



高职高专计算机类课程改革规划教材

国家社会科学基金“十一五”规划(教育学科)国家级课题成果

COMPUTER



XML程序设计 案例教程

XML CHENGXU SHEJI ANLI JIAOCHENG

郝俊寿 丁艳会 主编



机械工业出版社
CHINA MACHINE PRESS



赠电子课件

高职高专计算机类课程改革规划教材

国家社会科学基金“十一五”规划（教育学科）国家级课题成果

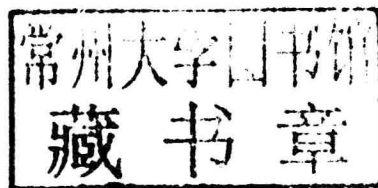
XML 程序设计案例教程

主 编 郝俊寿 丁艳会

副主编 王锐东 王凤岐

参 编 李文静 范哲超

主 审 包海山



机械工业出版社

本书内容紧扣国家对高职高专培养高级应用型、复合型人才技能水平和知识结构的要求,以若干个小型 Web 应用项目案例的开发思路为主线,采用模块分解、任务驱动、子任务实现、代码设计四层结构,通过对模块中每个任务相应知识点的详解,引导读者学习需求分析、分层设计、开发环境配置、功能实现以及整合测试等项目开发基本技能和 XML 语言相关知识。同时,本书采用学材小结、习题等方式对每个模块的理论和开发技能进行强化练习,以达到深化理解、熟练设计的目的。

本书内容分为 10 个模块:XML 概论、案例学习 XML 基础、XML 的 DTD 实例、Schema、使用 CSS 格式化 XML、XML 与 HTML、XSL 转换、XML DOM、XML 与数据交换和利用 XML 与 ASP.NET 构建小型 Web 应用程序——网上书店,涵盖了 IBM XML 认证考试要求的技能水平和知识范围,具有鲜明的职业特色,适合于高职高专、成人本专科和中职中专计算机相关专业程序设计语言类课程的教学,也可以作为计算机认证考试、高新技术考试的培训教材和 Web 应用开发人员、编程爱好者学习 XML 语言的参考书。

为方便教学,本书配备电子课件等教学资源。凡选用本书作为教材的教师均可登录机械工业出版社教材服务网 www.cmpedu.com 免费下载。如有问题请致信 cmpgaozhi@sina.com 或致电 010-88379375 咨询。

图书在版编目 (CIP) 数据

XML 程序设计案例教程/郝俊寿,丁艳会主编. —北京:机械工业出版社,2012.1
高职高专计算机类课程改革规划教材

国家社会科学基金“十一五”规划(教育学科)国家级课题成果

ISBN 978-7-111-36937-0

I. ①X… II. ①郝… ②丁… III. ①可扩展语言,XML—程序设计—高等职业教育—教材 IV. ①TP312

中国版本图书馆 CIP 数据核字(2011)第 277951 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

策划编辑:王玉鑫 责任编辑:王玉鑫 张 芳

封面设计:马精明 责任印制:杨 曦

北京鑫海金澳胶印有限公司印刷

2012 年 3 月第 1 版第 1 次印刷

184mm×260mm·13.5 印张·314 千字

0001—3000 册

标准书号:ISBN 978-7-111-36937-0

定价:26.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

社服务中心:(010)88361066

销售一部:(010)68326294

销售二部:(010)88379649

读者购书热线:(010)88379203

网络服务

门户网:<http://www.cmpbook.com>

教材网:<http://www.cmpedu.com>

封面无防伪标均为盗版

高职高专计算机类课程改革规划教材

编委会名单

主 任	包海山	陈 梅			
副主任	顾艳林	马 宁	那日松	艾 华	包乌格德勒
	恩和门德	来 全	李占岭	刘春艳	王瑾瑜
委 员	(按姓氏笔画排序)				
	丁艳会	马丽洁	马鹏烜	王 飞	王丽霞
	王应时	王晓静	王素苹	王鑫 _{内农大}	王鑫 _{内财院}
	付 岩	冉 明	包东生	田 军	田保军
	田 毅	白青山	刘宝娥	刘 静	孙 欢
	孙志芬	色登丹巴	邢海峰	何永琴	吴宏波
	吴和群	张丽萍	张利桃	张秀梅	张 芹
	张 娜	张 健	张维化	张惠娟	李友东
	李文静	李亚嘉	李红霞	李建锋	李 娜
	李 娟	李海军	杨东霞	杨忠义	杨 静
	迎 梅	陈俊义	陈瑞芳	孟繁军	孟繁华
	范哲超	侯欣舒	哈申花	胡姝璠	郝俊寿
	殷文辉	秦俊平	郭立志	高 博	高 歌
	崔 娜	曹文继	菊 花	萨日娜	塔林夫
	彭殿波	董建斌	蒙 君	赖玉峰	赖俊峰

项目总策划 包海山 陈 梅 王玉鑫

编委会办公室

主 任 卜范玉

副主任 王春红 郭喜聪

序

随着信息技术的发展,信息能力和传统的“读、写、算”能力一起正在成为现代社会中每一个人的基本生存能力。作为高等学校的学生,不仅要具备一般的信息能力,更应该具备较高的信息素养。因此,计算机类课程的改革一直是高等学校关注和研究的重点。

