

21世纪高等院校计算机规划教材

Web

语言与应用导论

主 编 杨德仁 保文星

副主编 薛 梅 丁红胜



Web



科学出版社

21 世纪高等院校计算机规划教材

Web 语言与应用导论

主 编 杨德仁 保文星

副主编 薛 梅 丁红胜

科学出版社

北 京

内 容 简 介

随着Web的问世和普及,其编程语言日趋显得重要,成为主流计算机软件即网络应用软件不可缺少的组成部分。本书包括Web编程语言概论、超文本标记语言(HTML)、无线标记语言(WML)、可扩展标记语言(XML)、可扩展超文本标记语言(XHTML)、动态网页编程技术(PHP和JSP)、搜索引擎优化(SEO)技术等标记语言及其相关技术。

本书可作为高等院校计算机科学与技术、电子信息科学与技术、管理信息系统、软件工程等相关专业的教材,也适合从事搜索引擎优化技术的网站优化工程师、拥有网站的企事业单位的网站管理员等相关工程技术人员参考和阅读。

图书在版编目(CIP)数据

Web语言与应用导论 / 杨德仁, 保文星主编. — 北京: 科学出版社, 2011.4

ISBN 978-7-03-030484-1

I. ①W… II. ①杨… ②保… III. ①主页制作—程序设计
IV. ①TP393.092

中国版本图书馆 CIP 数据核字 (2011) 第 038228 号

责任编辑: 韩卫军 荣洁莉 / 封面设计: 陈思思

科学出版社出版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

四川煤田地质制图印刷厂印刷

科学出版社发行 各地新华书店经销

*

2011年6月第 一 版 开本: 787×1092 1/16

2011年6月第一次印刷 印张: 13 1/4

印数: 1—1 500 字数: 300 000

定价: 36.00 元

前 言

随着 Web 技术的问世及其应用的普及, Web 语言和技术显得日趋重要, 已成为业界的主流开发语言。在变化多端的 Web 应用中, 无论是个人主页、社交社区、信息门户、博客和论坛, 还是企业运营性信息平台, Web 语言及其技术不可或缺, 其重要性日益彰显。

高校相关专业在教学过程中, 已经逐步采用了相关内容作为教学材料, 如 HTML, ASP, JSP, PHP, XML 等, 但大多单独开设此类课程, 占用课时量大。对 IT 类专业而言, 零星开设这类课程、又没有涉及到业界最新应用技术, 不足以向学生全面介绍相关知识。高校在设置和讲授相关课程时, 面临着一系列挑战。其中困境之一是没有合适的教材, 国内尚缺乏综合介绍 Web 语言及其应用技术的书籍, 不能满足 IT 类专业教学和教改需求。因此, 我们编写了《Web 语言与应用导论》一书。

本书分为四部分, 第一部分是 Web 语言与应用概述; 第二部分介绍了常用的标记语言, 第三部分介绍了用于编写动态页面的脚本语言; 第四部分介绍了相关应用技术和前景。本书从超文本标记语言 (HTML)、动态网页编程技术、可扩展标记语言 (XML)、可扩展超文本标记语言 (XHTML)、移动标记语言 (WML 和 XHTML MP)、Web 脚本语言 (PHP 和 JSP)、搜索引擎优化 (SEO) 技术和 Web 发展趋势等角度, 较为全面地阐述了 Web 语言及其应用技术, 以满足高校教学和工程实践应用需要。

本书由多年从事高校 IT 教育的教师和一线 IT 工程师联合编写。他们分别是宁夏医科大学的杨德仁教授、杨柳和董富江讲师, 北方民族大学的保文星教授、丁红胜和丁剑讲师, 上海电力学院的薛梅博士, 宁夏师范学院的李金娟讲师, 宁夏电信公司的张春涛和王生国工程师, 北京地杰灵珂科技发展有限公司的贾磊工程师。

