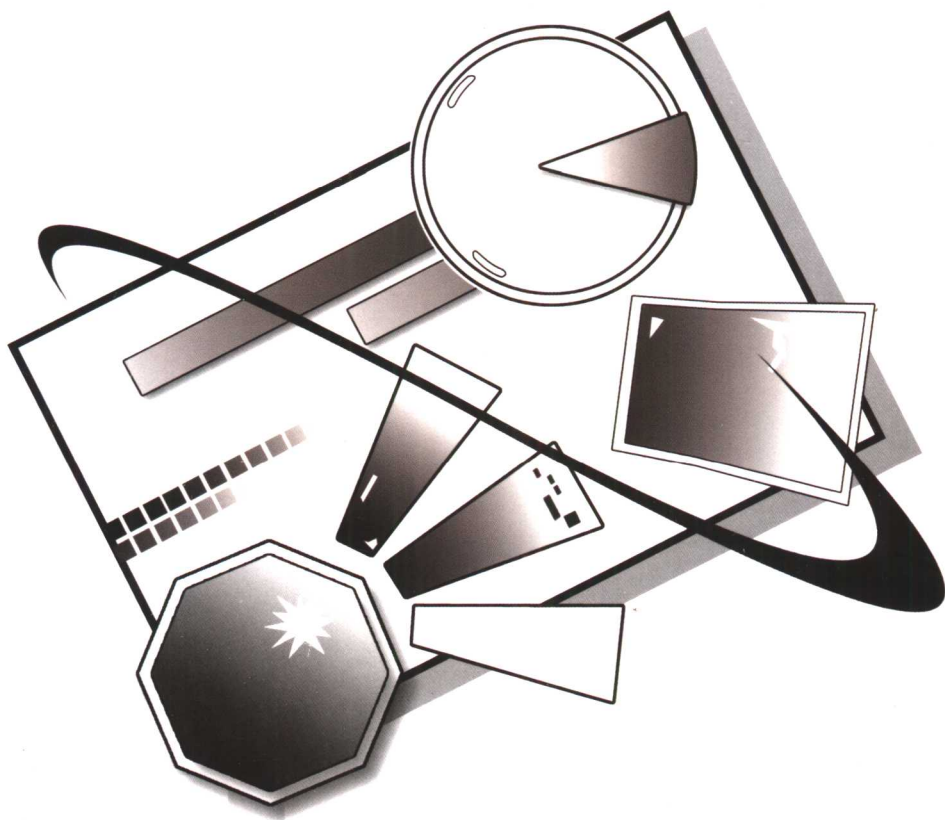




XML

实用教程



网上免费提供
实例源代码

丁跃潮 叶文来 陈杰 编著



机械工业出版社
CHINA MACHINE PRESS



XML 实用教程



XML



实用教程

王卫宁 叶卫明 编 陈 强 主编

中国铁道出版社



高等职业教育计算机专业推荐教材

XML 实用教程

丁跃潮 叶文来 陈 杰 编著



机械工业出版社

本书是学习 XML 语言的实用教材, 全书共分 9 章, 主要包括: XML 的基本概念、语法规则、文档类型定义 (DTD)、Schema 结构、层叠样式单 (CSS)、数据源对象 (DSO)、文件转换 (XSLT) 和文档对象模型 (DOM) 等。最后通过一个网上书店的实例讲解 XML 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。每章都包括学习目标、小结和习题, 便于读者巩固所学的知识。

本书适合具有一定计算机基础知识的读者阅读。它可作为计算机及相关专业高职高专 Web 设计或 XML 课程的教材, 还可供从事 Web 应用软件设计的科研人员参考。

图书在版编目 (CIP) 数据

XML 实用教程 / 丁跃潮等编著. —北京: 机械工业出版社, 2006.3

(高等职业教育计算机专业推荐教材)

ISBN 7-111-18467-X

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

中国版本图书馆 CIP 数据核字 (2006) 第 008169 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策 划: 胡毓坚