由包海山、陈梅策划并组织多所高等院校及高职高专院校编写的“高职高专计算机课程改革规划教材”,是根据面向 21 世纪培养高技能人才的需求,结合高职高专学生的学习特点,依据职业教育培养目标的要求,严格按照教育部提出的高职高专教育“以应用为目的,以必需、够用为度”的原则而设计、开发的系列教材。这套教材包括了信息技术公共基础课程、计算机专业基础课程和专业主干课程三部分内容,从高职高专的实际需求出发,重新整合了相关理论,突出了应用性和操作性,加强了能力的培养。

教材采用的“模块化设计、任务驱动学习”编写方式,对高等学校教材是一种新的尝试。实现任务驱动学习的关键是“任务”的设计,它必须是社会实际生产、生活中的一个真实问题,而不是为了验证理论而假设的虚拟事件。为了解决这个真实的问题,需要把它分解成一系列的“子任务”;每一个子任务的解决过程就是一个模块的学习过程。每个模块学习一组概念、锻炼一种技能;全部模块加起来,即完成一种知识的学习,形成一种相应的能力。任务驱动学习有利于学生从整体意义上理解每一个工作任务,掌握相关的知识和技能,形成解决实际问题的能力,提高学生的学习兴趣,是信息技术类课程有效的教学方式。

教材中每个模块安排的导读和要点提示要解决的问题,并用思维导图的形式给出了知识、技能和任务的分类和构成;知识导读部分体现了本模块需要学习的理论知识;子任务的划分安排了完成本模块总任务的各个步骤。利用模块最后的学材小结,学生可以自我检测对“理论知识”和“实训任务”掌握的程度;拓展练习可以为学有余力的学生提供个性化发展的方向。

参加本系列教材编撰工作的人员都是长期从事高职高专计算机教育和教学研究的专家和骨干教师,对高职高专的培养目标、学生的学习特点、计算机类课程的教学规律有着深刻的了解。我相信,本套教材的出版会对高职高专的计算机类课程的教学改革起到促进作用,对高职高专教学质量的提高将会产生显著的影响。

