



21 世纪 高职高专创新精品规划教材

# XML网页技术实用教程

主 编 余以胜

副主编 巫志勇 黄华林

创新  
精品

“教、学、做”一体化，强化能力培养  
“工学结合”原则，提高社会实践能力  
“案例教学”方法，增强可读性和可操作性



中国水利水电出版社  
www.waterpub.com.cn

21 世纪高职高专创新精品规划教材

# XML 网页技术实用教程

主 编 余以胜

副主编 巫志勇 黄华林



中国水利水电出版社  
www.waterpub.com.cn

## 内 容 提 要

本书采用项目化教学的理念设计教学大纲,并组织全书内容。全书共为7个项目,使用一个具体案例网上数字图书馆贯穿全部项目,每个项目都有相应的工作任务,为了完成这些任务,又将项目划分为若干个模块,每个模块都有精心设计的细分子任务。工作任务涵盖了XML文档基础,用CSS制作XML网页,XML和HTML的整合,显示XML的XSL样式技术,如何设计DTD,使用Schema定义文件格式,以及ASP.NET的XML网页开发等知识点,涉及XML网页设计的各个方面的知识。

本书内容丰富、概念清晰、深入浅出;在内容编排上,循序渐进、明晰易懂。为更好培养学生的职业能力,以真实工作任务及其工作过程为依据,设计了一个网上数字图书馆项目,进行教学内容的整合与优化。通过指导读者完成一系列的实际工作任务来达到课程的教学目标,重点培养读者解决实际问题的能力。既可以学习巩固、加深理解与掌握知识,也可以同步训练、强化读者的实际动手能力。

本书重在实践,实用性强,既适合作为高职高专院校、应用型本科院校等计算机相关专业的教材,也适合网页制作爱好者学习网页设计高端技术的自学参考书,也可以作为各类IT培训班学习XML入门技术的培训教材。

本书电子教案可以从中国水利水电出版社网站免费下载,网址为:  
<http://www.waterpub.com.cn/softdown/>。

## 图书在版编目(CIP)数据

XML网页技术实用教程/余以胜主编. —北京:中国水利水电出版社,2008

21世纪高职高专创新精品规划教材

ISBN 978-7-5084-5910-3

I. X… II. 余… III. 可扩充语言,XML—主页制作—程序设计—高等学校:技术学校—教材 IV.

TP312 TP393.092

中国版本图书馆CIP数据核字(2008)第145410号

|       |                                                                                                                                                                                                                                                                                                     |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 书 名   | 21世纪高职高专创新精品规划教材<br>XML网页技术实用教程                                                                                                                                                                                                                                                                     |
| 作 者   | 主 编 余以胜<br>副主编 巫志勇 黄华林                                                                                                                                                                                                                                                                              |
| 出版 发行 | 中国水利水电出版社(北京市三里河路6号 100044)<br>网址: <a href="http://www.waterpub.com.cn">www.waterpub.com.cn</a><br>E-mail: <a href="mailto:mchannel@263.net">mchannel@263.net</a> (万水)<br><a href="mailto:sales@waterpub.com.cn">sales@waterpub.com.cn</a><br>电话: (010) 63202266 (总机)、68367658 (营销中心)、82562819 (万水) |
| 经 售   | 全国各地新华书店和相关出版物销售网点                                                                                                                                                                                                                                                                                  |
| 排 版   | 北京万水电子信息有限公司                                                                                                                                                                                                                                                                                        |
| 印 刷   | 北京市天竺颖华印刷厂                                                                                                                                                                                                                                                                                          |
| 规 格   | 184mm×260mm 16开本 18.25印张 451千字                                                                                                                                                                                                                                                                      |
| 版 次   | 2008年9月第1版 2008年9月第1次印刷                                                                                                                                                                                                                                                                             |
| 印 数   | 0001—4000册                                                                                                                                                                                                                                                                                          |
| 定 价   | 29.00元                                                                                                                                                                                                                                                                                              |

凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换

版权所有·侵权必究

## 序

近年来,我国高等职业教育蓬勃发展,为现代化建设培养了大量高素质技能型专门人才,对高等教育大众化作出了重要贡献,顺应了人民群众接受高等教育的强烈需求。高等职业教育作为高等教育发展中的一个类型,肩负着培养面向生产、建设、服务和管理第一线需要的高技能人才的使命,在我国加快推进社会主义现代化建设进程中具有不可替代的作用。随着我国走新型工业化道路、建设社会主义新农村和创新型国家对高技能人才要求的不断提高,高等职业教育既面临着极好的发展机遇,也面临着严峻的挑战。