本书由杨德仁教授主编, 保文星教授为第二主编, 薛梅博士、丁红胜讲师为副主编。各章名称及其编写人员如下:

第 1 章 概论 (杨德仁)

第 2 章 超文本标记语言基础 (杨柳)

第 3 章 可扩展标记语言基础 (丁剑)

第4章 可扩展超文本标记语言基础（王生国）

第5章 无线标记语言和 XHTML MP 语言基础（张春涛）

第6章 PHP 动态页面语言基础（李金娟，贾磊）

第7章 JSP 动态页面语言基础（董富江）

第8章 Web 搜索引擎优化技术（杨德仁）

第9章 Web 2.0 技术应用与展望（薛梅）

全书由杨德仁教授和保文星教授统稿。辽宁省丹东市化纤技校的郝敬曼高级讲师友情校正了全文，在此表示诚挚感谢。

在本书编写过程中，编者得到了下列项目基金的支持：国家科技支撑计划项目(2007BAD33B03), 宁夏自然科学基金项目（NZ10103），科技攻关计划项目（KGX-01-10-01），宁夏医科大学特殊人才项目（XT200913），宁夏医科大学教改项目（标记语言应用与教学研究）。

本书的讲义稿在宁夏医科大学电子信息科学与技术专业和管理信息系统专业试用两年来，反应良好。本书可作为高等院校计算机科学与技术、电子信息科学与技术、管理信息系统、软件工程等相关专业的教材，也适合于 IT 工程技术人员阅读和参考。

因编写时间仓促，不当之处敬请读者指正。作者的联系方式：ydr@tom.com。

编 者

2011 年 1 月于银川

目 录

第 1 章 概论	1
1.1 计算机语言	1
1.1.1 编程语言	1
1.1.2 标记语言	2
1.1.3 脚本语言	2
1.1.4 形式语言	3
1.2 标记语言	4
1.2.1 SGML	4
1.2.2 HTML	4
1.2.3 XML	4
1.2.4 XHTML	4
1.2.5 WML	5
1.2.6 其他标记语言	5
1.3 Web 与脚本语言	5
1.3.1 互联网	5
1.3.2 Web	5
1.3.3 Web 浏览器和 Web 服务器	6
1.3.4 动态 Web 页面和脚本语言	6
1.3.5 应用服务器	8
1.3.6 Web 1.0 与 Web 2.0	8
1.4 软件体系结构及其演化	8
1.4.1 亚终端/主机模式	8
1.4.2 胖客户端/服务器模式	9
1.4.3 瘦客户端/服务器模式	9
1.4.4 RIA 模式	9
1.4.5 云计算模式	9
1.5 扩展阅读材料	10
1.5.1 ISO	10
1.5.2 W3C	10
1.5.3 Tim Berners-Lee	10

第 2 章 超文本标记语言基础	11
2.1 HTML 概述	11
2.1.1 从 SGML 到 HTML	11
2.1.2 HTML 与 URL	11
2.1.3 HTML 文档字符集	12
2.1.4 HTML 字符实体	13
2.1.5 编写 HTML 文档的方法	13
2.2 HTML 基本语法	14
2.2.1 HTML 元素和标签	14
2.2.2 HTML 注释	15
2.2.3 基本结构标签	15
2.2.4 文本格式标签	17
2.2.5 列表标签	20
2.2.6 超链接标签	23
2.2.7 图像标签	24
2.2.8 表格标签	25
2.2.9 表单标签	28
2.2.10 框架标签	30
2.3 HTML 事件	33
2.4 客户端动态 HTML 页面编程技术简介	35
2.4.1 JavaScript 基本概念	35
2.4.2 JavaScript 的作用	35
2.4.3 JavaScript 在 HTML 文档中的应用	36
2.5 Web 服务器与 HTML 文档发布	37
2.5.1 Web 服务器简介	37
2.5.2 静态网站的发布	38
第 3 章 可扩展标记语言基础	40
3.1 XML 概述	40
3.1.1 HTML 的缺陷	40
3.1.2 XML 的产生	40
3.1.3 XML 的特性	41
3.1.4 XML 的结构	42
3.1.5 XML 文档编辑器和浏览器	43

