

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

XML 实用教程

XML Programming

范立锋 主编

- 理论讲解通俗易懂
- 案例习题经典实用
- 重点技术全面包括



高校系列

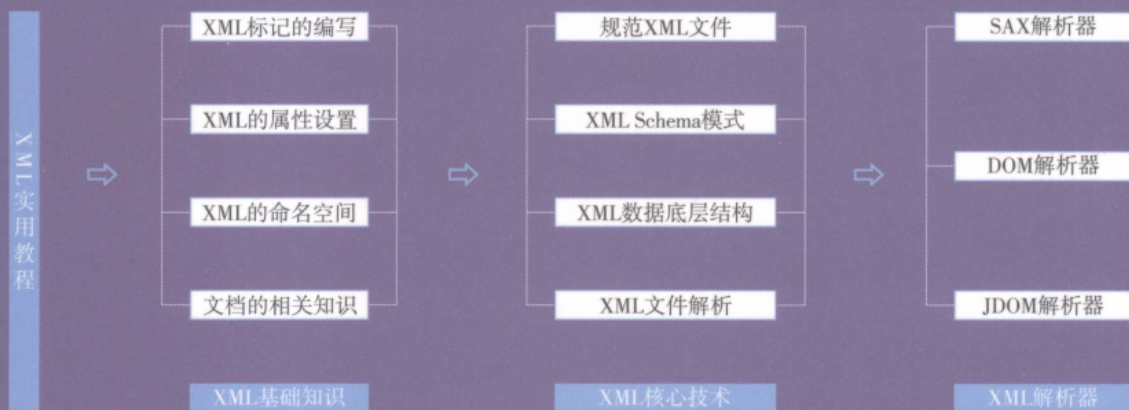
 人民邮电出版社
POSTS & TELECOM PRESS

XML Programming

XML 实用教程

本书针对XML初学者，重点介绍XML的核心技术及XML的相关技术。全书提供了大量的案例习题，与讲解知识点紧密结合；配有多媒体教学课件，方便教学；提供所有例题源代码、习题答案，方便学习；通过本书的学习，读者可以快速掌握XML的相关知识及具体应用。

本教材的结构框图



免费提供

PPT等教学相关资料



人民邮电出版社
教学服务与资源网
www.ptpedu.com.cn

教材服务热线：010-67170985

反馈/投稿/推荐信箱：315@ptpress.com.cn

人民邮电出版社教学服务与资源网：www.ptpedu.com.cn



ISBN 978-7-115-20527-8



9 787115 205278 >

ISBN 978-7-115-20527-8/TP

定价：24.00 元

封面设计：董志桢

人民邮电出版社网址：www.ptpress.com.cn

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

XML 实用教程

XML Programming

范立锋 主编



高校系列

人民邮电出版社

北京

图书在版编目(CIP)数据

XML实用教程 / 范立锋主编. —北京: 人民邮电出版社,
2009. 4

21世纪高等学校计算机规划教材

ISBN 978-7-115-20527-8

I. X… II. 范… 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(2009)第036640号

内 容 提 要

本书主要面向 XML 的初学者, 重点放在 XML 的相关重要概念和技术上, 并配有例题详解, 且所选例题具有一定的针对性。本书内容主要包括 XML 概述、XML 文件的规则、XML 数据的底层结构、XML Schema 模式、使用 CSS 格式化显示 XML 文件、使用 XSL 格式化显示 XML 文件、XML 的数据源对象等, 同时还介绍了两个 XML 解析器, 即 DOM 解析器与 SAX 解析器。

本书可作为高等学校计算机、电子商务以及信息类相关专业的 XML 相关课程的教材, 也可供广大 Web 应用程序开发者和用户学习参考, 也适合专职的 XSLT 程序员使用。

21 世纪高等学校计算机规划教材

XML 实用教程

-
- ◆ 主 编 范立锋
责任编辑 刘 博
 - ◆ 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@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京昌平百善印刷厂印刷
 - ◆ 开本: 787×1092 1/16
印张: 13
字数: 336 千字
印数: 1—3 000 册
 - 2009 年 4 月第 1 版
2009 年 4 月北京第 1 次印刷

ISBN 978-7-115-20527-8/TP

定价: 24.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223

反盗版热线: (010)67171154

前言

可扩展标记语言（XML）是由万维网协会（W3C）推出的新一代数据交换的标准，它主要用于定义 Web 网页上的文档元素和商业文档。对 XML 的研究和应用在业界正逐渐兴起，并在 Internet 上迅速发展。XML 最大的特点是能够结构化地描述数据，实现数据共享。作为标记语言，XML 又是一种元语言，用户可用 XML 来创建自己需要的其他标记语言（如出版社可以用来定义书籍标记语言）。正是由于 XML 的可扩展性，使得 XML 的应用能迅速渗透到各个领域，并使得广大用户对 XML 的学习正在全面展开。