中国教育技术协会学术委员会委员
内蒙古师范大学现代教育技术研究所所长

2008 年 12 月



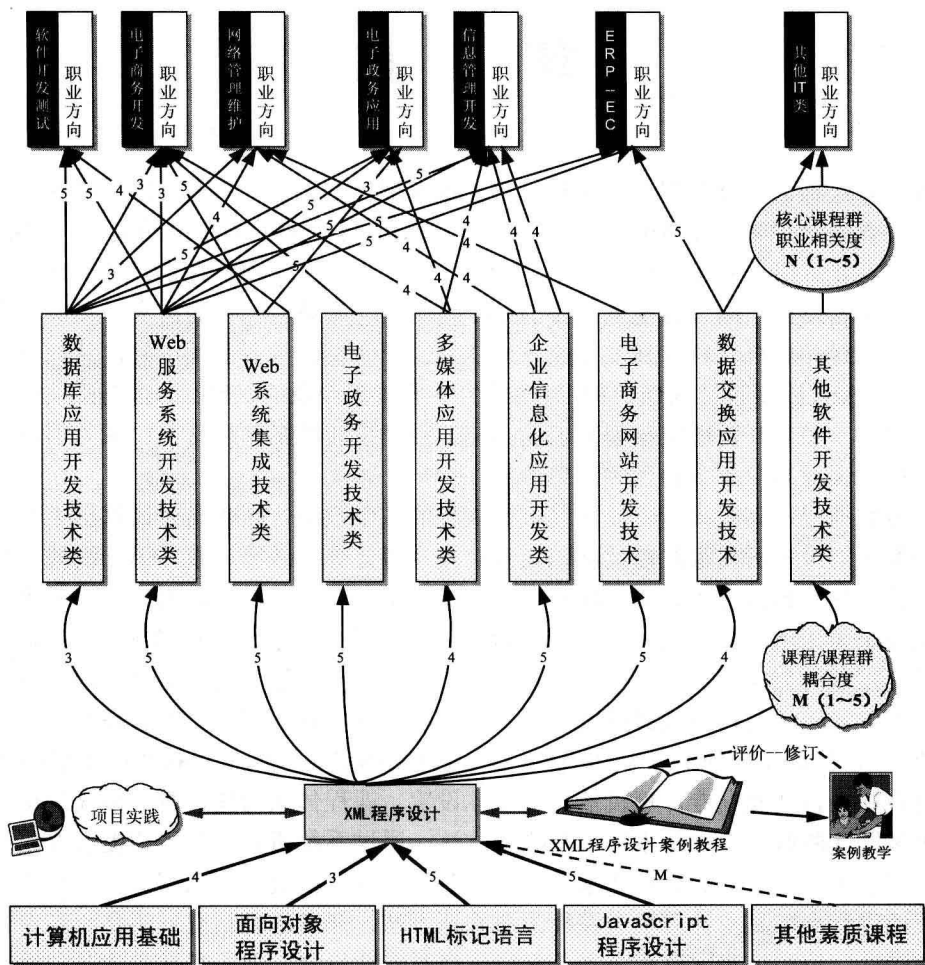
前 言

HTML 已经成为全球信息网的基础。它提供标准化的方法,将信息格式化并经由互联网传送给全世界的使用者。HTML 为人们传送和接受信息带来了革命性的变化。但是 HTML 设计的初衷主要是用来显示资料,因此,HTML 的焦点几乎完全放在信息应如何显示上,而不是信息的内容及其结构。这便是需要 XML 的原因。为适应社会的需求,目前多数高职高专院校除了针对计算机、软件开发专业开设 XML 程序设计专业课程以外,其他各相关专业也都开设了 XML 程序设计选修课程,以便学生掌握较为专业的 XML 语言编程技能和必要的软件开发知识。如何开发适应社会需求的 XML 程序设计类案例教材,已成为信息时代高职高专计算机类课程教学改革当务之急。

为了更好地促进高职高专院校计算机类课程的教学改革,高职高专计算机类课程改革规划教材编委会组织多所高等院校及高职高专院校从事计算机教研、教学第一线的专家和骨干教师,在认真分析和探讨教育部对高职高专各专业的培养目标、国家计算机等级考试和职业技能鉴定要求的基础上策划了“高职高专计算机类课程改革规划教材”。同时,编委会向中国教育技术协会申报了国家社会科学基金“十一五”规划(教育学科)国家级课题“信息技术环境下多元学与教方式有效融入日常教学的研究”的子课题“高职高专计算机类课程改革的研究”,目前课题研究正在进行中。本课题立项研究面向信息技术职业领域不同岗位层次如何有效融合高职高专计算机信息类专业设置、课程体系构建、教学模式改革和教材课件开发等多层次的教学设计基本理论和实现方法。通过系统研究,总结和提炼课题组成员以及有关专家学者已经取得的相关成果,探索高职高专计算机类专业课程标准建设的新思路,提出系统地进行高职高专计算机类课程改革的新方法,开发建设具有鲜明高职高专特色的系列教材和课件,旨在为我国高职高专计算机信息类专业设置、课程和教学改革、教材课件建设探索出一条坦途。

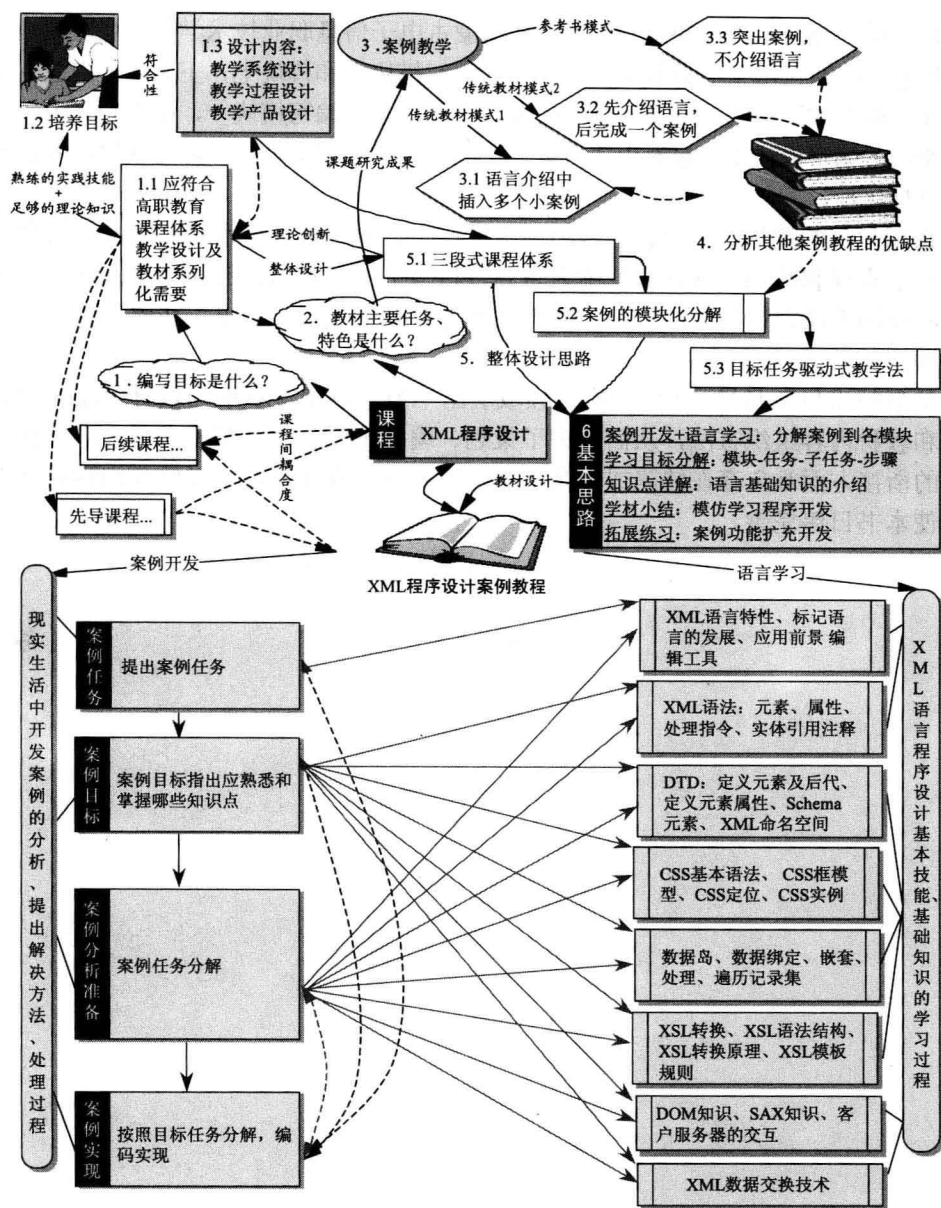
根据“课程间的耦合度”和“职业相关度”,我们将计算机信息类专业课程群体划分为职业基础能力层面、职业岗位能力层面和职业拓展能力层面等三个层次。“XML 程序设计”课程在高职高专计算机信息类专业的课程体系中属于职业基础能力层面,是学生学习当今流行的 Web 服务、Web 系统集成等领域的软件开发职业方向各种技能课程的基础,同时又是数据交换、信息处理和应用服务器配置的必备理论知识和操作技能的主要组成部分。可见,在高职高专计算机信息类各专业课程体系中,“XML 程序设计”课程作为多门职业岗位能力层面课程或职业拓展能力层面课程(课程群)的前导课程而起着非常重要的作用。但在高职高专教育层次,XML 程序设计课程一直没有一个大家都能够接受的标准,主要原因是 XML 语言环境涵盖的内容不断丰富,跨平台软件开发技术发展又非常迅速,这些不确定因素给课程标准的制定带来一定的困难。通过立项研究,我们认为以一个完整的软件开发案例为主线,将 XML 语言基础知识学习和软件开发基本技能实训并举作为课程标准的依据,在制定课程标准、开发教材课件以及课堂教学中应充分体现本课程的“基础性”和“职业性”特色。