3.2 XML 语法	43
3.2.1 声明	44
3.2.2 注释	44
3.2.3 元素与标记	44
3.2.4 命名空间	45
3.3 文档类型定义	46
3.3.1 DTD 概述	46
3.3.2 XML 文档的构建模块	46
3.3.3 元素声明	48
3.3.4 属性声明	50
3.3.5 实体声明	54
3.4 XML 解析器	55
3.4.1 DOM 解析器	55
3.4.2 SAX 解析器	58
3.4.3 DOM 与 SAX 的比较	60
3.5 XML 应用简介	60
 第 4 章 可扩展超文本标记语言基础	62
4.1 XHTML 基本语法	63
4.1.1 XHTML DTD	63
4.1.2 XHTML 字符集	64
4.1.3 XHTML 使用严谨的标记	64
4.1.4 XHTML 元素的属性	66
4.1.5 XHTML 元素的事件	67
4.1.6 XHTML 模型	67
4.2 XHTML 编程技术	68
4.2.1 XHTML 框架技术	68
4.2.2 XHTML 与 CSS 技术结合	69
4.3 浏览器技术	73
4.3.1 浏览器概述	73
4.3.2 浏览器内核技术	73
4.3.3 常见的浏览器	74
4.4 编辑器的使用	75
4.4.1 编辑器概述	75
4.4.2 Dreamweaver 的使用	75

4.5 升级 HTML 文档至 XHTML 文档	78
第 5 章 无线标记语言和 XHTML MP 语言基础	80
5.1 WML 的编辑、测试与应用环境	80
5.1.1 WML 网页编辑	80
5.1.2 WAP 网关及其配置	80
5.1.3 WML 程序结构	81
5.2 WML 语言基础	82
5.2.1 WML 字符的基本规则	82
5.2.2 卡片、卡片组及其属性	82
5.2.3 文本格式化及其元素	85
5.2.4 链接和图像	86
5.2.5 用户输入元素	87
5.2.6 事件	89
5.2.7 定时器元素	90
5.2.8 任务及其元素	91
5.2.9 变量及其设置	93
5.3 XHTML MP 简介	94
5.3.1 XHTML MP 的语法规则	94
5.3.2 MIME 类型和文件扩展名	95
5.3.3 XHTML MP 文档结构	95
5.3.4 XHTML MP 的元标签	96
5.3.5 XHTML MP 标签简介	97
5.4 XHTML MP 编程技巧	99
5.4.1 大图片问题及其处理	99
5.4.2 使用更短的 URL	99
5.4.3 设置超链接的访问快捷键	99
5.4.4 XHTML MP 编程示例	100
第 6 章 PHP 动态页面语言基础	101
6.1 PHP 概述	101
6.1.1 PHP 简介	101
6.1.2 PHP 环境的搭建	101
6.2 PHP 语法	102
6.2.1 基本语法	102

6.2.2	类型	103
6.2.3	变量	104
6.2.4	常量	105
6.2.5	运算符	106
6.2.6	控制结构	106
6.2.7	函数	109
6.2.8	异常处理	110
6.3	PHP 面向对象技术	111
6.3.1	类和对象	111
6.3.2	构造函数和析构函数	112
6.3.3	类的继承	113
6.3.4	重写	114
6.3.5	抽象类与接口	115
6.3.6	多态	117
6.4	MySQL 简介	118
6.4.1	MySQL 初步入门	118
6.4.2	MySQL 基本操作	120
6.4.3	MySQL 集合函数	121
6.5	MySQL 与 PHP	122
6.5.1	连接与关闭数据库	122
6.5.2	创建数据库和表	122
6.5.3	插入数据记录	123
6.5.4	选择数据记录	124
6.5.5	更新语句	126
6.5.6	删除语句	127
6.6	PHP 框架与模板	127
6.6.1	Smarty 模板引擎	128
6.6.2	MVC 介绍	130
6.6.3	常用框架简介	131
第 7 章	JSP 动态页面语言基础	133
7.1	JSP 技术简介	133
7.1.1	JSP 页面的执行过程	133
7.1.2	JSP 的特点	134
7.1.3	JSP 与其他动态网页开发技术的比较	134