本书是作者在总结了多年开发经验与成果的基础上编写的。本书全面、翔实地介绍了 XML 开发所需的各种知识和技巧。通过对本书的学习，读者可以快速、全面地掌握和使用 XML，并可达到融会贯通、灵活运用之目的。

XML 知识体系

XML 的知识体系如下图所示。



本书特点

本书注重教材的可读性和实用性，多数实例都是经过笔者精心挑选、编制而成的，它既能帮助读者理解知识，又具有启发性。本书在讲解过程中以“实用”为原则，介绍了 XML 节点之间的关系，另外，在介绍基础知识时，以“易学、易懂、易用”为原则，先对知识点作简要介绍，然后通过小实例来演示知识点。同时，本书还专注于解决问题的经典方法与技巧，保证读者看得懂，学得会，并能够快速进入到这个领域来。本书的特点如下。

（1）教材知识体系结构合理。本书知识安排强调整体性和系统性，知识表达强调层次性和有序性，概念和技术逐层推进，便于读者学习和理解。

（2）语言通俗易懂，读者容易理解。书中采用程序结构、页面交互图、表格等多种方式，来描述问题及解决问题的过程，使读者从多个角度来理解问题。

（3）注重理解，把握全局。本书对很多模糊的概念进行了翔实的描述，同时记录了许多笔者的技术经验和心得体会，以帮助读者思考问题和解决问题。而且，本书注重从全局上把握知识框架，在介绍宏观知识的同时穿插介绍 XML 的语法结构。

(4) 注重实际, 应用为主。本书注重实际应用, 尤其在介绍 SAX 和 DOM 解析器时, 都从应用的角度出发进行介绍, 并将应用模块化, 对每个模块都总结了典型实例和应用场合, 既方便了初级读者进行理解, 也方便了中高级读者进行查询。

(5) 分析本质, 前后呼应。XML 中很多技术都是相互呼应的, 本书在介绍时, 着重表达了知识体系的前后呼应。例如, 本书在介绍 XML Schema 中复杂类型时, 首先介绍如何定义复杂类型, 然后介绍复杂类型的默认规则, 以便于读者深入了解。

本书结构

本书介绍了 XML 的核心技术和 XML 的相关技术, 书中实例丰富, 而且每章后附有习题, 便于读者巩固所学内容。全书共分 9 章, 主要内容如下。

章 节	描 述
第 1 章 XML 概述	介绍 XML 的产生, XML 的基础知识, XML 的应用
第 2 章 XML 文件的规则	介绍 XML 的语法规则
第 3 章 XML 数据的底层结构	介绍什么是有效的 XML 文件, 以及 XML 文件的文档类型如何定义文件 DTD
第 4 章 XML Schema 模式	介绍 XML 文件的模式定义 XML Schema, 并与 DTD 作对比
第 5 章 使用 CSS 格式化显示 XML 文件	介绍如何利用 CSS 技术格式化显示 XML 文件, 使 XML 文件不以树状形式显示
第 6 章 使用 XSL 格式化显示 XML 文件	介绍如何利用 XSL 格式化显示 XML 文件, 并与 CSS 技术做对比
第 7 章 XML 的数据源对象	介绍如何将 HTML 中的对象和 XML 文件绑定, 如何利用结果集操纵数据岛, 以及 XML 文件如何和数据库交互
第 8 章 DOM 解析器	介绍如何利用 Java 语言的 DOM 解析器解析 XML 文件
第 9 章 SAX 解析器	介绍如何利用 Java 语言的 SAX 解析器解析 XML 文件

本书面向的读者群

本书可作为高等院校计算机、电子商务以及信息类相关专业 XML 相关课程的教材, 也可供广大 Web 应用程序开发者和用户学习参考, 也适合专职的 XSLT 程序员使用。

本书由范立锋主编, 参加本书编写的主要人员有: 王毅、李世欣、于琦、程峰、吕正超、吴新伟、霍晶馨、唐瑶等。

对于本书实例开发中用到的程序源代码, 读者可以在“人民邮电出版社教学资源与服务网 (www.ptpedu.com.cn)”上免费下载。

虽然本书作者在编写过程中力求精益求精, 尽了最大的努力, 但由于作者能力有限, 书中难免存在一些错误及不足之处, 敬请读者批评指正。

编 者

2009 年 2 月

出版者的话

现今社会对人才的基本要求之一就是应用计算机的能力。在高等学校,培养学生应用计算机的能力,主要是通过计算机课程的体制改革,即计算机教学分层、分类规划与实施;密切联系实际,恰当体现与各专业其他课程配合;教学必须以市场需求为导向,目的是培养高素质创新型人才。