因此,本书将围绕案例的任务、案例的目标、案例的分析、案例的实现等核心组织编写。



课程体系中的本课程、教材定位

作为以学习应用技术为目标的职业基础能力课程教材,“XML 程序设计案例教程”主要根据课题组提出并实现的高职高专计算机信息类专业课程“三段式课程体系、案例的模块化分解、目标任务驱动式教学法”的整体教学设计思想,针对利用 XML 程序设计操作技能和相关知识进行问题分解式介绍。在编写模式上采用目标任务驱动式教学法,让学生带着问题主动学习和实训:用一个完整的开发案例作为一门课程教学内容的引线,并将完整开发过程分解到介绍 XML 程序设计基础知识和基本技能的各个模块之中,然后根据教学目标将每个模块分解为若干个相对独立的学习/实训任务以及细化的多个子任务,最后对每个子任务中的每个开发步骤进行逐步介绍。对于各个任务中涉及的知识点进行适时适量讲解,将抽象的理论知识融入到实践活动中加以演绎和关联,力求达到高职高专教学目标。为强化教学内容,在每个模块后面采用学材小结、习题等学习内容,让学生在课堂内外对教学内容进行强化训练、深化理解。



案例的模块化分解与目标任务驱动式教学法

鉴于目前信息化应用对 IT 行业应用型人才“技能+知识”结构的需求, 基于 XML 程序设计技能对于软件开发与测试、网络管理维护、信息管理, 以及电子政务、企业应用、电子商务系统开发等计算机信息类各相关职业岗位方向的高职高专学生来说都是不可或缺的职业技能和理论知识, 但各类职业方向所需的技能和知识侧重面有所不同。因此, 在制定 XML 程序设计课程的教学目标、内容和课时数时应充分考虑其基础性、应用性、职业性和工程性特点。

本书由郝俊寿、丁艳会任主编。参加编写的教师及分工如下: 郝俊寿(内蒙古电子信息职

业技术学院)编写模块八和模块九,丁艳会(内蒙古电子信息职业技术学院)编写模块三~模块五,王锐东(内蒙古电子信息职业技术学院)编写模块六和模块七,王凤岐(内蒙古电子信息职业技术学院)编写模块二,李文静(内蒙古工业大学)编写模块一,范哲超(内蒙古机电职业技术学院)编写模块十。包海山(内蒙古财经学院)审阅了全稿,并对教材内容提出了修改意见和合理化建议。

在本系列教材的策划、组织、编写和出版过程中,中国教育技术协会学术委员会委员李龙教授给予了指导和帮助,并在百忙中为系列教材写了序。本书参考和引用了许多著作和网站内容,除确因无法查证出处的以外,我们在参考文献中都一一列出。在此,我们一并表示衷心的感谢。

由于信息技术日新月异,XML 应用领域层出不穷,再加上本系列教材旨在探索全新的教学模式和教材内容组织方法,从而加大了策划、编写难度。由于编者水平有限,在内容整合、项目的衔接方面难免存在缺陷或不当之处,敬请读者批评指正,以便我们再版时进行修订补充,使本书日臻完善。