7.2 JSP 开发环境搭建	135
7.2.1 工具简介	135
7.2.2 JDK 的安装与配置	135
7.2.3 Tomcat 的安装与配置	136
7.2.4 JSP 页面实例	136
7.3 JSP 基本语法	137
7.3.1 JSP 页面元素	137
7.3.2 注释	137
7.3.3 声明	138
7.3.4 表达式	139
7.3.5 脚本段	139
7.3.6 JSP 指令	140
7.3.7 JSP 动作	142
7.4 JSP 内置对象	144
7.4.1 request 和 response	145
7.4.2 session	146
7.4.3 pageContext, application, config, page, exception	147
7.5 变量的范围	148
7.5.1 page 级别	148
7.5.2 request 级别	148
7.5.3 session 级别	149
7.5.4 application 级别	149
7.6 JSP 高级技术简介	149
7.6.1 JSP 与 JavaBeans	149
7.6.2 JSP 与 Servlet	150
7.6.3 数据库访问技术	150
7.6.4 Java Web 框架技术	150
第 8 章 Web 搜索引擎优化技术	152
8.1 搜索引擎	152
8.1.1 搜索引擎的工作原理	152
8.1.2 信息检索技术	153
8.1.3 搜索引擎排名算法分类	154
8.1.4 Google 的几种排名算法	154
8.1.5 HITS 算法	157

8.1.6	微软的 BrowseRank 算法	158
8.1.7	Alexa 流量排名算法	159
8.1.8	谷歌搜索引擎的服务趋向	159
8.2	搜索引擎优化原理与策略	160
8.2.1	搜索引擎优化原理	161
8.2.2	搜索引擎优化策略	161
8.2.3	恶意的搜索引擎优化技术	162
8.3	网站结构及其优化	162
8.3.1	服务器与域名选择	163
8.3.2	网站地图及其提交	163
8.3.3	蜘蛛搜索协议	165
8.3.4	链接优化策略	166
8.3.5	谨防链接作弊	171
8.3.6	链接工具简介	171
8.4	关键词优化策略与技巧	172
8.4.1	选择关键词的策略与原则	172
8.4.2	关键词选取方法	173
8.4.3	关键词布局技术	175
8.4.4	谨防关键词作弊	175
8.5	网页优化策略	175
8.5.1	合理设计头标签内容	176
8.5.2	网页内容	177
8.5.3	谨防使用页面框架	178
8.5.4	页面代码的优化	179
8.5.5	谨防内容作弊	179
8.6	Google 搜索引擎优化指南	180
8.6.1	导航优化	180
8.6.2	链接优化	181
8.6.3	页面优化	183
8.6.4	优化工具	184
8.6.5	全面推广网站	184
8.7	SEO 效果检测工具	185
8.7.1	SEO 效果检测和排名工具	185
8.7.2	SEO 综合工具	185
8.7.3	网站访问统计工具	185

第 9 章 Web 2.0 技术应用与展望	187
9.1 Web 2.0 简介	187
9.2 Web 2.0 理论和技术	188
9.2.1 六度分隔理论	188
9.2.2 XML	189
9.2.3 Ajax	189
9.3 Web 2.0 设计模式	189
9.4 Web 2.0 编程思想	190
9.5 Web 2.0 应用简介	191
9.5.1 博客与移动博客	191
9.5.2 播客	193
9.5.3 站点摘要	193
9.5.4 维基	194
9.5.5 引用通告	195
9.5.6 标签	196
9.5.7 网摘	196
9.5.8 社会网络	197
9.5.9 对等联网	197
9.5.10 即时通讯工具	198
9.6 Web 3.0 展望	199
9.7 Web 2.0 实训	199