教材建设是整个高职高专院校教育教学工作的重要组成部分,高质量的教材是培养高质量人才的基本保证,高职高专教材作为体现高职高专教育特色的知识载体和教学的基本工具,直接关系到高职高专教育能否为一线岗位培养符合要求的高技术性人才。中国水利水电出版社本着为高校教育服务,为师生提供高品质教材的原则,按照教育部《关于全面提高高等职业教育教学质量的若干意见》的要求,在全国数百所高职高专院校中遴选了一批具有丰富的教学经验、较高的工程实践能力的学科带头人和骨干教师,成立了高职高专教材建设编委会。编委会成员经过几个月的广泛调研,了解各高职院校教学改革和企业对人才需求的情况,探讨、研究课程体系建设和课程设置,达成共识,组织编写了本套“21世纪高职高专创新精品规划教材”。

本套教材的特点如下:

1. 面向高职高专教育,将专业培养目标分解落实于各门课程的技术应用能力要求,建立课程的技术、技能体系,将理论知识贯穿于其中,并融“教、学、做”为一体,强化学生的能力培养。

2. 理论知识的讲解以基础知识和基本理论“必需、够用”为原则,在保证达到高等教育水平的基础上,注重基本概念和基本方法讲解的科学性、准确性和正确性,把重点放在概念、方法和结论的阐释和实际应用上,推导过程力求简洁明了。

3. 在教材中按照技术、技能要求的难易和熟练程度,选择恰当的训练形式和内容,形成训练体系;确定实训项目,并将实训内容体现在教材中。对于单独设置实训的课程,我们将实训分成基础实训和综合实训两个部分。综合实训中重点体现了工学结合的原则,提高学生的社会实践能力。

4. 在编写方式上引入案例教学和启发式教学方法,采用以实际应用引出的问题为背景来设计和组织内容,增强了教材的可读性和可操作性,激发学生的学习兴趣,使知识点更容易理解掌握,从而使学生能够真正地掌握相关技术,为以后的就业打好基础。

5. 教材内容力求体现经济社会发展对应用技术的新要求和新趋势,将新兴的高新技术、复合技术等引进教材,并在教材中提出了一些引导技术发展的新问题,以期引起思考和讨论,有利于培养学生技术应用中的创新精神和能力。

6. 大部分教材都配有电子教案和相关教学资源,以使教材向多元化、多媒体化发展,满足广大教师教学工作的需要。电子教案使用 PowerPoint 制作,教师可根据授课情况任意修改。相关教案和资源可以从中国水利水电出版社网站 [www.waterpub.com.cn](http://www.waterpub.com.cn) 下载。

本套教材凝聚了众多奋斗在高等职业教育教学、科研第一线的教师和科研人员多年的教学经验和智慧，教材内容选取新颖、实用，层次清晰，结构合理，概念清晰，通俗易懂，可读性和实用性强。本套教材适用于高职高专院校，也可作为社会各类培训班用书和自学参考用书。

我们期待广大读者对本套教材提出宝贵意见和建议，以便进一步修订，使该套教材不断完善。

21 世纪高职高专创新精品规划教材编委会

2008 年 4 月

# 前 言

目前,高等职业技术教育发展的趋势是走工学结合的道路,以工作过程导向的课程改革是将工作岗位技能培养导入课堂教学的重要举措。以项目化教学模式组织教学内容、实施面向工作过程导向的课程改革是计算机相关专业教学改革、课程建设的发展方向。

计算机相关专业的网站建设与网页设计课程体系中,XML技术的应用是网页制作的高端技术领域,当前涉及XML技术的书籍或教材有很多,如何将XML技术与网页制作相结合,如何将这部分技术内容纳入教学并使其规范化、项目化,是我们教学改革的尝试。

本书要求读者具有HTML语法基础,具有基本的Visual Basic编程基础,按照程序代码编写的基本方式学习XML网页制作技术。全书共分7个项目,子任务细分为18个模块,对每个项目,有明确的知识目标、能力目标和具体的教学任务,模块开始前有预备知识,中间有知识要点分解,项目完成后有知识与能力拓展,以对项目目标扩充,举一反三。具体内容安排如下:

项目1为图书信息XML文档的生成,主要讲XML文档的基本操作,分为两个模块:创建图书信息XML文档和图书信息XML文档的语法格式;项目2为读者借阅单网页设计,主要讲用CSS制作XML网页,分为三个模块:读者借阅单网页制作、书目单网页的修饰制作和读者借阅单首页制作;项目3为读者俱乐部网页设计,主要讲HTML和XML相结合制作网页,分为三个模块:读者俱乐部欢迎页面制作、读者会员注册页面制作和新书介绍页面制作;项目4为个人网上图书馆网页设计,主要讲用XSL实现个性化显示,分为三个模块:个人网上图书馆首页制作、借阅书籍分类展示页面制作和借阅书籍介绍页面制作;项目5为网上订购图书,主要讲设计一个DTD文件,分为两个模块:网上订书初级DTD文件和网上订书增强DTD文件;项目6为图书管理员档案,主要讲设计XML Schema文件,分为两个模块:简单的档案样本制作和完整的档案样本制作;项目7为读者在线咨询,主要讲在ASP.NET环境下对XML网页的开发,分为两个模块:设计读者在线咨询网页和显示XML文档中读者咨询信息。