编者

目 录

序

前言

模块一 XML 概论..... 1

任务一 了解标志语言的发展..... 2

子任务 1 了解 SGML 起源..... 2

子任务 2 了解 HTML 出现背景..... 2

子任务 3 了解 XML 的诞生..... 2

子任务 4 了解 XML 与 HTML 的主要区别..... 3

任务二 了解 XML 的特点和应用前景..... 3

子任务 1 了解 XML 的优缺点..... 3

子任务 2 了解 XML 的应用前景..... 4

任务三 了解 XML 开发和应用环境..... 5

学材小结..... 5

习题..... 5

模块二 案例学习 XML 基础..... 6

任务一 XML 基本操作案例——通信录..... 7

任务二 案例掌握 XML 语法知识..... 9

子任务 1 XML 的标记与元素..... 9

子任务 2 XML 的属性..... 11

子任务 3 XML 声明..... 12

子任务 4 处理指令..... 12

子任务 5 字符数据与实体引用..... 13

子任务 6 CDATA 段..... 13

子任务 7 XML 注释..... 14

子任务 8 格式良好的 XML 文档..... 14

任务三 XML 的名称空间..... 15

学材小结..... 16

习题..... 16

模块三 XML 的 DTD 实例..... 17

任务一 了解 DTD..... 18

子任务 1 什么是 DTD..... 19

子任务 2 为什么要使用 DTD..... 19

任务二 XML 文档的 DTD 实例——给通信

录增加 DTD 文件..... 19

子任务 1 给通信录增加内部 DTD..... 20

子任务 2 给通信录增加外部 DTD..... 21

子任务 3 给通信录增加公用 DTD..... 23

任务三 定义元素及其后代..... 24

子任务 1 元素类型声明..... 24

子任务 2 定义元素及其子元素..... 26

子任务 3 有顺序的子元素..... 28

子任务 4 重复元素..... 29

子任务 5 成组元素..... 29

子任务 6 或 (OR)..... 30

子任务 7 可选子元素..... 31

子任务 8 混合内容..... 32

子任务 9 空元素..... 32

任务四 定义元素属性..... 33

子任务 1 定义有效的元素属性..... 33

子任务 2 属性默认值..... 33

子任务 3 属性类型..... 34

子任务 4 CDATA 类型..... 34

子任务 5 枚举类型..... 35

子任务 6 ID 和 IDREF..... 35

子任务 7 NMTOKEN 和 NMTOKENS..... 37

子任务 8 NOTATION 类型..... 37

子任务 9 实体属性类型与参数实体..... 38

学材小结..... 39

习题..... 39

模块四 Schema..... 41

任务一 了解 Schema 的由来..... 42

子任务 1 了解 Schema 的诞生..... 42

子任务 2 了解 Schema 的发展历程..... 43

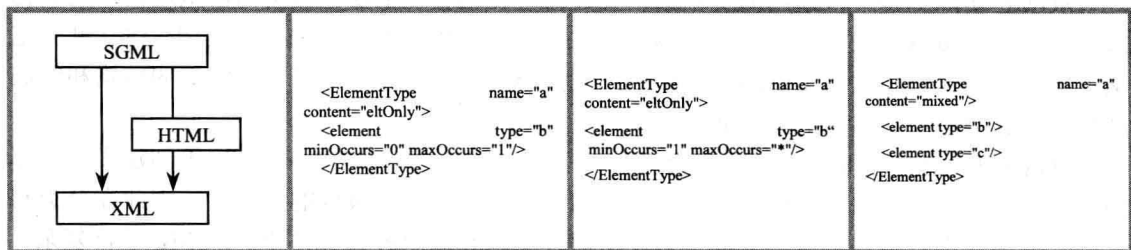
任务二 定义元素及其后代..... 45

子任务 1 Schema 的简单实例——给通信录增加 Schema	45	子任务 4 CSS 外边距	89
子任务 2 Schema 文件的结构	49	子任务 5 CSS 外边距合并	89
子任务 3 用 Schema 定义元素及其内容	50	任务四 CSS 定位	91
子任务 4 用 Schema 定义元素属性	53	子任务 1 CSS 定位概述	91
子任务 5 Schema 的特色——扩展数据类型	55	子任务 2 CSS 相对定位	93
子任务 6 Schema 中的注释	57	子任务 3 CSS 绝对定位	93
子任务 7 Schema 的应用	58	子任务 4 CSS 浮动	94
任务三 XML 命名空间	59	任务五 CSS 实例——英超联赛球队网页	98
子任务 1 前缀标识法	61	教材小结	101
子任务 2 如何声明命名空间	62	习题	102
子任务 3 什么是合法名称	63	模块六 XML 与 HTML	103
任务四 命名空间的使用	63	任务一 了解 XML 的局限性	104
子任务 1 命名空间作用于元素	63	任务二 IE 中数据岛的处理方式	104
子任务 2 命名空间作用于属性	65	任务三 XML 的数据绑定	105
子任务 3 巧用命名空间丰富表现效果	66	任务四 XML 的嵌套处理	107
子任务 4 命名空间与 DTD	69	任务五 DSO 技术的应用	108
教材小结	71	任务六 遍历记录集	108
习题	71	任务七 与 Script 语言相结合	108
模块五 使用 CSS 格式化 XML	72	子任务 HTML 中的 XML 数据岛记录集 页面管理	109
任务一 了解 CSS	73	任务八 数据岛实例——图书资料显示实例	110
子任务 1 CSS 概述	73	教材小结	115
子任务 2 样式解决了一个普遍的问题	73	习题	115
子任务 3 样式表极大地提高了工作效率	73	模块七 XSL 转换	116
子任务 4 多重样式将层叠为一个	74	任务一 了解 XSL 的背景知识	117
任务二 CSS 的基本语法	74	子任务 1 XSL 简介	117
子任务 1 CSS 语法	74	子任务 2 XML 文档转换原理	117
子任务 2 如何创建 CSS	79	子任务 3 XML 文档的转换步骤	118
子任务 3 CSS 背景	80	子任务 4 XSL 文档标准格式	118
子任务 4 CSS 文本	84	子任务 5 XSL 的语法结构	119
子任务 5 CSS 字体	84	子任务 6 XSL 的模板规则	119
子任务 6 CSS 列表	85	任务二 XSL 转换实例——网站购物车	122
子任务 7 CSS 表格	85	教材小结	125
子任务 8 CSS 轮廓	86	习题	125
任务三 CSS 框模型	86	模块八 XML DOM	127
子任务 1 CSS 框模型概述	86	任务一 了解 DOM 基础知识	128
子任务 2 CSS 内边距	88	子任务 1 了解接口的由来	128
子任务 3 CSS 边框	88	子任务 2 DOM 与 SAX 并存	129
		任务二 DOM 组成、DOM 树和接口	130