人民邮电出版社经过对教学改革新形势充分的调查研究,依据目前比较成熟的教学大纲,组织国内优秀的有丰富教学经验的教师编写一套体现教学改革最新形势的“高校系列计算机教材”。在本套教材的出版过程中,我社多次召开教材研讨会,广泛听取了一线教师的意见,也邀请众多专家对大纲和书稿做了认真的审读与研讨。本套教材具有以下特点。

1. 覆盖面广,突出教改特色

本套教材主要面向普通高等学校(包括计算机专业和非计算机专业),是在经过大量充分的调研基础上开发的计算机系列教材,涉及计算机教育领域中的所有课程(包括专业核心骨干课程与选修课程),适应了目前经济、社会对计算机教育的新要求、新动向,尤其适合于各专业计算机教学改革的特点特色。

2. 注重整体性、系统性

针对各专业的特点,同一门课程规划了组织结构与内容不同的几本教材,以适应不同教学需求,即分别满足不同层次计算机专业与非计算机专业(如工、理、管、文等)的课程安排。同时本套教材注重整体性的策划,在教材内容的选择上避免重叠与交叉,内容系统完善。学校可根据教学计划从中选择教材的各种组合,使其适合本校的教学特点。

3. 掌握基础知识,侧重培养应用能力

目前社会对人才的需要更侧重于其应用能力。培养应用能力,须具备计算机基础理论、良好的综合素质和实践能力。理论知识作为基础必须掌握,本套教材通过实践教学与实例教学培养解决实际问题的能力和知识综合运用能力。

4. 教学经验丰富的作者队伍

高等学校在计算机教学和教材改革上已经做了大量的工作,很多教师在计算机教育与科研方面积累了相当多的宝贵经验。本套教材均由有丰富教学经验的教师编写,并将这些宝贵经验渗透到教材中,使教材独具特色。