本书由余以胜任主编、巫志勇和黄华林任副主编。其中本书的项目1、项目2和项目7由余以胜编写,项目3、项目4由巫志勇编写,项目5、项目6由黄华林编写,全书由余以胜组织策划并负责统稿。

由于作者水平有限,XML技术和Web网页技术发展迅速,本书的覆盖面较广,书中难免会有错误、疏漏和不妥之处,敬请专家和读者批评指正。

编者

2008年7月

# 目 录

序

前言

|                                         |     |
|-----------------------------------------|-----|
| 项目 1 图书信息 XML 文档生成——XML 文档的基本操作 .....   | 1   |
| 预备知识 .....                              | 1   |
| 模块 1 创建图书信息 XML 文档 .....                | 16  |
| 模块 2 图书信息 XML 文档的语法格式 .....             | 26  |
| 知识与能力拓展 .....                           | 38  |
| 项目 2 读者借阅单网页设计——用 CSS 制作 XML 网页 .....   | 40  |
| 预备知识 .....                              | 40  |
| 模块 1 读者借阅书目单网页制作 .....                  | 43  |
| 模块 2 书目单网页的修饰制作 .....                   | 57  |
| 模块 3 读者借阅书目单首页制作 .....                  | 72  |
| 知识与能力拓展 .....                           | 81  |
| 项目 3 读者俱乐部网页设计——用 HTML 和 XML 制作网页 ..... | 82  |
| 预备知识 .....                              | 82  |
| 模块 1 读者俱乐部欢迎页面制作 .....                  | 84  |
| 模块 2 读者会员注册页面制作 .....                   | 97  |
| 模块 3 新书介绍页面制作 .....                     | 108 |
| 知识与能力拓展 .....                           | 118 |
| 项目 4 个人网上图书馆网页设计——用 XSL 实现个性化显示 .....   | 119 |
| 预备知识 .....                              | 119 |
| 模块 1 个人网上图书馆首页制作 .....                  | 123 |
| 模块 2 借阅书籍分类展示页面制作 .....                 | 135 |
| 模块 3 借阅书籍介绍页面制作 .....                   | 149 |
| 知识与能力拓展 .....                           | 164 |
| 项目 5 网上订购图书——设计一个 DTD 文件 .....          | 166 |
| 预备知识 .....                              | 166 |
| 模块 1 网上订书初级 DTD 文件 .....                | 168 |
| 模块 2 网上订书增强 DTD 文件 .....                | 180 |
| 知识与能力拓展 .....                           | 207 |
| 项目 6 图书管理员档案——设计 XML Schema 文件 .....    | 208 |
| 预备知识 .....                              | 208 |
| 模块 1 简单的档案样本制作 .....                    | 211 |
| 模块 2 完整的档案样本制作 .....                    | 222 |
| 知识与能力拓展 .....                           | 243 |

项目 7 读者在线咨询——ASP.NET 的 XML 网页开发..... 245

    预备知识 .....246

    模块 1 设计读者在线咨询网页.....248

    模块 2 显示 XML 文档中的读者咨询信息.....264

    模块 3 将读者咨询信息写入 XML 文档.....275

    知识与能力拓展 .....284

# 项目 1



## 图书信息 XML 文档生成

### ——XML 文档的基本操作

#### 一、知识目标

项目目标：本项目是学习 XML 的入门部分，在操作基础上介绍学习 XML 所必须具备的两方面知识：一方面，通过图书信息 XML 文档的制作，介绍对 XML 文档经常使用的基本操作；另一方面，通过图书信息 XML 文档的分解，介绍 XML 的基本语法知识。

教学目标：

- 掌握创建、表现一个 XML 文档的基本过程。
- 理解 XML 文档的基本结构。
- 理解并掌握 XML 的基本语法。
- 了解格式良好的 XML 文档的基本要求。

#### 二、能力目标

- 培养学生收集、处理信息，准备、加工 XML 文档资料的能力。
- 培养学生制作 XML 文档的能力。
- 培养学生阅读 XML 文档源码的能力。
- 培养学生优化 XML 文档格式的能力。
- 培养学生获取知识和应用知识的能力。

