

高等学校计算机应用人才培养系列教材

XML技术应用

贾素玲 王强 主编
姚琪琳 舒毕磊 柴庆慧 编著



清华大学出版社

中国计算机教育出版集团 中国铁道出版社

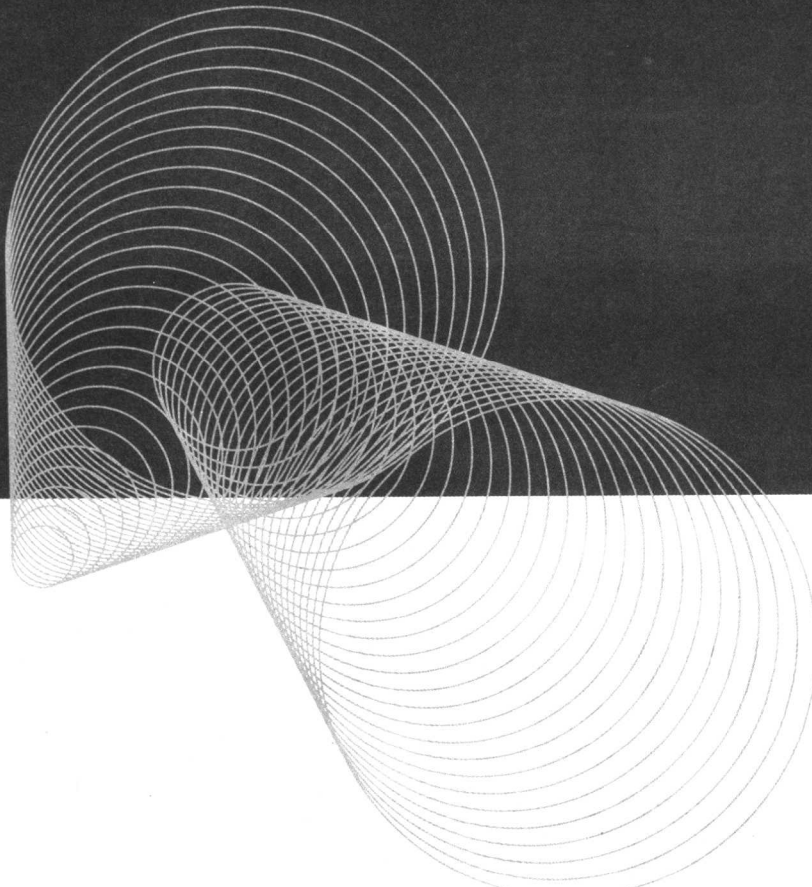
XML技术应用

XML 1.0 应用案例
XML 2.0 应用案例
XML 3.0 应用案例

高等学校计算机应用人才培养系列教材

贾素玲 王强 主编
姚琪琳 舒毕磊 柴庆慧 编著

XML技术应用



清华大学出版社
北京

内 容 简 介

本书按照万维网联盟(W3C)的成熟标准和发展趋势,带领读者循序渐进地学习 XML 的相关知识,并展示了大量的应用实例。全书的主要内容包括 XML 概述、XML 语法、文档类型定义、文档对象模型、XSL 转换、XML 中的链接、命名空间、XML Schema 以及 Web Service。同时,本书包括两个附录,附录 A 介绍了一个简单易用的 XML 集成开发环境 XML SPY 的安装及使用;附录 B 介绍了一组由 Java 语言提供的针对 XML 的 API。

本书以 XML 1.0 版本为基础,语言简洁、内容丰富、涵盖面广,并且包含丰富的实例程序,适合大学本科计算机、信息管理与信息系统及其相关专业的 XML 基础教材,同时也可作为相关研究人员和工程技术人员的参考书籍。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

XML 技术应用/贾素玲,王强主编;姚琪琳等编著. —北京:清华大学出版社,2007.7

(高等学校计算机应用人才培养系列教材)