5. 配套资源完善

所有教材均配有PPT电子教案,部分教材配有实践教程、题库、教师手册、学习指南、习题解答、程序源代码、演示软件、素材、图书出版后要更新的内容等,以方便教与学。

我社致力于优秀教材的出版,恳请大家在使用的过程中,将发现的问题与提出的意见反馈给我们,以便再版时修改。

人民邮电出版社

目 录

第 1 章 XML 概述 1

1.1 什么是 XML 1

1.1.1 XML 程序实例 1

1.1.2 XML 的特性 2

1.1.3 XML 与 HTML 的区别 4

1.2 XML 文件的操作 4

1.2.1 XML 文件的编辑 4

1.2.2 XML 文件的保存 5

1.2.3 XML 文件的运行 5

1.3 XML 基础知识 6

1.3.1 XML 的定义 6

1.3.2 XML 的设计目标 6

1.3.3 XML 的样式语言 7

1.3.4 XML 的链接语言 8

1.3.5 XML 的名称空间 9

1.3.6 XML 的文档对象模型 9

1.3.7 XML 的文档类型定义 9

1.3.8 XML 的文档类型模式 10

1.4 XML 解析器 10

1.4.1 XML 解析器简介 11

1.4.2 XML 解析器解析 XML 文件的步骤 11

1.5 XML 的应用领域 12

小结 12

习题 12

第 2 章 XML 文件的规则 14

2.1 XML 声明 14

2.1.1 XML 声明中的版本属性 14

2.1.2 XML 声明中的编码属性 15

2.1.3 XML 声明中的独立属性 16

2.2 XML 中标记设置 16

2.2.1 空标记的设置 17

2.2.2 非空标记的设置 17

2.2.3 非空标记的内容设置 18

2.2.4 标记的名称 18

2.2.5 根标记 19

2.3 特殊字符处理 20

2.4 CDATA 段 21

2.5 标记的文本数据 21

2.6 属性设置 23

2.6.1 属性的构成 23

2.6.2 属性的原则 23

2.7 注释设置 23

2.8 名称空间 24

2.8.1 有前缀和无前缀的名称空间 24

2.8.2 标记中声明名称空间 25

2.8.3 名称空间的作用域 25

2.8.4 名称空间的名字 26

2.9 区分大小写 27

2.10 标记的子标记的关系 27

小结 28

习题 28

第 3 章 XML 数据的底层结构 30

3.1 有效的 XML 文件 30

3.2 如何检查 XML 文件的有效性 30

3.3 XML 文件与 DTD 的匹配 33

3.4 XML 文档类型定义文件 DTD 34

3.4.1 DTD 的元素 34

3.4.2 DTD 的完整性 36

3.4.3 DTD 的属性定义 36

3.4.4 DTD 与名称空间 40

3.4.5 DTD 的实体	41
3.4.6 DTD 中的参数实体	43
3.4.7 DTD 中的注释	45
小结	45
习题	45

第 4 章 XML Schema 模式

4.1 什么是 XML Schema	47
4.2 XML Schema 的文档结构	48
4.3 XML Schema 标记及属性定义	49
4.3.1 简单类型标记的定义	49
4.3.2 自定义数据类型	49
4.3.3 复杂类型标记子标记的定义	53
4.3.4 复杂类型标记的属性的定义	54
4.3.5 空标记及混合内容标记的定义	54
4.4 全局声明和局部声明	55
4.5 XML Schema 的组定义	56
4.6 XML Schema 中的注释	57
4.7 XML 文件与 Schema 文件的匹配	57
4.8 XML 文件对于 XML Schema 模式的校验	58
小结	60
习题	60

第 5 章 使用 CSS 格式化显示

XML 文件

5.1 CSS 技术简介	62
5.2 XML 与 CSS 结合的方式	63
5.2.1 调用独立样式文件	63
5.2.2 将样式语句嵌入到 XML 文件	64
5.2.3 混合方法指定样式	65
5.2.4 使用多个样式文件	65
5.3 CSS 格式化 XML 文件	66
5.3.1 定义样式	66
5.3.2 文本显示方式设置	67
5.3.3 文本属性设置	70

5.3.4 设置元素	72
5.3.5 元素定位显示	75
5.3.6 图片设置	78
5.3.7 环绕文本	80
小结	80
习题	80

第 6 章 使用 XSL 格式化显示

XML 文件

6.1 XSL 的特点	82
6.2 XSL 变换步骤	82
6.3 XSL 的文件结构	84
6.4 XSL 的模式语言	85
6.4.1 相对定位和绝对定位	85
6.4.2 XSL 的路径匹配模式	88
6.5 XSL 的常用控制标记	97
6.5.1 循环标记<for-each>	97
6.5.2 选择标记<if>	99
6.5.3 选择标记<choose>	100
6.5.4 求值标记<value-of>	101
6.6 其他常用 XSL 标记	102
6.6.1 <template> 标记和<apply-templates> 标记	102
6.6.2 <copy>标记	104
6.6.3 <comment>标记	106
6.6.4 <script>标记和<eval>标记	106
6.6.5 <element>标记	107
小结	109
习题	109

第 7 章 XML 的数据源对象

7.1 数据岛的概念	111
7.1.1 内嵌的数据岛	111
7.1.2 外部数据岛	113
7.2 通过记录集管理数据岛	115
7.2.1 创建记录集	115

7.2.2 通过记录集访问数据岛中 各个标记	116	8.7 Attr 节点	153
7.2.3 操作记录集	117	8.8 DocumentType 节点	155
7.2.4 遍历记录集	118	8.8.1 获取 DTD 的基本信息	155
7.2.5 对记录集进行分页	120	8.8.2 获取实体	156
7.2.6 把记录存储到数组中	122	8.9 处理空白	157
7.3 数据管理	124	8.10 验证规范性和有效性	159
7.3.1 关系数据库中数据的存储 与查询	124	8.11 使用 DOM 生成 XML 文件	159
7.3.2 XML 数据中信息的存储与 查询	126	8.12 XML 与 JSP 的结合	162
7.4 XML 与数据库的交互	130	小结	165
7.4.1 数据库到 XML 文档的信息 交互	130	习题	165
7.4.2 XML 文档到数据库的信息交互	133		
7.5 SQL Server 对 XML 的支持	135	第 9 章 SAX 解析器	167
7.5.1 从 SQL Server 数据库中提取 数据到 XML	135	9.1 SAX 解析器	167
7.5.2 XML 数据保存到数据库	138	9.1.1 概述	167
小结	139	9.1.2 工作原理	167
习题	139	9.1.3 事件处理器	168
		9.1.4 事件的产生与处理	170
第 8 章 DOM 解析器	141	9.2 文件的开始与结束	171
8.1 DOM 解析器	141	9.3 XML 中的处理指令	173
8.1.1 DOM 标准	141	9.4 开始标记与结束标记	175
8.1.2 认识 DOM 解析器	142	9.5 文本数据	178
8.2 节点的类型	143	9.6 处理空白	180
8.2.1 Node 接口	143	9.7 名称空间	181
8.2.2 NodeList 接口	144	9.8 实体	183
8.2.3 NamedNodeMap 接口	145	9.9 文件的定位器	187
8.3 Document 节点	145	9.10 处理 SAX 解析过程中发现的 错误信息	189
8.4 Element 节点	147	9.11 SAX 过滤器	193
8.5 Text 节点	149	9.11.1 SAX 过滤器基本原理	193
8.6 CDATASection 节点	151	9.11.2 SAX 过滤器程序设计模式	193
		小结	197
		习题	197