子任务 1 DOM 的组成	130	子任务 1 XML 数据交换机制	158
子任务 2 DOM 树	130	子任务 2 XML 数据存取机制	160
子任务 3 DOM 的 4 个基本接口	132	任务二 XML 与数据库	162
任务三 XML 文档访问	134	子任务 1 XML 与关系数据库	162
子任务 1 创建 Document 对象	134	子任务 2 XML 与面向对象数据库	163
子任务 2 加载 XML 文档	134	子任务 3 基于 XML 数据库的模型分析	164
子任务 3 遍历 XML 文档	135	子任务 4 XML 在数据库中的应用模式	165
子任务 4 添加元素	137	任务三 XML 数据交换技术	166
子任务 5 删除元素	138	子任务 1 ADO 控件技术	166
子任务 6 修改元素	138	子任务 2 HTTPXML 对象技术	171
任务四 SAX 基础知识	138	子任务 3 DB2XML 转换工具	173
子任务 1 SAX 分析器接口简介	138	子任务 4 ODBC2XML 转换工具	174
子任务 2 生成应用程序框架	140	子任务 5 XOSL 转换工具	175
子任务 3 引入需要的类	141	子任务 6 ASP2XML 组件	176
子任务 4 创建分析器	141	子任务 7 InterAccess 软件包	176
子任务 5 实现 DocumentHandler 接口	142	子任务 8 XML Servlet 软件包	178
子任务 6 输出异常	143	子任务 9 WDDX Web 分布式数据交换	179
子任务 7 简单格式化输出	143	任务四 数据库对 XML 的支持	181
子任务 8 处理文档事件	144	子任务 1 MS SQL Server	181
任务五 应用程序的编译与执行	145	子任务 2 Oracle 8i	183
任务六 客户机与服务器的数据交互	146	子任务 3 Informix	184
子任务 1 C/S 的通信模式	146	学材小结	185
子任务 2 基于 XML 的 C/S	146	习题	185
子任务 3 向服务器发送数据	146	模块十 利用 XML 与 ASP.NET 构建小型 Web 应用程序——网上书店	186
子任务 4 XMLHTTP 对象的 Open 方法	148	任务 利用 XML 和 ASP.NET 构建动态网站	187
子任务 5 XMLHTTP 对象的 Send 方法	148	子任务 1 网上书店项目概述	187
子任务 6 服务器端的数据处理	149	子任务 2 网站后台数据文件	187
子任务 7 服务器端的响应消息	149	子任务 3 网站站点架构	190
子任务 8 客户机处理响应信息	150	子任务 4 具体实现	190
任务七 DOM 实例——北京申奥	151	学材小结	202
学材小结	156	习题	202
习题	156	参考文献	203
模块九 XML 与数据交换	157		
任务一 XML 的数据交换机制与数据存取机制	158		

模块一

XML 论坛



本模块导读

HTML 已经成为全球信息网的基础,它提供标准化的方法将信息格式化,并经由互联网传送给全世界的使用者。HTML 为人们传送和接受信息带来了革命性的变化。但是 HTML 主要是被设计为资料显示所用,因此,HTML 的焦点几乎完全放在信息应如何显示上,而不是信息的内容及其结构。这便是我们需要 XML 的原因。

本模块的主要目的是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。

本模块要点

- 了解标记语言的发展和现状
- 掌握 XML 文档的数据结构和基本语法
- 掌握 XML 的特点

任务一 了解标志语言的发展

子任务 1 了解 SGML 起源

XML 有两个先驱——SGML 和 HTML，这两种语言都是非常成功的标记语言。SGML 的全称是 Standard Generalized Markup Language（标准通用化标记语言），它是国际定义电子文档结构和内容描述的标准。

SGML 实际上是一种通用的文档结构描述符号化语言，主要定义文档模型逻辑结构和物理结构。一个 SGML 文件由 3 部分组成：语法定义、文档类型定义和文档实例。语法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 文件的主体部分。