#### 三、教学任务

- 创建一个图书信息 XML 文档。
- 解读图书信息 XML 文档的语法格式。

## 预备知识

#### 一、标记语言及 XML 的出现

在标记语言没有出现之前，计算机中的数据在人们的印象中是极其神秘的二进制比特。

然而,标记语言的出现慢慢改变了这种认识。特别是近些年来,随着 WWW (World Wide Web) 的迅猛发展以及 HTML 的出现和广泛应用,数据变成了文本,变成了任何人都能速写的东西。而且,由于人们对数据的表现、数据的共享等多方面要求的不断提高,导致了 XML 的产生。作为新一代的网络发布语言,XML 也是一种标记语言,而且它与另外两种主要的标记语言 SGML 和 HTML 有着千丝万缕的联系。在这里,将简要地介绍 SGML、HTML 和 XML 概况以及它们的相互关系。

### (一) SGML 的产生和应用

1969 年,IBM 的研究人员 Ed Mosher、Ray Lorie 和 Charles F. Goldfarb 发明了第一种标记语言——GML (Generalized Markup Language, 通用标记语言)。GML 是一种自参考的语言,它可以用于标记任何数据集合的结构,同时它也是一种元语言 (Meta-language),即能够描述其他语言及其语法和词汇表的语言。1974 年,Goldfarb 提出了在 GML 中实现的语法合法性分析器的概念,即在处理一个文档之前,先读取该文档的类型定义,然后由此检查标记语言使用的正确性,而不必实际处理这个文档。此后,GML 发展成了 SGML (Standard Generalized Markup Language, 标准通用标记语言)并在以 Goldfarb 为首的一批研究人员、程序员和用户的共同努力下,SGML 得到不断的发展、完善和应用。1986 年,SGML 被国际标准化组织 (ISO) 接受为国际性的数据存储和交换的标准,收录在 ISO-8897 中。



提示

标记是一种传输元数据 (即关于数据集本身的信息) 的方法。标记语言使用文字串或标记来界定和描述这些数据。标记文档的过程实际上就是标识出文档的某些部分具有特殊含义的过程。

尽管从 GML 产生到 SGML 成为国际标准走过了相当漫长的历程,但当 GML 发展成 SGML 后就一直在市场上处于领先地位。SGML 被美国政府及其合同商、大型制造公司、信息技术发布商广泛采用。出版商也经常使用 SGML 制作各类纸张文档,如书籍、报告、参考手册等,然后将这些 SGML 文档转换为适当的格式交给排版者和印刷者处理。SGML 还被用来将技术规范应用于生产制造。对于 Web 语言的发展,SGML 更是有着特殊的贡献。无论是现在普遍应用于网络中的 HTML 还是未来网络发布的主流语言——XML,都起源于 SGML。因此,考虑到 Web 对人类商业和通信行为产生的巨大影响,在技术发展史上,GML 的问世可称得上是一件极其重要的事件。

但问题是,SGML 为了适应各种复杂应用的需求,在其成为国际标准时,已经变得异常庞大和错综复杂。它的复杂性及其实现所需要的大量资金限制了它的用户范围,除了为数不多的大型企业和组织外,大多数商业用户和个人用户无法享受此项技术所带来的益处。因此,尽管 SGML 已经产生和发展了近 20 年,但其应用并没有得到普及,其影响力也远不及后来的 HTML。

### (二) HTML 的产生和应用

1989 年,一位名叫 Tim BernersLee 的研究人员提出,用超链接文本文档实现在欧洲原子能研究所 (CERN) 内部的信息共享。在同事 Anders Berglund 的建议下,Tim 采用了 SGML 形式的语法。他们从 SGML 的一种简单文档类型入手,开发出了一种超文本格式,这就是现

在的 HTML (Hypertext Markup Language, 超文本标记语言)。

尽管与发展了 20 年的 SGML 相比, HTML 的开发显得有些匆忙, 但它却非常有效。Tim 把他的超文本系统称为 World Wide Web (简称 WWW 或 Web), 现在它是已有的超文本系统中最多样化的、最受欢迎的一种 Internet 服务。HTML 的简洁性和 Web 规范使得世界各地的程序员能够迅速地创建用于 Web 的系统和工具。现在人们普遍认为, 简单易用是 HTML 获得巨大成功的重要原因。

HTML 在抛弃 SGML 复杂性的同时, 也从 SGML 那里继承了很多重要特性。除了少数例外情况, HTML 的元素类型是一般化的、描述性的, 用它编制的文档能够在屏幕上显示, 也能在图形界面中显示, 甚至可以通过扬声器来告诉视力障碍人士文档的内容。HTML 与 SGML 一样, 使用简单的“<”和“>”来标识标记。这就意味着可以使用一般的文本编辑器或字处理器来创建 HTML 文档, 而且得到的 HTML 文档与现有的几乎所有计算机系统兼容。