责任编辑: 李利健

责任印制: 洪汉军

北京鑫海金澳胶印有限公司印刷

2006 年 3 月第 1 版 · 第 1 次印刷

787mm×1092mm 1/16 · 14.25 印张 · 348 千字

0001—5000 册

定价: 21.00 元

凡购本图书, 如有缺页、倒页、脱页, 由本社发行部调换

本社购书热线电话 (010) 68326294

编辑热线: 88379739

封面无防伪标均为盗版

高等职业教育计算机专业推荐教材

编委会成员名单

主 任 王元元

编 委	丁跃潮	黄陈蓉	黄国兴
	李咏梅	逯燕玲	王爱梅
	奚李峰	杨世平	张桂芸

编者的话

根据有关部门对我国信息产业发展的客观需求及劳动力市场现状的调查,在计算机应用和软件专业领域培养技能型紧缺人才,是当务之急。近年来,不仅高等职业技术类院校,而且相当一部分本科类工程技术院校(特别是相当数量高等学校的二级学院、民办院校),都把招收和培养计算机专业技能型紧缺人才列为教育改革的重要举措。为一些院校提供“适时、适度、优选、优质”的计算机专业的高等职业教育系列教材,正是我们组织编写这套“高等职业教育计算机专业推荐教材”(以下简称“推荐教材”)的目标。“推荐教材”由四个模块的30多本教材组成。这些模块是:基础知识模块、程序设计模块、实用技术模块、实践模块。

这套“推荐教材”是“适时”的,因为它努力适应我国信息产业发展和劳动力市场的客观需求,适应计算机行业技术的现状,强调教学内容的先进性和实用性。这套教材十分关注信息技术的最新发展,突出本专业领域的新知识、新技术、新流程和新方法。其中程序设计模块和实用技术模块充分体现了这一特色,所涉及的19本教材既有基础的平台、语言,如《Linux 操作系统》、《C 语言程序设计与实践》,也有最新的《Visual C#.NET 面向对象程序设计教程》、《XML 实用教程》、《JSP 应用教程》等工具,还有十分接近实际工作需要的《Oracle 数据库应用教程》、《计算机网络管理》、《电子商务概论》等实用教材。

这套“推荐教材”是“适度”的,因为它不是简单地摒弃基础理论,而是注意强调理论联系实际,努力做到专业技能型人才能从中学学习到必要和相对系统的基础理论知识,把各种能力的培养和全面素质的提高放在首要的位置。“推荐教材”中基础知识模块的设置,充分体现了这一特色,它囊括了从数学基础、电子基础、硬件技术基础到系统软件基础、应用技术基础、网络技术基础、信息安全基础等10本教材。

这套“推荐教材”是“优选”的,因为充分考虑了现有高中毕业生的认知水平和已有知识,为学生提供适应劳动力市场需要和有职业发展前景的、模块化的教材体系。在学习内容、教学组织等方面留给教师和学生选择和创新的空間,便于教师组织和构建开放式的课程体系,适应学生个性化发展的需要,在灵活的模块化课程结构中自由发展。“推荐教材”的四个模块对重要内容都安排了看似重复的多种教材,供教师和学生去选择。例如,可以在《C 语言程序设计与实践》、《Visual Basic.NET 可视化程序设计教程》、《Visual C#.NET 面向对象程序设计教程》、《Java 程序设计教程》中任意选择一到两门,也可以在《ASP 基础及应用教程》、《JSP 应用教程》中任选一门。

这套“推荐教材”是“优质”的,因为它们作者多数是从事高等职业教育的计算机专业教师,具有长期的计算机实际工作和教育工作经验。这套教材的优质,还体现在它的改革和创新精神上。其中《计算机电路基础》对传统的电路、模拟电路和数字电路课程教材作了重大的改变;《计算机组装与维修教程》则是一门纯实践的课程教材。我们欢迎使用这套教材的师生,指出教材中存在的问题并提出修改意见。

高等职业教育计算机专业推荐教材
编委会

前 言

可扩展标记语言 (eXtensible Markup Language, XML) 作为一种计算机技术出现不过是短短几年的事情。虽然它是一种表面上非常简单的技术,但对计算机软件、数据库管理和 B2B (Business to Business) 通信世界的广泛适用性和深远影响令人惊讶。这种基于标记的可扩展语言提供了强大的功能,不但可以为构建数据提供一种易于处理的高度结构化的方式,而且能够存储、交换关于数据意义的信息,是跨平台的信息交换的主要途径。人们越来越发现 HTML 这种语言已经远远不能满足网络潜在应用能力的需要,而 XML 良好的格式、灵活的 Web 应用、便捷的数据处理、面向对象的特点,以及开放的标准和数据的选择性更新等特点却越来越受到人们的青睐。

本书系统介绍了设计和使用 XML 文档所必需的基本技术,以及很多现在非常重要的相关技术。通过本书的学习,使学生对 XML 语言有一个全面的了解,掌握相关的主要知识点,在软件设计中运用 XML 技术灵活地进行信息的描述、交换和显示。全书共分 9 章,内容如下:

第 1 章介绍 XML 标识语言的引导知识。包括标记语言发展、XML 的特点、应用前景、数据结构、开发和应用环境,最后建立简单的引例。

第 2 章讲解 HTML 基本语法概要、XML 语法、XML 的相关技术和 XML 名称空间,并讲述编辑 XML 文档的步骤。

第 3 章论述 DTD 的作用,详细说明 DTD 元素、属性和实体的语法规则,并讲解 DTD 的使用方法。

第 4 章介绍 XML Schema 的基础知识,Schema 中元素、属性和类型等构件,模式定义基本要求以及命名空间的概念。

第 5 章介绍 CSS 的作用,介绍 CSS 语法及常见的 CSS 属性的设置,并以实例的方式讲述 XML 中如何使用 CSS。

第 6 章介绍 XML 作为数据源对象的相关技术,数据岛的概念、数据绑定的方法、支持 DSO 的 HTML 标记、XML 文件的嵌入和 XML 数据源对象的综合运用。

第 7 章讲述 XSL 的基本概念,着重讲解 XSLT 转换语法、XSLT 文档的编写、模板的规则及应用模板、XSLT 的转换执行过程、实现 XSLT 的方式。

第 8 章讲述文档对象模型的概念和基本知识,以实例形式说明客户端和服务端加载 XML 的方法和出错处理,并介绍了 XML 的节点接口。

第 9 章介绍 Microsoft .NET Framework 的基本概念,.NET 进行处理 XML 文档的方式,使用数据集存储和处理 XML 文档,综合应用 .NET 技术建设一个网上书店。

第 1~5 章属于 XML 基础知识,满足一般的简单的应用;第 6~8 章属于中级内容,实现对 XML 的较全面掌握和系统应用;第 9 章属于高级应用入门。在教学过程中,各学校可以根据课时的多少取舍相关内容。

全书力求使学生能够把 XML 与网页设计技术结合起来,特别是与数据库技术相结合,灵活地进行信息的描述、交换和显示。各章节语法介绍简明易懂,避免了复杂的理论介绍;

对实例的代码作了必要的注解，全部实例源代码可在机械工业出版社网站上（<http://www.cmpbook.com>）下载。

本书由丁跃潮、叶文来、陈杰编写。其中，第 1、6、8 章由丁跃潮编写，第 2、3、5 章由陈杰编写，第 4、7、9 章由叶文来编写，全书由丁跃潮统稿。

书中遗漏或错误之处，敬请读者批评指正。

编者

目 录

编者的话

第 1 章 引论	1
1.1 标记语言的发展	1
1.1.1 SGML 的起源	1
1.1.2 HTML 的出现	2
1.1.3 XML 的诞生	3
1.2 XML 的特点和应用前景	6
1.2.1 XML 的特点	6
1.2.2 XML 的应用前景	9
1.3 XML 文档的数据结构	11
1.3.1 层状结构数据模型	11
1.3.2 XML 语法基本点	12
1.3.3 XML 文档的结构	13
1.4 XML 开发和应用环境	15
1.4.1 XML 编辑工具	15
1.4.2 XML 验证工具	17
1.4.3 XML 解析器	17
1.4.4 XML 浏览工具	18
1.5 简单引例的建立	18
1.6 小结	21
1.7 习题	21
第 2 章 XML 基础	23
2.1 HTML 语法概要	23
2.1.1 HTML 基本语法	23
2.1.2 HTML 表格	26
2.1.3 HTML 表单	28
2.2 XML 标记语法	32
2.2.1 非空标记	32
2.2.2 空标记	33
2.2.3 XML 标记命名规则	34
2.3 XML 文档语法	34
2.3.1 XML 声明	34
2.3.2 处理指令	35

2.3.3 XML 元素	35
2.3.4 XML 的属性	36
2.3.5 实体引用与 CDATA 标记	37
2.3.6 XML 的注释	38
2.4 XML 的相关技术	39
2.4.1 CSS	39
2.4.2 DTD 和 XML Schema	39
2.4.3 XML DOM	39
2.4.4 XML XSLT	40
2.4.5 XLink、XPointer 和 XPath	40
2.5 XML 文档内的名称空间	40
2.5.1 名称空间的产生	40
2.5.2 名称空间的使用	41
2.5.3 名称空间的作用域	42
2.6 XML 文档的编辑	42
2.7 小结	44
2.8 习题	44
第 3 章 XML 文档类型定义	46
3.1 DTD 的作用	46
3.2 DTD 元素	46
3.2.1 DTD 元素的定义	47
3.2.2 元素出现顺序和子元素	48
3.3 DTD 属性	48
3.3.1 DTD 属性语法	48
3.3.2 属性类型	49
3.3.3 属性的默认形态	52
3.4 DTD 实体	52
3.4.1 内部实体	52
3.4.2 外部实体	54
3.4.3 参数型实体	55
3.5 使用 DTD	55
3.5.1 内部 DTD 的使用	56
3.5.2 外部 DTD 的使用	56
3.5.3 内外 DTD 结合的使用	57
3.6 实例	58
3.7 小结	62
3.8 习题	62
第 4 章 XML Schema 结构	64
4.1 Schema 概述	64