第 1 章

XML 概述

可扩展标记语言（eXtensible Markup Language，XML）在 1996 年开始形成其雏形，并由相关人士向万维网联盟（World Wide Web Consortium，W3C）提案。1998 年 2 月，W3C 正式推出了 XML（XML1.0）。XML 的前身是标准通用标记语言（Standard Generalized Markup Language，SGML）。

在 Internet 的背景下，XML 已经以其应用简单、使用灵活的优势确立了其地位，并得到了迅速的发展。如今，XML 不仅已经广泛用于计算机及网络的各个方面，还应用于机械、物理、化学、数学等领域，并发挥着越来越重要的作用。

1.1 什么是 XML

XML 是一种简单的数据存储语言，它使用一系列简单的标记描述数据，并可以标记任何一种事物。XML 同时也是一组规范，它的用户都遵守这一规范来开发自己的软件或信息，这样，不同的计算机之间就可以相互交流信息。

XML 继承了 SGML 和 HTML 的功能，是一种用于定义标记的语言，又称为“元语言”。XML 的定义只是由框架语法组成，即当用户创建一个 XML 文档时，不必使用固定的标记就可以创建任意喜欢的标记名称，这就是可扩展标记语言中“扩展”的意义。因此，可以使用 XML 描述任意类型的事物。

1.1.1 XML 程序实例

下面先看一个简单 XML 文件，以对 XML 有一个初步的认识。

```
<?xml version="1.0" encoding="gb2312"?>
<persions>
  <persion>
    <name>小张</name>
    <sex>male</sex>
    <birthday>1983.10.25</birthday>
  </persion>
  <persion>
    <name>小王</name>
    <sex>female</sex>
    <birthday>2003.10.20</birthday>
  </persion>
</persions>
```



上述代码通过 IE 浏览器直接运行后的结果如图 1-1 所示。

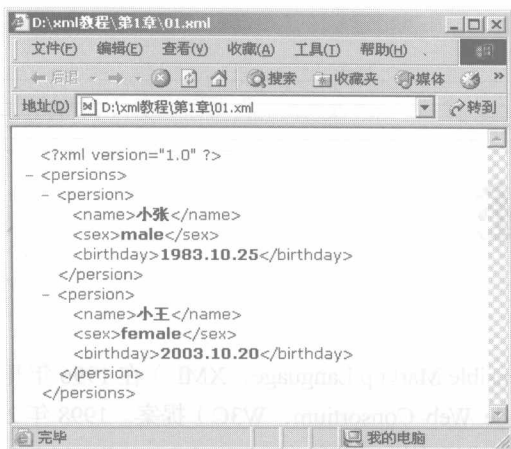


图 1-1 通过 IE 浏览器运行 XML 文件

图 1-1 所示为一个类似源文件的树状结构图，这与 HTML 文档的显示结果完全不同。这是因为 HTML 中的标记是公认的具有固定意义的标记，HTML 文件可以被浏览器直接解析，并能够以一定的样式显示其中的数据。而 XML 文件则侧重于描述数据，其中的标记大多是用户自己定义的，浏览器不能明白其含义，所以，浏览器默认的只是以树状结构显示的 XML 文件内容。

图 1-1 中的 XML 文件是一个规范的 XML 文件，第 1 行是 XML 的声明部分，它声明了 XML 的版本、采用的编码等信息。需要注意的是，XML 的声明必须写在程序的第 1 行，规范的 XML 文件必须含有该语句；接下来 12 行的内容是 XML 文件的主体部分，所使用的标记都是由用户自己定义的，其中，最外面的标记<persons>和</persons>称为根标记。在 XML 文件中，根标记必须存在且唯一，其他的标记都包含在根标记内部，且必须成对出现。

在 XML 规范中，规定 XML 的标记是区分大小写的，所以，下面两个标记被认为是不同的标记。

```
<List>  
<list>
```

1.1.2 XML 的特性

XML 最突出的优点是它可以按照用户自己的意愿随意创建标记。这种优点使得用户能够根据自己的特殊需要制定出适用于自身的一套标记，便于结构化地描述其领域的各类信息，从而提供了一种处理数据的最佳方式。无论是数据的显示和存储，还是数据的传输和处理，都是 XML 的强项。下面将分别介绍 XML 文件的突出特性。