ISBN 978-7-302-14958-3

I. X… II. ①贾… ②王… ③姚… III. 可扩充语言,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(2007)第 044234 号

责任编辑:索 梅 顾 冰

责任校对:李建庄

责任印制:王秀菊

出版发行:清华大学出版社 地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn> 邮 编:100084

c-service@tup.tsinghua.edu.cn

社 总 机:010-62770175 邮购热线:010-62786544

投稿咨询:010-62772015 客户服务:010-62776969

印 刷 者:北京市清华园胶印厂

装 订 者:三河市溧源装订厂

经 销:全国新华书店

开 本:185×260 印 张:16.5 字 数:399 千字

版 次:2007 年 7 月第 1 版 印 次:2007 年 7 月第 1 次印刷

印 数:1~4000

定 价:24.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:022482-01

打开本套丛书的朋友,如果你抱有获得 IT 应用成功的愿望或想尽快加入令人羡慕的 IT 行业发展的期盼,相信这套丛书将使你豁然开朗。

信息技术是当代人类社会中发展最快、渗透性最强、应用面最广的先导技术,这些日新月异的新技术不仅改变了世界,也改变了人们的生活。掌握计算机基础知识,提高计算机应用能力,不再是计算机专业人士的特长,并且已经成为许多行业必备的基本技能。计算机技术不仅仅是一门独立的学科,而且日益成为其他学科飞速发展的助推器。对于当代的大学生,或者想进入五彩缤纷的 IT 应用领域的朋友,要跟上信息时代的步伐,就必须掌握现代科学技术,调整自己的知识结构,使自己具备开拓创新的意识和能力,以适应当前社会发展。

信息时代离不开信息化人才,掌握计算机基础知识和提高应用能力,是信息化人才培养的一个重要环节。我们知道,印度是当今的计算机软件出口王国,软件产业的发展成就令人赞叹。之所以如此,除了政府长期实施的一系列扶持政策外,主要还是得益于持之以恒、行之有效的系列化的优秀教材和教育培训。

本套系列教材的形成,一是根据教育部人才培养的指导方针,以培养 IT 应用人才为目标,在引进推广印度 IT 培训教材的同时,借鉴国内外的计算机专业、信息管理专业人才培养的经验,力求课程的设置重点突出、循序渐进,将知识学习与能力培养相结合,使理论与实践完美融合;二是以企业对信息化人才的需求为依据,把面向对象、数据库、软件体系结构、软件工程的思想融入教材体系中,将基本技能的培养与主流应用技术相结合,培养具有扎实基础的实用型人才;三是在多年从事信息化人才培养和信息系统项目开发的经验基础上,充分理解企业人才需求层次和大学传统人才培养模式的错位,把课程体系的理论知识学习成功转变为应用能力的掌握,使大学真正成为企业的人才资源库。这就是我们开发这套系列教材的最终目标。

计算机基础和语言类的教材,可以说是多如牛毛,那么本套系列教材又凭什么在市场竞争中获得优势呢?

一是知识系列化。本套系列教材以“面向对象、数据库、软件工程、信息系统开发、……、项目管理”的思想为主线,以 Java、XML 为主流技术,形成系列化的能力培养阶梯,使得学生能从一个初学入门者,逐渐成长为合格的 IT 应用技术开发人才。

二是理论实践化。本套系列教材从应用实践的需要入手,合理组织每门课程的结构和内容,在总体框架下,通过大量的案例训练使学生掌握程序设计语言的核心技术和应用技巧,使理论知识在实践中得到升华,在不知不觉中能力得到积累和提高。

三是案例实战化。本套系列教材的编著者既是教师又是软件工程师,具有丰富的教学经验和软件项目开发经验,善于把握计算机技术与学生能力需求之间的尺度,按照循序渐进、突出重点的原则,从多个承担的科研项目中精心抽取和设计教材中引入的案例,使案例

与技术更接近实战的要求。

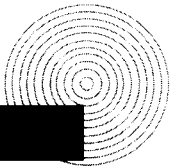
四是重点内容课件化。除学生用书外,本套系列教材还配有完善的课件,既可以直接为教师上课服务,也能为学生提供本课程的学习要点,引导学生深刻理解每一章节的主体内容,轻松完成相关知识的学习和案例设计。