4.2	Schema 的格式和使用	65
4.2.1	简单实例	65
4.2.2	Schema 的使用	67
4.3	Schema 元素声明	69
4.3.1	Schema 根元素	69
4.3.2	element 元素	69
4.3.3	设置默认值和固定值	70
4.3.4	引用元素和替代	72
4.4	Schema 属性声明	73
4.4.1	创建属性	73
4.4.2	为属性指派类型	74
4.4.3	属性的默认值和固定值	74
4.5	Schema 的数据类型	75
4.5.1	简单类型	75
4.5.2	复杂类型	80
4.6	Schema 的名称空间	84
4.6.1	命名冲突	84
4.6.2	使用前缀解决命名冲突问题	84
4.6.3	使用名称空间	85
4.6.4	URI、URL 和 URN	86
4.6.5	名称空间的作用域	86
4.6.6	在 XSDL 中使用的名称空间	87
4.7	使用 XMLSpy 建立 Schema 文档	88
4.7.1	建立根节点	89
4.7.2	为根节点 book 添加子节点及属性	90
4.7.3	为 XML 文档指定 Schema	91
4.8	小结	91
4.9	习题	92
第 5 章	使用 CSS 格式化 XML	94
5.1	CSS 简介	94
5.2	CSS 的基本语法	95
5.3	CSS 的使用	95
5.3.1	内部 CSS 的使用	95
5.3.2	外部 CSS 的使用	97
5.3.3	内外结合 CSS 的使用	98
5.4	常用的 CSS 属性设置	99
5.4.1	字体属性	99
5.4.2	颜色与图像属性	100
5.4.3	文本属性	101

5.4.4	方框边界属性	102
5.5	CSS 选择器	106
5.6	CSS2 新增功能	108
5.6.1	@规则	108
5.6.2	新增属性	109
5.6.3	其他伪类	109
5.7	实例	110
5.8	小结	113
5.9	习题	113
第 6 章	XML 数据源对象	114
6.1	数据岛、XML 数据源对象与数据绑定	114
6.1.1	数据岛和 XML 数据源对象	114
6.1.2	数据绑定	115
6.1.3	XML 与 HTML 结合的优势	118
6.2	支持数据绑定的 HTML 元素	118
6.3	将 XML 文件数据嵌入 HTML	119
6.3.1	内部嵌入 XML	120
6.3.2	使用外部的 XML 文件	121
6.4	结合脚本语言的 XML	124
6.4.1	HTML 中的 XML 数据岛记录的编辑与添加	124
6.4.2	HTML 中的 XML 数据岛记录集的页面管理	128
6.5	综合应用实例	131
6.6	小结	134
6.7	习题	134
第 7 章	XSL 转换	135
7.1	XSL 概述	135
7.1.1	XSLT 的使用	135
7.1.2	XSLT 和 CSS 的比较	136
7.1.3	XPath 的含义	136
7.1.4	Internet Explorer MSXML	136
7.2	XSLT 格式和使用	137
7.2.1	简单的 XSLT 示例	137
7.2.2	XSL 中的元素分析	138
7.2.3	XSL 的执行过程	140
7.3	XSLT 转换语言	142
7.3.1	Stylesheet 元素	142
7.3.2	template 和 apply-templates 元素	142
7.3.3	XPath	144
7.3.4	模板体	146

