 1-2 W3C Amaya

4. XML Spy

XML Spy 是 ALTOVA 公司开发设计的专用 XML 编辑管理工具。一共有 3 个版本：XML Spy Enterprise Edition/Professional Edition/ Home Edition。用户可以获得一个 3 个月免费试用版，读者可以在 <http://www.altova.com> 中获得。图 1-3 是 XML Spy 的用户界面。

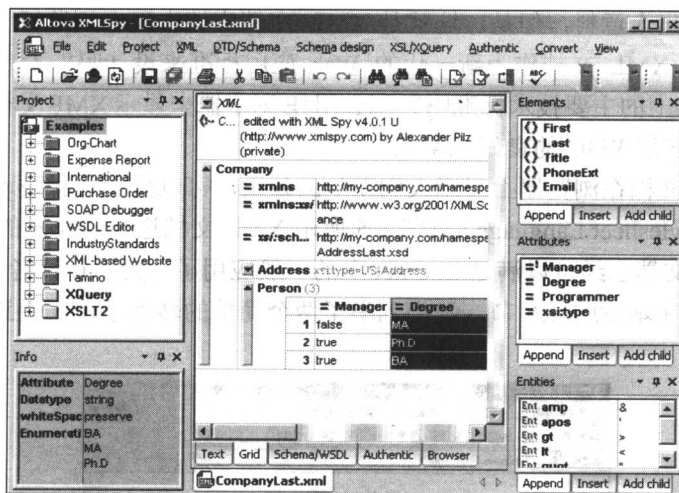


图 1-3 XML Spy 用户界面

1.4 浏览 XML 文档

Internet 上的主要浏览器，一个是微软公司的 Internet Explorer，另一个是网景公司的 Navigator。微软公司的浏览器 IE5.5 及其以后的版本均支持 XML。随着 XML 的发展，网景公司的浏览器也开始支持 XML。

除了上述两个浏览器外，支持 XML 的还有 W3C 自己推荐的 Awaya 和 Opera。

在浏览器上显示的 HTML 文件，就是众所周知的网页。上节中的 HTML 文件在 IE 浏览器上的显示效果如图 1-4 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 1-5 所示。从图 1-5 看出，XML 文档在 IE 浏览器上显示出元素的树形结构，这类似于 Windows 资源管理器左窗格的文件夹形式。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“+”，则表示这个元素及其下一级子元素被封闭起来。

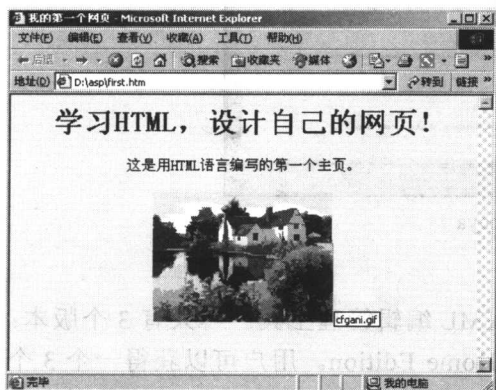


图 1-4 HTML 程序的浏览效果



图 1-5 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上的样子。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？怎么用 IE 浏览器来显示 XML 文档会成这个样子？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的效果，可以采用几种技术。图 1-6 是使用 XSL (eXtensible Stylesheet Language) 对图 1-5 所示 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。为了让读者对 XML 的转换有一个初步的认识，在本章的 1.6 节，我们列举了两个简单的转换示例，更详细的内容请参考本书的后续章节。

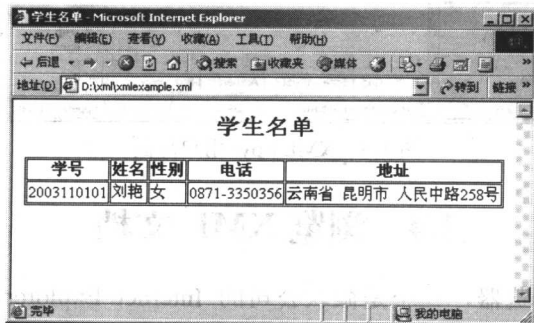


图 1-6 经过 XSL 转换的文档的显示效果

1.5 XML 的相关技术

1998 年至今,短短几年,XML 技术规范从 1.0 第一版本发展到 1.0 第三版本。XML 不仅在 Internet 领域,而且在各种行业数据描述和表示方面都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,并得到了广泛应用。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。

1.5.1 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。DTD 是 W3C 建议的 XML 有效性检验的标准,但是其可以使用的数据类型十分有限。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。2004 年 10 月 28 日,W3C 发布了 XML Schema 2.0 规范。XML Schema 2.0 提供的丰富的数据类型,足以满足 XML 的各类应用。

1. DTD

DTD (Documents Type Definition),即文档类型定义。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 (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

W3C 一直致力于适合描述 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 3 个模块:结构、数据类型和初级读物。结构描述了 XML 文档的结构和内容模型的约束及规则来操纵文档元素的有效性。数据类型为确定 XML 的元素和属性提供了多种数据类型,允许 XML 软件在操作数据、数字、时间、串等信息时表现得更好。初级读物是为了方便读者理解 XML Schema 而写的一个非标准文档,使用大量的例子来描述 XML Schema 1.0。

XML Schema 1.0 给 XML 技术带来更大的灵活性,加速了在重要领域中 XML 技术的应用,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 版本发展到今天的 4.0 版本,在 MSXML 4.0 中,引入