总之,本套系列教材的指导思想是力求内容新颖、概念清晰、结合实践需要,突出应用能力的培养,使学生在循序渐进的学习中,达到软件项目开发的技术能力要求,成为满足企业信息化需要的人才。

本套系列教材在编写过程中得到了多方专业人士的指导、支持和帮助,在此表示由衷的感谢。尽管我们在教材编著时力求准确,但难免存在不当之处,恳请各位同仁和读者批评指正。

编 者

2007年3月10日



前言

毫无疑问,21 世纪初是 Web 的时代。随着 Internet 的飞速发展,各种新技术应运而生。标记语言特别是 HTML 的出现改变了计算机的发展方向,以文档对象为基础、以浏览器为载体、内容更丰富且更具实时性的图形界面取代了单调的命令行界面,没有任何网络基础的普通用户也可以通过鼠标点击而轻松地阅读新闻和收发邮件。然而,HTML 本身却存在着很多缺陷。首先,HTML 是一种样式语言,它目前在 Internet 中扮演的只是数据表示的角色,随着信息量的增多,HTML 变得越来越难以胜任;其次,HTML 对浏览器的过度依赖性也形成了 HTML 标准的严重不统一,从而导致许多信息表示只能由某种特定的浏览器来解释。HTML 的这些不足使得人们重新思考 HTML 在 Internet 上的角色,并开始研究一门新的语言来弥补 HTML 的缺陷,XML 的产生就正是这种思考的最终结果。

XML 是一种专门为 Internet 所设计的标记语言,它的重点是管理数据本身,数据的表示形式交给其他技术来解决。这种明确的分工带来了更高效的程序设计、更快速的搜索引擎、更统一的数据表示方式以及更方便的数据交流。

XML 是一种定义语言的语言,它克服了 HTML 的缺点,给了程序员更自由的空间。近几年,XML 在除数据表示以外的其他方面也得到越来越广泛的应用,从数据存储到数据交换再到系统整合,XML 都发挥着强大的作用。基于 XML 的新技术也如雨后春笋般层出不穷,Web 服务、AJAX 等技术相继形成,并逐渐开始发展成熟。W3C(万维网联盟)也不断为 XML 制定新的标准,对其进行完善。毫不夸张地说,XML 一统 Web 天下的局面已经指日可待。

因此,作为计算机专业的学生,了解并掌握 XML 是十分必要的。本书面向 XML 的初学者,避开 XML 的高级应用不谈,而是从最简单、最基础的内容入手,使读者把基本的知识牢牢掌握,这样再学习和研究高级应用就会容易得多。

本书共分为 8 章。第 1 章从标记语言入手,引出 XML 的概念和相关技术,并介绍了如何创建一个最简单的 XML 文档。第 2 章重点介绍 XML 语法,并学习如何创建格式良好的 XML 文档。第 3 章介绍文档类型定义(DTD),DTD 可以详细说明允许或者不允许哪些元素出现在文档中,以及这些元素的确切内容和结构。在第 3 章中,读者将学习如何定义元素、实体和属性,如何将 DTD 附加到文档上,如何使用 DTD 验证文档的正确性以及如何编写自己的 DTD 来解决遇到的问题。第 4 章介绍文档对象模型(DOM),DOM 提供了一系列接口,使应用程序可以方便地访问 XML 文档,本章将重点介绍这些接口。第 5 章为 XSL 转换,XSL 是比 CSS 更复杂更强大的样式语言,XSL 转换是 XSL 的一种,它可以将 XML 树状结构转化为其他形式的树状结构。第 6 章介绍了 XLink 和 XPointer,XLink 提供多方向的

超文本链接,XPointer 不仅可以链接到特定文档,还可以链接到特定文档的个别部分。第 7 章介绍了命名空间和 Schema,命名空间可以解决名称冲突的问题,而 Schema 则弥补了 DTD 的一些不足。第 8 章介绍了 XML Web Service,并引出 WSDL 与 SOAP。