第1章 概 论

1.1 计算机语言

语言是人类思维和交际的工具，也是人类保存认识成果的载体。语言的定义不一，如语言是传递信息的符号系统和人类最重要的交际工具；用于表达事物、动作、思想和状态的系统；作用于人与人的关系时，是表达相互反应的中介。语言作用于人和客观世界的关系时，是认识事物的工具；作用于文化时，是文化信息的载体。严谨地讲，语言是人类的交流和通信系统。书面语言用符号构造单词，单词的集合便是语言的词汇。单词的组合方式被定义为语言的句法或语法，而单词的意义称为语言的语义。

在计算机科学中，人类语言被称为自然语言。但计算机不足以完全理解自然语言，而必须有其专用语言，即计算机语言。计算机语言是人类为使计算机完成各种任务而设计和开发的。每种计算机语言都有其特定的关键字和组织程序指令的语法规则，即计算机语言让程序员能够准确地定义计算机所需要使用的数据，并精确地定义在不同情况下所应当采取的行动。

计算机语言有许多种，包括机器语言、编程语言、脚本语言和标记语言等。

1.1.1 编程语言

编程语言又称程序设计语言，是一套指导计算机完成具体任务的词汇及其语法规则。大多数被广泛使用或经久不衰的语言拥有负责其标准化的组织。按其发展，编程语言分为以下几代。

第一代语言是机器语言，主要特征是面向机器，即依赖于硬件环境。第一代语言是低级编程语言，其代表性语言是机器语言和汇编语言。

机器语言指计算机实际理解的语言，只包括数字。机器语言使用指令代码（绝对地址码、绝对操作码和存储空间分配），为机器直接接受，不需翻译。每种 CPU 有其唯一的机器语言。

汇编语言用符号写程序，需翻译成机器指令。用户指令集根据处理器体系结构不同而不同。它允许程序员用名称（指令集）代替数字，因此容易编写。

第二代语言是面向算法过程的高级编程语言，其主要特征是脱离机器。这类语言没有彻底摆脱硬件和操作系统的限制，即程序在不同类计算机上运行前，需要“移植”（调试和重新编译）。

第三代语言是结构化编程语言，具有描述过程和数据结构的能力，是高级编程语言。这类语言独立于操作系统。

第四代语言是声明性、交互性和非过程式的编程语言，如面向对象的编程语言。第四代语言具有如下特点：用户界面友好；编程将越来越简单化和专业化，类似于直接的口头或书面的人类自然语言指令。非专业程序员也能用它编写程序。

第五代语言指人工智能和神经网络。

这些语言各有特点，因此在解决问题时需要慎重选择。**Fortran** 适合处理数字型数据，而不好组织大量的程序。**Pascal** 适于编写结构严谨和可阅读的程序，但不灵活。**C++** 具有强大的面向对象特征，但因复杂而难以掌握。**Java** 是首选的面向对象编程语言，但书写代码比较繁杂，好在目前已经有一些界面和功能友好的开发工具，如 **Eclipse** 等。

无论使用那种编程语言编写程序，都需要把程序转换成机器语言，以便计算机理解和使用。有两种转化方法，即编译程序和解释程序。如果所使用的翻译机制是将所要翻译的程序代码作为一个整体翻译，那么这个翻译过程就被称为编译。因此，编译器处理人可阅读的程序文本（叫做源代码），输出可执行文件即机器语言，由计算机中央处理器直接运行。如果程序代码是在运行时才即时翻译，那么这种翻译机制就被称作解释。经解释的程序运行速度往往比编译的程序慢，但通常更具灵活性，因为它们能够与执行环境互相作用。