1. 实现不同应用程序之间的数据交互

XML 是跨平台型的，它提供了一种在不同的应用程序之间进行数据库交换的公共标准，是一种公共的交互平台，如图 1-2 所示。

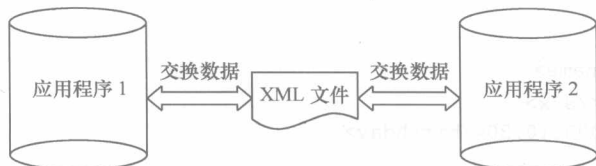


图 1-2 不同的应用程序之间的数据交换

2. 一种数据多种样式

HTML 是显示 Web 数据的专用语言, 而 XML 则是用来存储并传输数据的标记语言。一个 XML 文件并不能决定数据的显示样式, 数据的显示部分必须由其他语言来确定 (通常由 CSS 来确定), 这样就可以按照用户的意愿来给一份数据任意添加多种样式, 如图 1-3 所示。

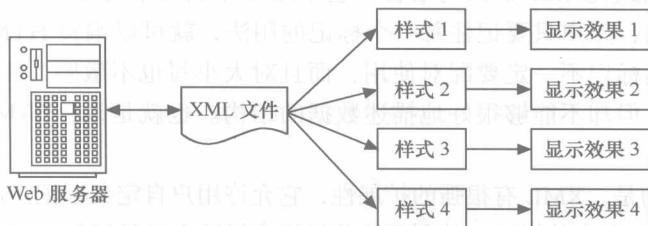


图 1-3 XML 数据样式示意图

XML 将信息的数据部分和信息的数据显示部分分离开来, 这样就可以给同一份数据添加多种样式, 从而得到多种显示效果。这是使用 HTML 无法做到的 (HTML 必须制作多份文件才能得到多种显示效果), 因为 HTML 没有将信息的数据部分和信息的数据显示部分区分开。

3. 实现数据的分布式处理

XML 是一种针对 Internet 设计的标记语言, 它可以在 Internet 上自由传送。当 XML 数据被发送给客户后, 客户可以通过应用软件从 XML 文档中提取数据, 进而对数据进行编辑和处理。XML 文档对象模型允许用脚本或其他编程语言处理 XML 格式的数据。这种情况下的数据处理可以在客户端完成, 而没有必要返回到 Web 服务器上完成, 此方式节省了 Internet 上的数据带宽。而对于原来的 HTML 来说, 即便是对其中的一个字的更改都必须在服务器上进行, 都要导致整个页面的全部数据的重新传送。基于 XML 的分布式数据处理图, 如图 1-4 所示。

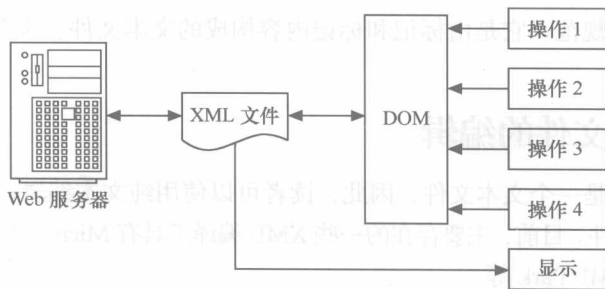


图 1-4 基于 XML 的分布式数据处理图

XML 数据模型的优点就是将原来由服务器处理的许多工作分配到客户端上进行处理, 从而降低了服务器的负担, 优化了服务器的性能。

4. 简单易学、功能强大

HTML 语法结构比较简单, 这种简单性对于 HTML 的推广和普及起到了不可忽视的作用。同样地, XML 继承了 HTML 简单易学的特性, 其简单性也是 XML 标记语言创立之初的一个重要的标准。另一方面, XML 的功能十分强大, 如单独一份 XML 文档在一个样式文件的配合下, 就可以将 XML 中的数据显示成为页面, 达到和 HTML 文件相同的显示效果。

1.1.3 XML 与 HTML 的区别

XML 同 HTML 一样，都是由标记及其所标记的内容构成的文本文件，但它们又有很大的区别。

HTML 是用来格式化 Web 数据的语言，它采用国际上通用的标记。HTML 中的每一个标记都有其固定的用法，用户只要记住每一个标记的用法，就可以编辑 HTML 文件。HTML 的语法也相当宽松，其标记不一定要配对使用，而且对大小写也不敏感。HTML 把数据和其显示格式捆绑在一起，但却不能够很好地描述数据的结构。也就是说，HTML 本质上是一种格式显示语言。