第 3~5 章为本书的重点,应该着重掌握。

由于时间仓促,本书不足之处恳请读者批评指正。

编 者

2007 年 6 月

读者意见反馈

亲爱的读者：

感谢您一直以来对清华版计算机教材的支持和爱护。为了今后为您提供更优秀的教材，请您抽出宝贵的时间来填写下面的意见反馈表，以便我们更好地对本教材做进一步改进。同时如果您在使用本教材的过程中遇到了什么问题，或者有什么好的建议，也请您来信告诉我们。

地址：北京市海淀区双清路学研大厦 A 座 602 室 计算机与信息分社营销室 收
邮编：100084 电子邮箱：jsjjc@tup.tsinghua.edu.cn
电话：010-62770175-4608/4409 邮购电话：010-62786544

教材名称：XML 技术应用

ISBN：978-7-302-14958-3

个人资料

姓名：_____ 年龄：_____ 所在院校/专业：_____

文化程度：_____ 通信地址：_____

联系电话：_____ 电子信箱：_____

您使用本书是作为：☐指定教材 ☐选用教材 ☐辅导教材 ☐自学教材

您对本书封面设计的满意度：

☐很满意 ☐满意 ☐一般 ☐不满意 改进建议_____

您对本书印刷质量的满意度：

☐很满意 ☐满意 ☐一般 ☐不满意 改进建议_____

您对本书的总体满意度：

从语言质量角度看 ☐很满意 ☐满意 ☐一般 ☐不满意

从科技含量角度看 ☐很满意 ☐满意 ☐一般 ☐不满意

本书最令您满意的是：

☐指导明确 ☐内容充实 ☐讲解详尽 ☐实例丰富

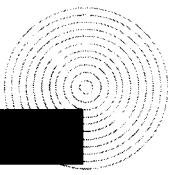
您认为本书在哪些地方应进行修改？（可附页）

您希望本书在哪些方面进行改进？（可附页）

电子教案支持

敬爱的教师：

为了配合本课程的教学需要，本教材配有配套的电子教案（素材），有需求的教师可以与我们联系，我们将向使用本教材进行教学的教师免费赠送电子教案（素材），希望有助于教学活动的开展。相关信息请拨打电话 010-62776969 或发送电子邮件至jsjjc@tup.tsinghua.edu.cn 咨询，也可以到清华大学出版社主页（<http://www.tup.com.cn> 或 <http://www.tup.tsinghua.edu.cn>）上查询。



目 录

第 1 章 XML 概述	1
1.1 什么是 XML	1
1.1.1 标记语言	1
1.1.2 XML 的起源	2
1.2 XML 的优点	3
1.2.1 XML 的特性	3
1.2.2 XML 的优点	5
1.3 XML 的处理过程	6
1.3.1 编辑器	6
1.3.2 语法分析和处理程序	6
1.3.3 用户端应用程序	6
1.3.4 XML 处理过程的总结	7
1.4 XML 的设计目标	7
1.5 XML 的相关技术	8
1.5.1 文档类型定义(DTD)	9
1.5.2 文档对象模型(DOM)	9
1.5.3 级联样式表(CSS)	9
1.5.4 可扩展的样式语言(XSL)	10
1.5.5 URL 和 URI	10
1.5.6 可扩展的链接语言(XLL)	11
1.6 创建并显示简单的 XML 文档	11
1.6.1 Hello XML World	11
1.6.2 显示 XML 文档	12
本章小结	14
习题 1	14
第 2 章 XML 语法	15
2.1 XML 文档结构	15

2.1.1	序言区	15
2.1.2	主体区	16
2.1.3	尾声区	17
2.2	元素和标记	17
2.2.1	元素的命名	17
2.2.2	标记	18
2.2.3	空元素	20
2.2.4	元素的嵌套	20
2.3	属性	21
2.3.1	属性的命名	21
2.3.2	属性的值	22
2.4	引用	22
2.4.1	实体引用	23
2.4.2	字符引用	23
2.5	处理指令	24
2.6	注释	24
2.7	CDATA 节	25
2.8	创建格式良好的 XML 文档	26
2.8.1	XML 文档的声明	26
2.8.2	元素	26
2.8.3	属性	29
2.8.4	其他	30
2.8.5	文档全文	31
	本章小结	37
	习题 2	37