1.1.2 标记语言

标记语言是为处理、定义和表示文本而设计的语言。标记语言是一种将文档及其相关信息结合起来以展现文档结构和数据处理细节的编码。这种语言规定了用于格式文档布局和风格的代码即标签。

标记最早用于出版业，是作者、编辑以及出版商之间用于描述出版作品的排版格式所使用的。如今标记语言广泛应用于网页和网络应用程序中，如超文本标记语言和可扩展标记语言是众所周知的标记语言。

标记语言分为 3 类，即表示性标记语言、过程性标记语言和描述性标记语言。

描述性标识，源于传统出版业的“标记”手稿，即在原稿边缘加注一些符号来指示打印格式。长期以来，这项工作由专业人员以及校对人员来完成，对原稿标示出使用什么字体以及字号，然后再将原稿交给其他人进行手工排版。

表示性标记语言在编码过程中标记文档的结构信息。如在文本文件中，文件的标题需要用特定格式（居中，放大等）表示，这就需要标记文件的标题。文字处理以及桌面出版产品有时能自动推断出这类的结构信息，而纯文本编辑器不能解决这个问题。

过程性标记语言，一般专用于文字表达，通常对于文本编辑者可见，能够被软件依其出现顺序依次解读。比如，为了格式化一个标题，需要在标题文本之前插入一系列的格式标识符，以指示计算机转换到居中的显示模式，同时加大及加粗字体，并在标题文本之后会缀上格式中止标识。过程性标记语言被广泛应用于专业出版领域，专业出版商根据要求使用不同的指标语言以达到出版要求。

1.1.3 脚本语言

脚本语言，又称为动态语言，是一种编程语言控制软件应用程序。脚本通常以文本形式保存，在被调用时被进行解释或编译。脚本语言编程速度快、文件小、灵活性高、通常是解释执行。

以简单方式快速完成某些复杂的事情通常是创造脚本语言的重要原则。基于这项原则，脚本语言通常比 **C**，**C++** 或 **Java** 等高级编程语言简单。与由计算机处理器编译的其

他高级语言不同,脚本语言是在运行时由其他程序注释的高级编程语言,是为了缩短传统的编写、编译、链接和运行过程而创建的编程语言。

脚本语言的主要特性是:语法和结构通常比较简单,容易学习和使用,脚本编程速度更快,脚本文件明显小。脚本通常是解释执行的,速度可能很慢,运行时耗内存。程序的开发产能优于执行效能,即以执行效率为代价,但其优势在当前程序员工资趋高和硬件成本趋低时很明显。

早期的脚本语言经常被称为批处理语言或工作控制语言。随着互联网技术的发展和应用,脚本语言不断问世。面向 Web 的脚本语言通常嵌入页面文件内,为 Web 页面增加功能,如不同的菜单风格、图形显示或提供动态广告,或利用数据库数据动态生成页面。Web 的脚本语言分为客户端脚本语言和服务器端脚本语言。客户端脚本语言如 VB 和 JavaScript 等能影响浏览器中的数据,服务器端脚本语言如 ASP, JSP 和 PHP 等能利用数据库数据生成动态页面。

高级编程语言和脚本语言在许多方面互相交叉而没有明确界限。相比而言,脚本语言的特点是:介于标记语言和编程语言之间,其函数与编程语言比较相似,也涉及变量,语法和规则松散、简单,属于解释性语言,不需要编译,由脚本引擎(解释器)负责解释。

1.1.4 形式语言

在数学、逻辑和计算机科学中,形式语言是用精确的数学或机器可处理的公式定义的语言。如语言学中的语言一样,形式语言也有两个方面,即语法和语义。专门研究形式语言语法的数学和计算机科学分支叫做形式语言理论,而不涉及语义。

形式语言是一个字母表上的某些有限长字符串的集合。一个形式语言可以包含无限多个字符串。

语言的形式定义,字母表 Σ 为任意有限集合, ϵ 表示空串,记 Σ^0 为 $\{\epsilon\}$,全体长度为 n 的字串为 Σ^n , Σ^* 为 $\Sigma^0 \cup \Sigma^1 \cup \dots \cup \Sigma^n \cup \dots$,语言 L 定义为 Σ^* 的任意子集。