7.4	实现 XSLT 的方式	154
7.4.1	在客户端的实现	154
7.4.2	在服务器端的实现	156
7.5	使用 XMLSpy 建立 XSLT 文档	156
7.6	小结	157
7.7	习题	158
第 8 章	XML 文档对象模型	160
8.1	DOM 基础知识	160
8.1.1	DOM 的概念和组成	160
8.1.2	DOM 结构模型	161
8.1.3	创建 Document 对象	163
8.2	客户端加载 XML	164
8.2.1	加载外部 XML 文件	164
8.2.2	加载 XML 文档字符串	169
8.2.3	加载 XML 文件及其转换文件	171
8.3	服务器端加载 XML	172
8.3.1	利用 ASP 加载 XML 文件	172
8.3.2	利用 ASP 文件加载 XML 文件及其转换文件	174
8.4	ParseError 对象	174
8.4.1	ParseError 属性	174
8.4.2	出现错误时的处理	175
8.5	DOM 的节点接口	175
8.5.1	新增 XML 节点	176
8.5.2	删除 XML 元素和属性	179
8.5.3	访问和修改 XML DOM 节点	180
8.6	小结	182
8.7	习题	182
第 9 章	XML 与 .NET 构建小型 Web 应用	184
9.1	Microsoft .NET Framework 介绍	184
9.1.1	NET 构架	185
9.1.2	从 MSXML 到 .NET 的 XML	185
9.1.3	XML 分析模式	186
9.2	Visual Studio.NET 集成开发环境	186
9.2.1	VS.NET 主界面	187
9.2.2	XML 编辑器	187
9.3	用 .NET 处理 XML 文档	189
9.3.1	.NET 下 XML 相关类简介	189
9.3.2	用 XmlTextReader 类读取和解析 XML 文档	191
9.3.3	用 XmlTextWriter 类书写 XML 文档	194

9.3.4 存储和处理 XML 文档	196
9.3.5 XML 与数据库协调工作	200
9.4 利用 XML 和 .NET 构建动态网站	203
9.4.1 网上书店项目概述	203
9.4.2 网站后台数据文件	204
9.4.3 网站站点架构	206
9.4.4 具体实现	206
9.5 小结	211
9.6 习题	212
参考文献	213

第1章 引 论

学习目标

- 了解标记语言的发展和现状
- 掌握 XML 的特点
- 理解 XML 的应用前景
- 掌握 XML 文档的数据结构和基本语法
- 了解常用的 XML 开发工具，学会使用一种综合性工具

信息革命和信息技术的发展，改变了人们的生产、生活方式，促进了社会的信息化发展。随着网络技术的发展，网上信息交流已逐渐成为人们交流思想、感情、成果等的主要渠道，互联网成为与报纸、广播、电视并驾齐驱甚至更为优越的第四媒体。在网络应用中，人们逐渐发现 HTML 这种语言已经远远不能满足网络潜在应用能力的需要，这种静态网页受到了挑战，带来 JavaScript、VBScript、ASP、JSP、PHP、ASP.NET 等 Web 编程技术的不断发展。

现有的标识语言 HTML 和与其配合的脚本语言都是有限的。可扩展性标识语言——XML 的出现，把网络表达语言推进了一大步，使得标记语言可以由编程人员按需要进行扩展，其扩展能力是无限的，这就给 Web 编程带来了革命性的进展。

本章介绍 XML 这一全新的标识语言的引导知识，包括标记语言的发展、XML 的特点、应用前景、数据结构、开发和应用环境。最后介绍简单引例的建立。

1.1 标记语言的发展

XML 是标记语言发展的一个新阶段。为了更好地理解 XML，需要回顾一下标记语言的发展过程。标记语言是随着文档处理和信息交换技术的发展而进化的。

1.1.1 SGML 的起源

SGML 的全称是“Standard Generalized Markup Language”（标准通用标识语言）。它是国际上定义电子文档结构和描述内容的标准。

早在 1969 年，IBM 公司的工作人员 Ed Mosher、Ray Lorie 和 Charles F. Goldfarb 就开发了一种文档描述语言 GML，用来解决不同系统中文档格式不同的问题。GML 是 IBM 许多文档系统的基础，是一种自参考的语言，可以用来描述任何其他语言的语法和词汇，能够置标任何数据集合的结构。1974 年，Goldfarb 提出了在 GML 中实现的语法合法性分析器的概念。经过多年的发展，1984 年，国际标准化协会（ISO）开始对这项提案进行讨论，终于在 1986 年演变成一个国际标准（ISO8897），并被称为 SGML。

SGML 实际上是一种通用的文档结构描述的符号化语言，主要用来定义文档模型的逻辑和物理结构。一个 SGML 语言文件由三部分组成，即语法定义、文档类型定义（Document Type

Definition, DTD) 和文档实例。其中, 语法定义部分定义了文档类型定义和文档实例的语法结构; 文档类型定义部分定义了文档实例的结构和组成结构的元素类型; 文档实例是 SGML 语言程序的主体部分。