第 3 章 文档类型定义 39

3.1	什么是 DTD	39
3.2	DTD 的调用	40
3.2.1	内部 DTD 的声明	41
3.2.2	外部 DTD 的声明	43
3.2.3	内部 DTD 和外部 DTD 的联合使用	45
3.3	DTD 的结构	45
3.4	元素的声明	46
3.4.1	#PCDATA	46

3.4.2	空元素	48
3.4.3	子元素	49
3.4.4	混合型	59
3.4.5	ANY	59
3.5	实体的声明	59
3.5.1	内部一般实体	60
3.5.2	外部一般实体	65
3.5.3	内部参数实体	68
3.5.4	外部参数实体	71
3.6	属性的声明	73
3.6.1	设置属性的类型	74
3.6.2	属性的默认设置	83
	本章小结	87
	习题 3	88
第 4 章	文档对象模型	89
4.1	DOM 概述	89
4.1.1	什么是 DOM	89
4.1.2	DOM 的结构	90
4.1.3	DOM 的工作方式	91
4.1.4	DOM 的规范	91
4.1.5	DOM 规范的约定	92
4.1.6	正确理解 DOM	92
4.2	DOM API	93
4.2.1	什么是 DOM API	93
4.2.2	DOM 接口继承树	94
4.2.3	DOM API 核心接口介绍	95
4.3	利用 DOM API 处理结构化文档	106
4.3.1	遍历 XML 文档	106
4.3.2	处理结点	106
4.4	DOM 的简单应用	109
4.4.1	Java XML 解析器	109
4.4.2	遍历文档	110
4.4.3	添加结点	114
4.4.4	删除结点	124

4.4.5 替换结点	130
本章小结	134
习题 4	134
第 5 章 XSL 转换	136
5.1 什么是 XSL	136
5.2 XSLT 概述	136
5.2.1 转换语言	136
5.2.2 树状结构	137
5.2.3 在什么时候使用 XSLT	138
5.2.4 规则、模板和模式	139
5.2.5 XSLT 文档	139
5.3 XSL 模板	140
5.3.1 xsl:stylesheet 元素、xsl:template 元素和 match 属性	140
5.3.2 xsl:apply-templates 元素、xsl:value-of 元素和 select 属性	142
5.3.3 xsl:for-each 元素	143
5.3.4 结点的值	144
5.4 匹配结点的模式	145
5.4.1 匹配根结点、子结点以及其他后代结点	146
5.4.2 匹配元素名称	147
5.4.3 通配符	147
5.4.4 使用 ID 匹配单个元素	147
5.4.5 使用@匹配属性结点	147
5.4.6 使用 comments()匹配注释结点	149
5.4.7 使用 processing-instruction()匹配处理指令结点	149
5.4.8 使用 text()匹配文本结点	150
5.4.9 使用或(OR)操作符“ ”	150
5.4.10 使用[]进行测试	151
5.5 选择结点的 XPath	153
5.5.1 XPath 和表达式	153
5.5.2 结点轴	153
5.5.3 表达式类型	155
5.5.4 路径定位举例	159
5.6 XSL 输出	160
5.6.1 属性值模板	160

5.6.2	xsl:element 元素	161
5.6.3	xsl:attribute 元素和 xsl:attribute-set 元素	161
5.6.4	生成处理指令、注释及文本	163
5.6.5	使用 xsl:copy 元素复制上下文结点	163
5.6.6	使用 xsl:number 元素对结点进行数字处理	164
5.6.7	使用 xsl:sort 元素对输出元素进行排序	166
5.6.8	使用 xsl:variable 元素定义常量	167
5.6.9	使用 xsl:if 和 xsl:choose 元素进行选择	168
5.6.10	模板的调用	171
5.6.11	选择输出方式	174
5.7	合并多个样式表	174
	本章小结	175
	习题 5	176
第 6 章	在 XML 中使用链接	179
6.1	XML 中的链接	179
6.2	XLink 概述	179
6.2.1	什么是 XLink	179
6.2.2	XLink 的设计原则	180
6.3	链接元素	181
6.3.1	简单链接	181
6.3.2	链接行为	183
6.3.3	扩展链接	184
6.3.4	弧元素	186
6.3.5	行外链接	188
6.4	XPointer 概述	189
6.5	使用 XPointer 访问信息	191
6.5.1	定位点	191
6.5.2	根结点	192
6.5.3	定位轴	192
6.5.4	谓词	193
6.5.5	范围	193
	本章小结	193
	习题 6	194