对语言的研究包括 3 方面,即语言的表示、语言的有穷描述性和表示结构。语言可以由文法产生,即文法是描述语言的一种模型。根据文法中产生式的形式,乔姆斯基把文法分成 4 种,即正则文法、上下文无关的文法、上下文有关的文法和短语结构文法,分别对应于 4 种语言。

构造文法的方法不一,难以用文法说明语言的特点,更主要的是难以识别语言的句子。因此,语言的识别模型显得至关重要。正则语言用有限自动机或正则表达式识别,有限自动机分析计算机高级语言的词法、字符串查询的基础。上下文无关的语言用下推机识别,这是分析计算机高级语言句子结构的基础。上下文有关的语言用线性有界自动机识别。短语语言用图灵机识别,图灵机是计算机理论研究、算法复杂度分析的基础。总之,形式语言是描述语言的语言,即一种元语言。

本书将重点介绍与 Web 相关的标记语言、脚本语言及其若干应用技术。

1.2 标记语言

1.2.1 SGML

标准通用标记语言(Standard Generalized Markup Language, SGML)是通用的组织和标记文档元素的语言,用来定义文献模型的逻辑和物理类结构。SGML 是国际标准化组织(ISO)于 1986 年发布的国际标准。SGML 本身没有任何具体的格式,而是明确了标记元素的规则,然后这些标记才能被解释成格式元素。

SGML 文档由 3 部分组成,即语法定义、文件类型定义(Definition Type Document, DTD)和文件实例。语法定义了文件类型定义和文件实例的语法结构,文件类型定义了文件实例的结构和组成结构的元素类型,文件实例是 SGML 语言程序的主体部分。

在实际使用中,每个 DTD 定义了一类文件。例如,所有新闻稿件都可用同一个 DTD。因此,人们习惯上把具有某一特定 DTD 的 SGML 语言称为某某标记语言,例如用于国际互联网的 HTML 语言。

SGML 被广泛用于管理那些面临频繁修改和需要用不同格式打印的大量文档。因为它是一个大型复杂系统,所以难以普及。然而 Web 的发展重新引起了人们对 SGML 的兴趣,因 Web 离不开 HTML,而 HTML 是根据 SGML 规则定义和注释标签的方法。

1.2.2 HTML

超文本标记语言(Hyper Text Markup Language, HTML)是 W3C 为方便作者创建网页和用户在网页浏览器中阅读信息而设计的标记语言,用于创建 Web 文档。HTML 借助标签和属性定义 Web 文档的结构和布局。HTML 允许页面作者标记其文档和在其中插入链接,语法不够严谨。

在 HTML 文档中可嵌入脚本语言代码如 JavaScript,以便实现动态显示功能。

1.2.3 XML

可扩展标记语言(Extensible Markup Language, XML)是 W3C 开发的一种针对 Web 文档的规范和语言,是 SGML 的简化版本。它允许设计者创建自己定制的标记、定义、转换、验证和在不同应用之间注释数据。

创建 XML 旨在简化 SGML 烦杂的结构,强化 HTML 过于简单而不够严谨的语法。虽然 XML 创立之初只被视作一项基础技术,但其发展早已超出设计者的构想。不论是学术界还是商业界都将其视为下一代网络的基石,XML 已经成为一股不可抵挡的技术潮流。

1.2.4 XHTML

可扩展超文本标记语言(eXtensible Hyper Text Markup Language, XHTML)表现方式与超文本标记语言类似,符合 XML 语法规则,语法严谨。

XHTML 是基于 XML 的标记语言,是扮演 HTML 角色的 XML。XHTML 在本质上是桥接(过渡)技术,融 XML 的灵活性与 HTML 的简单特性于一体(与后两者有交