SGML 是很多大型组织, 如电信、飞机、汽车公司、化工企业和军队的文档标准。它是与平台无关的、结构化的、可扩展的语言, 这些特点使它在很多公司受到欢迎, 并被用来创建、处理和发布大量的文本信息。

1.1.2 HTML 的出现

1989 年, 在欧洲量子实验室 (CERN) 的研究人员 Tim Berners-lee 开发了基于 SGML 的非常简单的超文本版本, 被称为 HTML (Hyper Text Markup Language), 它是 SGML 的一个简化子集。HTML 继承了 SGML 的许多重要的特点, 如结构化、实现独立和可描述性, 但是同时它也存在很多缺陷, 例如, 它只能使用固定和有限的标记, 而且它只侧重于对内容的显示。当计算机的应用水平大幅度地提高的时候, 人们对简单的 HTML 已不再满足。因此, 开始了对其进行修改、升级。就这样, HTML 从 1.0 一直升到 4.0 的版本, 扩展了许多命令, 使 HTML 能够解决更多的问题。同时, DHTML、VHTML、SHTML 等也快速发展起来。作为 WWW (World Wide Web) 的一个组成部分, HTML 以简单精练的语法、极易掌握的通用性与易学性, 使 Web 网页可以吸引广大网民, 使互联网得以普及和发展。

然而, HTML 本身有很多缺陷: 首先, HTML 是把数据和显示格式放在一起存放的, 因此不能只使用数据而不需要格式, 分离这些数据和格式较为困难。其次, HTML 对超级链接支持不足, 它属于单点的链接, 功能上有一些限制。再次, HTML 缺乏空间立体描述; 处理图形、图像、音频、视频等多媒体能力较弱, 图文混排功能简单, 不能表示多种媒体的同步关系, 这就影响了 HTML 的大规模应用, 特别是在复杂的多媒体数据处理上的应用。最后, HTML 的标记有限, 不能由用户自己扩展自己的标记。下面来看 HTML 的一个简单例子。

【例 1-1】简单的 HTML 网页, 程序代码为 code1_1.htm。

```
<html>
<head>
<title>简单的 HTML</title>
</head>
<body>
<h1 align="center">XML 专题系列讲座</h1>
<table border="1" width="97%" id="table1">
  <tr>
    <td width="38">序次</td>
    <td width="96">内容</td>
    <td width="54">主讲人</td>
    <td width="106">时间</td>
    <td>地点</td>
  </tr>
  <tr>
    <td width="38">1</td>
    <td width="96">XML 导引</td>
```

```

        <td width="54">丁跃潮</td>
        <td width="106">2005 年 6 月 4 日</td>
        <td>计算机学院教学楼 317 教室</td>
    </tr>
    <tr>
        <td width="38">2</td>
        <td width="96">XML 文档验证</td>
        <td width="54">叶文来</td>
        <td width="106">2005 年 6 月 9 日</td>
        <td>理学院教学楼 204 教室</td>
    </tr>
    <tr>
        <td width="38">3</td>
        <td width="96">XML 文档显示</td>
        <td width="54">陈杰</td>
        <td width="106">2005 年 6 月 13 日</td>
        <td>计算机学院教学楼 108 教室</td>
    </tr>
    <tr>
        <td width="38">4</td>
        <td width="96">XML 网页制作</td>
        <td width="54">叶文来</td>
        <td width="106">2005 年 6 月 17 日</td>
        <td>信息工程学院教学楼 315 教室</td>
    </tr>
    <tr>
        <td width="38">5</td>
        <td width="96">XML 应用实例</td>
        <td width="54">陈杰</td>
        <td width="106">2005 年 6 月 21 日</td>
        <td>信息工程学院教学楼 315 教室</td>
    </tr>
</table>
<ul>
    <li>信息发布日期: 2005 年 5 月 28 日</li>
    <li>联系人: 张小英、梅玉婷</li>
    <li>电子邮件: <a href="mailto:xyzhang@jmu.edu.cn">xyzhang@jmu.edu.cn</a></li>
    <li>网址: <a href="http://www.jmu.edu.cn/cs">http://www.jmu.edu.cn/cs</a></li>
</ul>
</body>
</html>

```

1.1.3 XML 的诞生

随着网上数据的增多, HTML 存在的缺点就变得不可忽略。因此, W3C (World Wide Web Consortium, 万维网络联盟) 提供了 HTML 的几个扩展来解决这些问题。后来, W3C 的成