与 HTML 不同的是，XML 有很强的扩展性，它允许用户自定义标记。用户可以根据自己的需要制定出一套适用于自身的标记，并利用这些标记来描述自己领域的信息。尽管用户可以自己定义标记，但是，XML 仍有其严格的语法规则。在 XML 中，标记必须成对使用，并严格区分名称的大小写。XML 的特点就是将信息的内容和它们的显示样式区分开来，而且 XML 是被设计用来描述数据的，其核心是数据的内容。

目前，XML 并没有代替 HTML，还在与 HTML 一起使用。XML 极大地扩展了 Web 页的能力，使 Web 页能够实现如下功能。

- 传递任意类型的文档。
- 用其他方法排序、过滤、重新排列、查找以及管理信息。
- 显示高度结构化的信息。

总之，XML 可以与 HTML 进行互操作，这大大增加了 Web 页面的灵活性。

1.2 XML 文件的操作

XML 文件是一组规范，它是由标记和标记内容构成的文本文件。本节将介绍如何创建和运行 XML 文件。

1.2.1 XML 文件的编辑

一个 XML 文件就是一个文本文件，因此，读者可以使用纯文本编辑器（如 Windows 的“记事本”）来编辑 XML 文件。目前，主要存在的一些 XML 编译工具有 Microsoft XML Parser、XMLSpy、XML Writer、JavaTM XML Park 等。

一个 XML 文件由两个主要的部分组成，即声明和根标记。一个 XML 文件的第 1 行为声明部分，前面不能留有空白，也不能有其他任何语句。XML 的声明结构以“<?xml”开头，以“>”结束，要注意的是，“<”与“?”、“?”与“xml”以及“?”与“>”之间都不要留有空格，否则会发生错误。

一个基本的 XML 声明如下：

```
<?xml version="1.0" ?>
```

XML 的声明部分可以包含一些属性，如“version”、“encoding”等，它指定了 XML 文件的版本号及字符编码等。

一个 XML 文件中有且只有一个根标记，其他标记必须封装在根标记中（注释内容可以不包

含在根标记中)。开始标记与结束标记的名称必须完全一致,包括字母的大小写。例如,下面一行就是错误的 XML 语句。

```
<example_tag>content</Example_tag >
```

XML 中的标记必须正确的嵌套,即,一个标记必须完全包含在另一个标记内,不能交叉使用。下面的代码是不允许出现的。

```
<persion>
  <name>
    小王
  </persion>
</name>
```

综上所述,在编写 XML 文件时需要注意以下几点。

- 根标记必须唯一。
- 标记大小写敏感。
- 标记必须成对出现,正确嵌套。
- 属性值必须用引号括起来。

1.2.2 XML 文件的保存

XML 文件的扩展名为“.xml”。如果 XML 文件指定了文件的编码,则在保存时也必须使用同样的编码,这样 XML 解析器才能识别 XML 中的标记,并能正确地解析出所标记的内容。例如,保存一个名为“example.xml”的文件时,若文件所采用的编码是“UTF-8”,则保存时文件的编码也应为“UTF-8”,如图 1-5 所示。



图 1-5 保存文件

1.2.3 XML 文件的运行

XML 文件一般是配合其他应用程序而使用的,要想单独运行 XML 文件,最简单的方法是用 IE 浏览器直接打开 XML 文件。例如,在用 IE6.0 打开编辑后的 XML 文件时,如果该文件是一个规范的 XML 文件,浏览器就会显示出 XML 的源文件;否则将会显示出错信息。因此,用户可以用这个方法检查 XML 文件是否规范。1.1.1 小节中的实例在 IE6.0 中的显示结果,如图 1-6 所示。

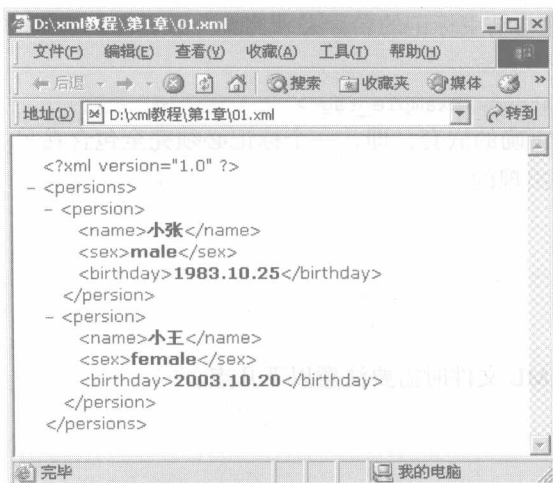


图 1-6 程序显示结果

1.3 XML 基础知识

前两节内容对什么是 XML, 以及 XML 的特点做了相关的介绍, 本节将介绍 XML 的一些基础知识。

1.3.1 XML 的定义

W3C 对 XML 的定义是: “可扩展标记语言 (XML) 是 SGML 的子集, 其目标是允许普通的 SGML 可以在 Web 上以目前 HTML 的方式被服务、接收和处理。XML 被设计成易于实现, 且可在 SGML 和 HTML 之间互相操作。”