HTML 的简易性和平台无关性使得它的应用遍及 WWW 的各个角落。但是, 随着网络应用的不断发展, HTML 简单易用的优点正变成它的致命缺点, 已经越来越不适应未来网络发展的需要, 具体表现在下面所列举的几个方面。

### 1. HTML 扩展困难

HTML 只采用了一部分 GML 思想, 即文档表示应该标准化, 所以它只使用了一个固定的元素类型集。这就使得 HTML 仅仅是符合 SGML 语法的一种固定格式的超文本标记语言, 它不可能扩展, 用户也不能在其中增加有意义的并能供他们使用的标记。

随着 Web 越来越流行, 许多人开始不满足于 HTML 固定的文档类型。浏览器销售商发现, 通过对 HTML 非兼容性的扩展可以获取更大的市场份额, 于是大量非标准的 HTML 扩展出现了。在这方面, 两大浏览器生产商 Netscape 公司和 Microsoft 公司就是典型的代表。他们发明了许多只能被自己的浏览器所识别的标记, 但这往往让用户和编程人员感到迷茫。非标准的扩展破坏了 Web 的交互操作性。

### 2. HTML 的交互性差

现在, 用户对网络应用的交互性要求越来越高, 但是使用 HTML 却并不容易做到。频繁地请求处理使 Web 服务器不堪重负, 尽管使用 DHTML 可以把一些处理交给浏览器完成, 但在这方面做得还远远不够。更为重要的是, 当想对同一数据以不同的视角进行观察时, HTML 的办法是请求服务器另外再传送一个页面, 这无疑更加重了服务器的负担。

把网络浏览器作为一个应用平台, 以 HTML 作为数据标准, 网络上的应用程序依赖服务器上 CGI (公共网关接口) 脚本来处理网页上的数据, 这样的结构是非常清晰的, 但 HTML 却使服务器一端要处理的数据量过大, 从而降低了网络速度和网络应用的处理效率。

### 3. HTML 的语义性差

HTML 最初是被设计用来标记文档以反映文档结构的, 同时它也包含一些用于显示格式的标记, 再加上超文本链接, HTML 在 Web 上获得了巨大的成功。HTML 在页面布局、外观方面很擅长, 但是却极度缺乏对内容, 也就是信息含义的表达能力。HTML 标记几乎全部用于设计网页。所以 HTML 主要是设计来做人机交流用的。

HTML 只是一种表达的技术, 它并不一定能提示 HTML 标记中所隐含的含义。举一个简单的例子, <H2>Apple</H2>这条代码在网络浏览器中有特定的表现, 但是 HTML 却没有告诉我们它到底是什么。尽管 Apple 只是一个英文单词, 但它在不同的环境之中可能会有不同的意

义,它可能是指 Apple 计算机公司,也可能是指一种水果,还可能是一个姓氏。但它在显示的页面上指的是什么,HTML 不能给我们明确的答复。

正是由于 HTML 没有考虑标记的语义性,因此在以下几个方面存在极大的不足。

- 电子数据交换 (EDI)

在 HTML 中,产品名和产品价格可能隐藏在文档的任何地方,没有任何特别之处,计算机无法将其从文档中识别出来,因此也就不可能对其进行更深入的处理。但是,下面将介绍的 XML 却可以将产品名和产品价格用类似于<ProductName>和<ProductPrice>的标记将它们标识出来,这样针对不同的应用就能理解它们的含义并进行相应的处理。XML 彻底弥补了商业规则和数据内容及结构密不可分的传统 EDI 的缺陷,从这个意义上说,XML 实际上带来的是一场 EDI 技术革命。

- 数据库

要在 HTML 中展示数据库中的数据,一般的过程是:通过某种手段从数据库取出数据,然后将它们写到用于展示的 HTML 文档中,加上各种标记,提交给用户。这样做的结果就是用户得到一些“死”数据。即使用户仅仅想以另外一种方式浏览数据,也必须通过重新发出请求才能实现。

- 搜索引擎

返回的结果太多了,这大概是每个使用过搜索引擎的人都有的体会。你需要足够的耐心和时间去检查哪些是你需要的结果。问题在于现在的搜索引擎只能检索单词的频率、文档的标题以及在某些情形下描述页面内容的元标记等,这是 HTML 的缺陷带来的限制。

- 单向的超链接

虽然 HTML 提供的超链接机制取得了巨大的成功,但这种机制的如下特征却注定了它的局限性:

- (1) 链接在它的某一端(源端)定义。

- (2) 超链接确定了它的另一端(目的端),虽然服务器可以有很大的自由度来寻找和创建它的目的端。

- (3) 用户只能从源端出去到目的端。

- (4) 超链接的效果由浏览器而不是由超链接本身来决定。

实际上,在早些的若干专用超媒体系统中,已经有一些更灵活的机制来表达更复杂的链接关系,如多个目的端、多向链接等,使用 XML 可以满足这样的需求。

#### 4. HTML 缺乏对双字节或多国文字的支持

HTML 缺乏对双字节或多国文字的支持,或者说支持不够。例如,在设计 HTML 网页时,经常发现中文信息在不同的平台下出现格式不齐的问题。XML 将使 Web 建立者创建真正的国际性站点变得比以前更容易。开发者不需要学习任何脚本语言就能使用 Unicode 在 XML 文档中生效,文档在用户的浏览器中用合适的字符集显示出来。



提示

Unicode 是一个国际接受的标准,用于描述世界上所有的字母、字型、字符和表意符,包括 ASCII 字符、ISO 拉丁字符、日本、韩国、中国、印度、希腊、阿拉伯和其他字符,甚至允许字符集的混合。本书将在后面对 Unicode 进行简要的介绍。

在 HTML 中,也不能使用中文标记,而在 XML 中标记可以由编写者自由定义,定义出的标记可以充分反映编写者希望的含义。因此在 XML 中可以通过使用中文标记来更准确地表达其意思。

### (三) XML 的出现

随着 Web 应用越来越广泛和深入,人们渐渐觉得 HTML 不够用了,HTML 过于简单的语法使它在表达复杂信息方面显得力不从心,再加上前面提到的 HTML 的缺陷,使得 HTML 似乎已经到了穷途末路的地步。尽管 HTML 推出了一个又一个新版本,并增加了脚本、表格、帧等表达功能,但始终满足不了人们不断提出的更高要求。同时,由于计算机技术近年来的飞速发展,使得实现比当初发明 HTML 时复杂得多的 Web 浏览器成为可能,所以开发一种新的 Web 语言的时机已经成熟。

SUN 公司的 Jon Bosak 和一些精心挑选的 SGML 结构信息专家向 W3C 提出了“网络上的 SGML”计划。W3C 支持并赞助了他们的工作,并于 1996 年 7 月正式成立了专门的 XML 工作组。该工作组的主席就是 Jon Bosak,技术负责是 James Clark。工作的早期遇到了较大的阻力,因为也存在反对 SGML 的人。一些制定 XML 标准的 W3C 代表甚至声称“网络上的 SGML”是不可能实现的。但是工作组并没有就此退缩,他们打算让 SGML 以全新的面目出现在网上,故他们为新一代的网络语言起了一个令人耳目一新的名字——XML (eXtensible Markup Language, 可扩展标记语言)。

工作组制定了一个宏伟的计划来展示 XML 的特色,这个计划的实施分为三部分:

- XML 的句法。
- XLL (可扩展链接语言): XML 的语义链接。
- XSL (可扩展类型语言): XML 的表现。

并且 XML 将按照如下的目标来设计:

- XML 应该在 Internet 上直接使用。
- XML 应该支持广泛的应用程序。
- XML 应该与 SGML 兼容。
- 处理 XML 文档的程序应当易于编写。
- XML 可选功能的数目应该保持为最小值,最好为 0。
- XML 对读者来说应该是可读的,并且逻辑清晰。
- XML 设计应当可以快速地准备好。
- XML 设计应该正式而简洁。
- XML 文档应该易于生成。
- XML 标记的精炼性不是十分重要。

这样就得到了一种固有标记很少,却又具有足够的冗余度可以基本实现 HTML 和 SGML 功能总和的标记语言,它在文档设计方面比 HTML 具有更大的灵活性,在耗费时间和资金方面比 SGML 具有更低的成本。

经过 XML 工作组的不断努力,1998 年 2 月 10 日, W3C 正式批准公布 XML1.0 标准。从此, XML 诞生了。

## 二、XML 的情况

庞大复杂限制了 SGML 的应用, 过于简单使 HTML 面临绝境, XML 正是吸取了两者的不足而产生和发展起来的。下面介绍 XML 与 SGML、HTML 的关系以及 XML 的概念和术语。

### (一) 什么是 XML

XML 是根据 SGML 制定出来, 其全称为 eXtensible Markup Language, 意为“可扩展标记语言”, 是用于标记电子文件使其具有结构性的语言。XML 是 SGML 的一个子集, 是一个精简的 SGML, 它将 SGML 的丰富功能与 HTML 的易学易用性结合到了 Web 的应用中。

### (二) XML 与 HTML 的比较

前面已经谈到了, XML 与 HTML 一样, 都是源自 SGML。但 XML 保留了 SGML 的大部分功能(约 80%), 使复杂程度降低了不少(约 20%)。由于 XML 从 SGML 那里继承了 SGML 最为人称道的特性之一——可扩展性, 因此, XML 具有 HTML 语言所欠缺的巨大的伸缩性与灵活性。XML 不再像 HTML 那样使用固定数量、固定名称的标记, 而是给了用户很大的自由度, 使他们可以定义不限数量的标记来描述文件中的任何数据元素, 从而突破了 HTML 固定标记集合的约束, 使文件的内容更丰富、更复杂并组成一个完整的信息体系。