第 7 章 命名空间和 XML Schema	196
7.1 什么是命名空间	196
7.2 命名空间的语法	198
7.2.1 用 xmlns 属性定义命名空间	198
7.2.2 多个命名空间	199
7.2.3 属性	201
7.2.4 默认的命名空间	202
7.3 命名空间和有效性	203
7.4 Schema 与 DTD	203
7.5 Schema 的有效性检验	206
7.6 Schema 的语法	207
7.6.1 数据类型	207
7.6.2 minOccurs 和 maxOccurs	209
7.6.3 属性的声明	210
本章小结	211
习题 7	212
第 8 章 XML Web Service	213
8.1 Web Service 概述	213
8.1.1 Web Service 的发展经历	213
8.1.2 Web Service 的特点	214
8.1.3 Web Service 的体系结构	215
8.2 Web Service 协议	217
8.3 WSDL	218
8.3.1 WSDL 的结构	219
8.3.2 WSDL 的主要元素	220
8.3.3 WSDL 示例	222
8.3.4 绑定类型	223
8.4 SOAP	224
8.4.1 SOAP 简介	224
8.4.2 SOAP 的设计目标	225
8.4.3 SOAP 的示例	225
8.4.4 SOAP 的消息交换模式	226
8.4.5 SOAP 消息	227

8.4.6 SOAP 的编码规则	228
8.4.7 HTTP 上的 SOAP	229
8.4.8 在 RPC 中使用 SOAP	230
8.5 UDDI	230
本章小节	232
习题 8	233
附录A XMLSPY 简介	234
A.1 XMLSPY 概述	234
A.2 安装 XMLSPY 5	234
A.3 使用 XMLSPY 的主界面	237
附录B XML 的 Java API	242
B.1 JAXP	242
B.2 JAXB	244
B.3 JAXM	245
B.4 JAXR	246
B.5 JAX-RPC	247
参考文献	248

第 1 章 XML 概述

本章将介绍什么是 XML、XML 的特性是什么以及它的优势体现在什么地方等内容,同时帮助读者理解与 XML 相关技术的关键部分是如何协同工作的。希望通过本章的学习,能够使读者对 XML 有一个大概的了解。

1.1 什么是 XML

在描述什么是可扩展标记语言(extensible markup language, XML)以及 XML 的起源之前,首先来了解什么是标记语言(markup language)。

1.1.1 标记语言

标记语言就是使用某种“记号”来表示某种特殊信息的语言,它是一套标记符号和相关语法的集合。它通过其特定的标记符号和语法传达某种标记信息。标记语言由一些格式标记(code)或者控制标记(tag)组成。这些标记决定了信息显示的格式或数据的意义,但是,标记语言本身不能单独存在,它只能用于装饰它所包括的文本内容,如以下这一段代码:

```
<bold>足球队资料</bold>
```

`<bold></bold>`就是所使用的标记,“足球队资料”就是它所修饰的文本内容,它本身是没有意义的,但是结合了这些文本信息后,就对这些文本信息的显示格式进行了控制,在这里就是使得文本“足球队资料”以加粗的形式显示。同样再看下面一段代码:

```
<国家>意大利</国家>
```

`<国家></国家>`也是一种标记,它用来说明它所包括的文本信息的意义。

目前大多使用两种标记语言:专用的标记语言与通用的标记语言。专用的标记语言针对特定应用程序与设备,产生专用的编码,这些标记语言是为了满足特定的需求而建立的。通用标记语言则描述文件中文字的意义与结构,但并不定义如何使用文字,这类标记语言并不是为了满足任何特定应用程序而设计,而是为了让许多不同的、一般性的程序使用。通用标记语言所描述的文件比起专用标记语言描述的文件有较佳的移植性。