1.3.2 XML 的设计目标

根据 W3C 对 XML 的定义, 其目标也反映出了 XML 的一些特性。

(1) XML 应该可以在 Internet 上直接使用, 以支持各类不同的应用程序。设计 XML 的目标就是使 XML 成为一个标准, 并供其他类型软件使用。

(2) XML 应该与 SGML 兼容。XML 是 SGML 的一个具有特殊用途的子集。这个特点使得 SGML 的软件工具可以轻易地加入到 XML 中来并同其共同运作。

(3) 可以很容易地撰写处理 XML 文件的程序。要使 XML 能够广泛地应用, 那么浏览器以及其他负责处理 XML 文件的程序应该可以很容易地被撰写。因为, 推出 XML 的主要原因之一就是撰写负责处理 SGML 文件的程序太复杂。

(4) XML 中的选择性功能的数目应该少, 理想情况是零。在 XML 中, 使用最少数目的选择性功能可以让开发人员在撰写程序来处理 XML 文件时变得较为容易。SGML 中有很多选择性功能, 使得 SGML 更灵活, 功能更强大, 但也导致了 SGML 的高复杂性, 这也是 SGML 无法定义网站文件的主要原因。XML 的这个目标也保证了 XML 的跨平台性。

(5) XML 文件应该是结构合理、内容清楚的。XML 的设计目标是使之成为一个世界上使用

者与应用程序之间往来信息的通用媒介。XML 类似于一个数据库，但又区别于数据库，这是因为 XML 文件是以纯文字撰写的，且拥有逻辑的树状结构的文件。所以，XML 文件应该被人们或计算机轻易地阅读。

(6) XML 的设计应该简洁。由于 SGML 的高复杂性，使其不能被广泛应用，所以，XML 只有在程序设计人员与使用者愿意接受它的情况下，才能成为可实行的标准。因而，XML 要尽力避免这一现象，使其成为一种很容易被人们使用的标记语言。

(7) XML 的设计应该语法严格并且简洁。这一目标使得 XML 的处理软件变得更为容易，进而促进了 XML 被接受的可能性。

(8) XML 文件应该容易建立。XML 的出现是为了让所有用户都能够使用，所以 XML 应该简单明了，且应该使任何人都能够轻松地创建 XML 文件。

(9) 简化 XML 标签并不重要。XML 不应该为了简洁而简化，XML 标签不能太过简洁而使其难以阅读。

1.3.3 XML 的样式语言

XML 的特点就是数据与样式的分离，它不提供显示数据的功能。如果用户想要浏览器只显示 XML 中的数据，就需要告诉浏览器以一定的样式进行显示，甚至是最基本的样式。

W3C 提供了两种通用的样式语言，层叠样式表 (Cascading Style Sheet, CSS) 和可扩展样式语言 (eXtensible Style Language, XSL)。下面通过例子简单说明一下，其具体内容将在第 5 章和第 6 章详细介绍。

1. 使用 CSS 显示 XML 数据内容

XML 文件见 1.1 节中的 XML 文件，格式化显示该 XML 文件的 CSS 样式文件代码如下。

```
name{ display:block; font-size:18px;}
sex{ display:block;font-size:18px;}
birthday{ display:block;font-size:18px;}
```

这个简单的 CSS 样式文件定义了 XML 文件中标记 <name>、<sex>、<birthday> 所标记的内容的显示样式。要使用该样式，就必须将 XML 文件与之关联。关联的方法是在 XML 文件中声明之后添加一条指令语句，具体代码如下。

```
<?xml version="1.0" ?>
<?xml-stylesheet type="text/css" href="show.css" ?>
<persions>
<persion>
..... <!--省略的代码-->
```

小张
male
1983.10.25
小王
female
2003.10.20

程序运行效果如图 1-7 所示。

图 1-7 使用 CSS 的显示结果

2. 使用 XSL 显示 XML 数据内容

使用 XSL 格式化显示 XML 文件类似于使用 CSS，首先也要编辑好 XSL 文件，然后再通过指令关联到想要显示的 XML 文件。例如，仍然使用 1.1 节中的 XML 文件，XSL 文件的代码如下。

```
<?xml version="1.0" ?>
<!-- XSL 声明 -->
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<!-- 根模板 -->
<xsl:template match="/">
<!-- 嵌套 HTML 标记 -->
```