HTML 只是 Web 显示数据的通用方法, 而 XML 提供了直接处理 Web 数据的通用方法; HTML 着重描述 Web 页面的显示格式, 而 XML 着重描述 Web 页面的内容。尽管在外观上, HTML 文档与 XML 文档相差不多, 但 XML 文档却有着 HTML 文档无法比拟的语义性(在 XML 文档中可以使用中文)。请看下面的例子。

例 1-1 用 HTML 表现图书信息的文档(BOOK.HTML)。

```
<html>
  <head> <title> 用 HTML 文档表现的图书信息</title> </head>
  <body>
    <h3 align="center">使用 HTML 的文档<HR/> </h3>
    <table border="1" cellpadding="5" align="center">
      <tr>
        <th>书籍名称</th>
        <th>作者</th>
        <th>出版商</th>
        <th>价格</th>
      </tr>
      <tr align="center">
        <td>数字图像处理学</td>
        <td>陈秋琦</td>
        <td>电子工业出版社</td>
        <td>49.00</td>
      </tr>
      <tr align="center">
        <td>C++ Builder 开发技术</td>
```

```

<td>李东,王宏</td>
<td>水利水电出版社</td>
<td>35.00 </td>
</tr>
<tr align="center">
<td>VC++编程技术</td>
<td>高强</td>
<td>人民邮电出版社</td>
<td>35.00 </td>
</tr>
</table>
</body>
</html>

```

以 BOOK.HTML 保存文档后,以 IE 对其浏览,效果如图 1-1 所示。

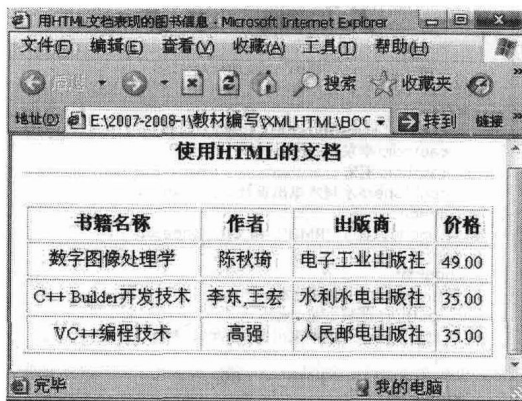


图 1-1 图书信息的 HTML 文档浏览效果

作为对比,用 XML 来设计包含相同信息的文档,具体代码如下例所示。

```

<?xml version="1.0" encoding="gb2312" ?>
<!-- 图书信息存放在"BOOK.XML"文件中 -->
<booklibrary>
<book>
<title> 数字图像处理学 </title>
<author> 陈秋琦 </author>
<publisher> 电子工业出版社 </publisher>
<price moneytype="RMB"> 49.00 </price>
</book>
<book>
<title> C++ Builder 开发技术 </title>
<author> 李东 </author>
<author> 王宏 </author>
<publisher> 水利水电出版社 </publisher>
<price moneytype="RMB"> 40.00 </price>
</book>
<book>

```

```
<title> VC++编程技术 </title>
<author> 高强 </author>
<publisher> 人民邮电出版社 </publisher>
<price moneytype="RMB"> 35.00 </price>
</book>
</booklibrary>
```