SGML 从 20 世纪 80 年代初开始使用。正如 XML 一样，SGML 也可用于创建成千上万的标记语言，它为语法置标提供了异常强大的工具，同时具有极好的扩展性，因此在分类和索引数据中非常有用。目前，SGML 多用于科技文献和政府办公文件中。SGML 非常复杂，其复杂程度远远超乎网络上的日常应用。不仅如此，SGML 还非常昂贵。

子任务 2 了解 HTML 出现背景

1989 年，一位名叫 Tim Berners-Lee 的研究人员提出，用超链接文本文档实现在 CERN 欧洲原子能研究所内部的信息共享。在同事 Anders Berglund 的建议下，Tim 采用了 SGML 形式的语法。他们从 SGML 简单文档入手，开发出了 HTML（Hypertext Markup Language，超文本标记语言）。它是 SGML 的一个简化子集。HTML 继承了 SGML 的许多重要特点，如结构化、实现独立和可描述性，同时也存在很多缺点。例如，它只能使用固定的有限标记，而且只侧重于对内容的显示。

子任务 3 了解 XML 的诞生

Internet 提供了全球范围的网络互联与通信功能，Web 技术的发展更是一日千里，其丰富的信息资源给人们的学习和生活带来了极大的便利。特别是应运而生的 HTML（超文本置标语言），以简单易学、灵活通用的特性，使人们发布、检索、交流信息都变得非常简单，从而使 Web 成为最大的环球信息资源库。然而，电子商务、电子出版、远程教育等基于 Web 的新兴领域的全面兴起使得传统的 Web 资源更加复杂化、多样化，数据量的日趋庞大对网络的传输能力也提出更高的要求。同时，人们对 Web 服务功能的需求也达到更高的标准，如用户需要对 Web 进行智能化的语义搜索和对数据按照不同的需求进行多样化显示等个性化服务；公司和企业要为客户创建和分发大量有价值的文档信息，以降低生产成本，以及对不同平台、不同格式的数据源进行数据集成和数据转化等，这些需求越来越广泛和迫切。

传统的 HTML 由于自身特点的限制,不能有效地解决上述问题:作为一种简单的表示性语言,它只能显示内容而无法表达数据内容。而这一点恰恰是电子商务、智能搜索引擎所必需的。另外,HTML 语言不能描述矢量图形、数学公式、化学符号等特殊对象,在数据显示方面的描述能力也不尽如人意。最重要的是,HTML 只是 SGML 的一个实例化的子集,可扩展性差,用户根本不能自定义有意义的置标供他人使用。这一切都成为 Web 技术进一步发展的障碍。

SGML 是一种通用的文档结构描述置标语言,为语法置标提供了异常强大的工具,同时具有极好的扩展性,因此在数据分类和索引中非常有用。但 SGML 太过复杂,不适合网络的日常应用,加上开发成本高、不被主流浏览器所支持等原因,使得 SGML 在 Web 上的推广受到阻碍。在这种情况下,开发一种兼具 SGML 的强大功能、可扩展性以及 HTML 的简单性的语言势在必行,由此诞生了 XML 语言。标记语言的层次结构如图 1-1 所示。

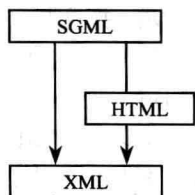


图 1-1 标记语言的层次结构

XML (eXtensible Markup Language, 可扩展置标语言) 是由 W3C 于 1998 年 2 月发布的一种标准。它同样是 SGML 的一个简化子集,将 SGML 的丰富功能与 HTML 的易用性结合到 Web 的应用中,以一种开放的自我描述方式定义了数据结构,在描述数据内容的同时能突出对结构的描述,从而体现出数据之间的关系。这样所组织的数据对于应用程序和用户都是友好的、可操作的。

子任务 4 了解 XML 与 HTML 的主要区别

XML 与 HTML 的主要区别有:

- 1) XML 是用来存放数据的。
- 2) XML 不是 HTML 的替代品,XML 和 HTML 是两种不同用途的语言。
- 3) XML 是被设计用来描述数据的,其重点是:什么是数据,如何存放数据;HTML 是被设计用来显示数据的,其重点是:显示数据以及如何更好地显示数据。
- 4) HTML 是与显示信息相关的,XML 则是与描述信息相关的。

任务二 了解 XML 的特点和应用前景

子任务 1 了解 XML 的优缺点

XML 的优势之一是它允许各个组织、个人建立适合自己需要的置标集合,并且这些置标可以迅速地投入使用。这一特征使得 XML 可以在电子商务、政府文档、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,针对不同的系统、厂商提供各具特色的独立解决方案。

XML 的最大优点在于它的数据存储格式不受显示格式的制约。一般来说,一篇文档包括 3 个要素:数据、结构以及显示方式。对于 HTML 来说,显示方式内嵌在数据中,这样在创建文本时,要时时考虑输出格式,如果因为需求不同而需要对同样的内容进行不同风格的显示时,要从头创建一个全新的文档,重复工作量较大。此外,HTML 缺乏对数据结构的描述,对于应用程序理解文档内容、抽取语义信息都有诸多不便。