以文件名 BOOK.XML 保存文档后，用 IE 浏览其效果如图 1-2 所示。

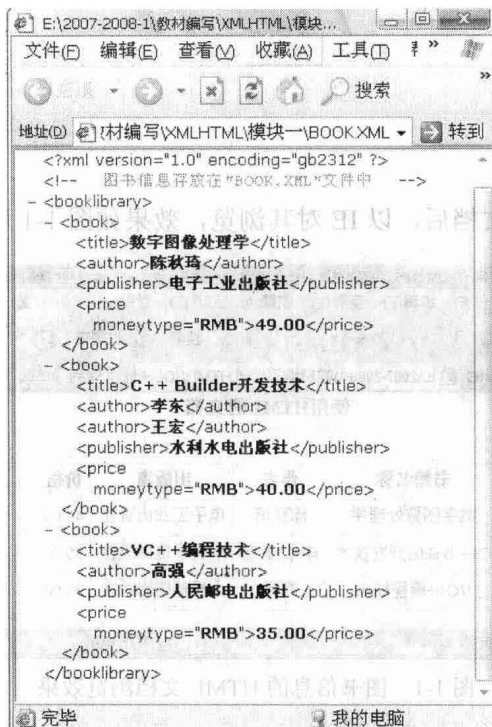


图 1-2 图书信息的 XML 文档浏览效果

通过上述两个例子的对比，可以发现对 HTML 文档的代码任意抽取一条，很难明白其中的含义，但通过 IE 浏览可知，其含义在其表头中指示；而对 XML 文档，任意抽取一条，其含义则由其本身可知。当然，从图 1-1 与图 1-2 的比较中，可能觉得 XML 文档在浏览器中没有 HTML 文档那样一目了然。其实这只不过是因为还没有为这个 XML 文档设计显示样式，也就是说，现在看到的仅仅是 XML 源文档的样子，在这个文档的基础上，可以通过设计不同的样式来获得不同的显示效果。也就是说，XML 实现了文档内容与表现形式的分离。这对于 HTML 来说同样是一件不可能的任务。

通过对比，开始感到 XML 的优越性了吧。XML 可以让信息提供者根据需要，自行定义标记名称。实际上，XML 不但允许设计者定义标记名，而且允许定义属性名，还可以包含描述法，从而使 XML 文档的结构可以复杂到任意程度。XML 主要有三个要素：Schema（模式）、XSL（eXtensible Stylesheet Language，可扩展样式语言）和 XLL（eXtensible Link Language，可扩展链接语言）。其中 Schema 规定了 XML 文档的逻辑结构，定义了 XML 文件中的元素、元素的属性以及元素和元素属性之间的关系，它可以帮助 XML 的分析程序校验 XML 文档标记的合法性；XSL 是用于规定 XML 文档样式的语言，它能在客户端使 Web 浏览器改变文档

的表示法,从而不需要再与服务器进行交互通信;XML 将进一步扩展目前 Web 上已有的简单链接。

良好的数据存储格式、可扩展性、高度结构化、便于网络传输是 XML 主要的四大特点,决定了其卓越的性能表现。由于 XML 能针对特定的应用定义自己的标记语言,这一特征使得 XML 可以在电子商务、政府文档、报表、司法、出版、CAD / CAM、保险机构、厂商和中介组织信息交换等领域中一展身手。根据不同的系统、厂商提供各具特色的独立解决方案。

### (三) XML 与 SGML 的比较

XML 不是凭空发明出来的,它建立在 SGML 基础之上。SGML 是一个非常强大、非常通用的标准,但其强大功能的副作用是实现的复杂性,也使其用户仅局限于那些需要它全部功能的大公司。通常,SGML 的用户是具有几万页信息的组织。而 XML 是 SGML 的一个子集,它的目的是使 SGML 足够简单。XML 是 SGML 严格意义上的子集,也就是说,所有的 XML 文档都是有效的 SGML 文档,然而,并不是所有的 SGML 文档都是有效的 XML 文档。

如图 1-3 所示是 XML 与 SGML 关系的示意图。

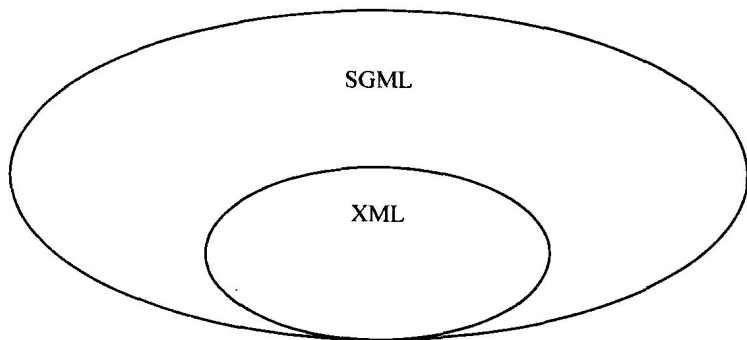


图 1-3 XML 与 SGML 的关系

### (四) XML 的主要特征

XML 作为新一代的网络语言,当然有其特殊的规定性,而在 XML 中使用的一些概念和术语就是这种特殊规定性的表现之一,它们构成了 XML 的主要特征。下面就将解释 XML 几个最常用的概念和术语。

#### 1. 元素

元素是 XML 文档的组件。元素可由其他元素、其他数据等组成。

XML 元素间的关系:由于 XML 元素必须规定根元素,所以根元素以下的元素都为子元素,而处在同一层上的元素之间为兄弟元素关系。为了数据存储方便,应确保子元素内容与根元素相关,并且各个兄弟元素间最好也有逻辑上的关联。

XML 元素内容:XML 文件是用于存储数据的,所以它具有一定的内容,而它的内容都是包含在各个 XML 元素之中。在这里,内容可以是空白,也可以是简单的文本或单一数据,还可以是复杂的多个数据列,甚至可以是其他的分类子元素。

XML 元素命名规则:XML 对其元素的命名限制较少,只要是数据读取软件支持的语言,