XML 把文档的三要素独立开来,分别处理。首先把显示格式从数据内容中独立出来,保存在样式单文件(Style Sheet)中,这样如果需要改变文档的显示方式,只要修改样式单文件即可。XML 的自我描述性质能够很好地表现许多复杂的数据关系,使得基于 XML 的应用程序可以在 XML 文件中准确、高效地搜索相关的数据内容,忽略其他不相关部分。XML 还有许多其他优点,如它有利于不同系统之间的信息交流,完全可以充当网际语言,并希望成为数据和文档交换的标准机制。

当然,XML 作为一个新建立的标准,还有许多不足之处:它在强调了数据结构的同时,语义表达能力上略显不足,如定义了<地址>这样一个置标,如果不是在文档中实际定义内容,就无法知道是要表达家庭住址还是 E-mail 地址。另外,XML 的有些技术尚未形成统一的标准,充分支持 XML 的应用处理程序很少,甚至浏览器对 XML 的支持也是有限的。

因此,XML 还并不能完全取代 HTML,毕竟 HTML 是最为方便、快捷的网上信息发布方式。况且 HTML 是描述数据显示的语言,而 XML 是描述数据及其结构的语言,二者在功能上也是截然不同的。

子任务 2 了解 XML 的应用前景

不管怎样,Web 的应用将随着 XML 的发展而更加精彩。

1. 商务的自动化处理

XML 的丰富置标完全可以描述不同类型的单据,如信用证、保险单、索赔单以及各种发票等。结构化的 XML 文档发送至 Web 的数据可以被加密,并且很容易附加上数字签名。因此,XML 有希望推动 EDI(Electronic Data Interchange)技术在电子商务领域的大规模应用。有兴趣的读者可以访问网站 <http://www.xmledi.org>。

2. 信息发布

信息发布在企业的竞争中起着重要作用。服务器只需发出一份 XML 文件,客户可根据自己的需求选择和制作不同的应用程序以处理数据。加上 XSL(eXtensible Stylesheet Language)的帮助,使广泛的、通用的分布式计算成为可能。

3. 智能化的 Web 应用程序和数据集成

XML 能够更准确地表达信息的真实内容,其严格的语法降低了应用程序的负担,也使智能工具的开发更为便捷。来自不同应用程序的数据也能够转化到 XML 这个统一的框架中,进行交互、转化和进一步的加工。

XML 的优点备受瞩目,它的发展方兴未艾,可以预见,未来的 Web 将是 XML 的 Web。

任务三 了解 XML 开发和应用环境

开发 XML 有许多工具可以使用:

- 1) Notepad 是最直接、最简单的文本编辑工具,在 Windows 附件中就可以找到。
- 2) Microsoft XML Notepad 是微软公司专门为设计 XML 文档而提供的编辑软件,可以借助它验证 XML 文档的有效性,详细说明和下载地址见 <http://msdn.microsoft.com/xml/NOTEPAD/intro.asp>。
- 3) Visual InterDev 该软件是用来开发 Web 应用程序的,不仅仅局限于 XML,还可以开发 ASP、HTML、XSL 样式单等。
- 4) Microsoft XML Tree Viewer 利用这个软件可以把 XML 文档的内容用树的结构形式显示出来,下载地址为 <http://msdn.microsoft.com/xml/demos/default.asp>。
- 5) Microsoft XML Validator 该软件可以检查 XML 文档是否是“格式良好的”以及其有效性,并对错误发出警告,它的下载地址是 <http://msdn.microsoft.com/xml/demos/default.asp>。
- 6) Microsoft XSL Debugger 样式单文件的复杂性使开发人员在编写时容易出现错误,这个软件就是帮助用户调试样式单文件的,把复杂枯燥的调试过程用可视化界面显示出来。下载地址是 http://msdn.microsoft.com/xml/_archive/xsl-debugger/xsl-debugger.htm。
- 7) WordPerfect 一种对开发 XML 和 SGML 提供高级支持的文字处理器,具有所见即所得的开发环境,提供向导、自动化控件的插入和文档的自动生成等功能,它是一款收费的商业产品,支持 Windows 95/98/2000 和 Linux 平台。详情可访问 <http://www.corel.com>。
- 8) SixPack 该软件为 XML 的解析和开发提供简洁的界面,支持 Macintosh 平台,并公开源代码,详情可访问 <http://www.trafficstudio.com/sixpack/info.htm>。
- 9) Xray 一种具有实时错误检查的 XML 编辑器。它根据 DTD 或者 XML Schema,允许用户创建格式良好的 XML 文档或验证文档的有效性,并且支持多文档编辑,是一款免费软件,支持 Windows 95/98/NT/2000 平台。详情可访问 <http://architag.com/xray>。

学 材 小 结

作为新一代的网络发布语言,XML 为各种 Web 应用的飞速发展奠定了坚实的基础,其强大而灵活的数据描述机制为各类开发人员提供了极大的便利。通过本模块的介绍,读者应对什么是 XML、XML 产生的原因、XML 与 HTML 及 SGML 相比所具有的优势、XML 的应用领域及应用前景,以及开发 XML 的常用工具有一个初步的认识和了解。

习 题

1. 比较 HTML 与 XML 的异同。
2. 开发 XML 的常用工具有哪些?