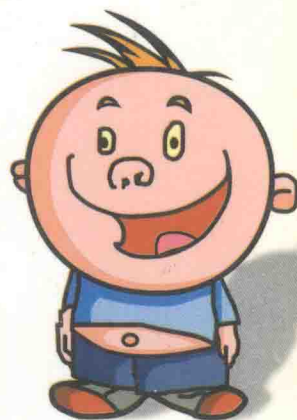


计算机编程起步 **ABC** 系列丛书

XML

编程起步

清宏计算机工作室 编著



计算机编程起步 ABC 系列丛书

XML 编程起步

清宏计算机工作室 编著



机械工业出版社

XML 是近几年来热门话题,而且随着 XML 以及相关标准的制定,使它日益成为各界关注的焦点。本书旨在向初学者介绍这一技术。

全书分为预备篇、开始篇和加速篇三大部分,共 11 章。预备篇主要介绍一些学习 XML 语言必备的知识;开始篇则向读者较为详细地介绍 XML 语言系统的语法知识和相关的类型声明 DTD、schema 和应用 CSS, XSL 进行 XML 的显示的知识,同时还有关于 HTML 数据岛的介绍;加速篇则是通过几个比较典型的例子来综合前两篇所学的知识。

本书在对语法的讲解中附有大量的示例以方便读者的学习,同时每章的最后附有一些典型的习题用来巩固读者所学到的知识。

图书在版编目(CIP)数据

XML 编程起步/清宏计算机工作室编著. —北京:机械工业出版社, 2002. 1

(计算机编程起步 ABC 系列丛书)

ISBN 7-111-09471-9

I. X... II. 清... III. 可扩充语言, XML-程序设计 IV. TP312

中国版本图书馆 CIP 数据核字(2001)第 079591 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

责任编辑:边 萌

责任印制:路 琳

中国建筑工业出版社密云印刷厂印刷·新华书店北京发行所发行

2002 年 1 月第 1 版第 1 次印刷

1000mm×1400mm B5·9.25 印张·360 千字

0 001—5000 册

定价:30.00 元 (含 1CD)

凡购本书,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) 68993821、68326677-2527

编 者 的 话

XML 是 W3C 组织基于以数据的建立和管理为目标的初衷定义的一种互联网上交换数据的标准。相对于师出同门的 HTML 来说,它具有易于编辑,不再限于标记语言的范畴,强调形式描述与内容描述的分离,赋予标记一定的语意,便于网上信息检索,世界通用,传输快速和链接方便等一系列的优势,所以一经推出即成为网络发展中倍受各界瞩目的新星。为此也吸引了大量的读者学习 XML。本书向初学者系统详细地介绍 XML 的相关背景 and 知识。

本书的第一部分为预备篇(第 1~2 章),主要介绍学习 XML 语言的必备的预备知识,包括 Internet 的历史和 XML 的出现与发展以及一些相关的 XML 编辑、浏览等工具,同时还会给出一个 XML 实例,使读者对 XML 有一个初步的认识。

第二部分为开始篇(第 3~10 章),是本书的主要部分。重点介绍 XML 的语法知识,以及其他相关的各项规范,主要包括 XML 的基本语法,用 DTD 实现数据结构定义,用 schema 实现元素定义、名称空间(namespace),用 CSS 实现 XML 的显示,XSL 实现 XML 的显示和数据处理以及数据岛等章节。

第三部分为加速篇(第 11 章),通过几个比较典型的例子综合在第二部分所学的知识。通过图书订单的例子,一个数据岛的例子和在第 1 章中提到的一个例子来展示 XML 的简单应用和开发的过程。

某些章节中有的问题用来复习和巩固读者在该章节所学习到的知识,并在本书最后的附录中对于部分问题给出简单的解答。相信有助于读者的学习。

本书配套光盘中含有此书中的全部源程序,这些程序都是经过严格的调试和测试的。读者可将其下载到硬盘上运行。多媒体演示能帮助读者更快、更好地了解 XML。

由于编者的水平有限及时间关系,书中难免出错,希望读者谅解并批评指正。

编 者

前 言

随着 21 世纪“软件蓝领”工程的启动，迫切需要对广大非计算机专业大专院校的学生以及高中毕业等就业人员进行培训，从而迅速为社会输送大批能够胜任较简单编程的工作人员（软件蓝领）。

为此，本工作室策划并编著了《计算机编程起步 ABC 系列丛书》，通过浅显生动的讲解，贴近实际的大量实例，以 ABC 逐步引导，使得零起点的读者通过短期培训、自学，即可胜任简单编程，并为更上一层楼打下坚实的基础。

内容编排分三部分：

预备篇（使读者初步了解编程，确立信心、激发兴趣）

简介开发工具、操作系统、应用领域和使用特点；完整实例的演示让读者初步了解开发工具的操作和编程的基本规则。

开始篇（使读者掌握基本概念、实用技术与技巧）

开发工具的安装、环境配置；生动形象地介绍基本概念和知识；以 ABC 逐步引导介绍实例的方式，使读者明白各功能技术用在哪里，如何应用。各章结尾的“问题与提示”，可使读者进一步巩固所学的知识要点。

加速篇（提高读者编程水平）

综合实例给出：功能概述、编程思想、技术要点、开发流程、实现步骤以及详细讲解，让读者初步了解和掌握较大型、全局性开发的概念和方法。

配套光盘提供生动形象的多媒体演示、实例检索、源代码下载，便于读者学习和实际应用。

本套丛书适用于广大编程初学者自学，也可作为软件蓝领人员的培训教程。

我们恳请广大读者对书中的不足之处提出宝贵意见。

清宏计算机工作室

目 录

前言

编者的话

预备篇	1	3.2 元素	36
第 1 章 XML 概述	2	3.2.1 起始和结束标记	36
1.1 Internet 的历史	3	3.2.2 空标记	37
1.2 XML 的出现和发展	3	3.2.3 字符数据	38
1.2.1 Web 语言的发展	3	3.2.4 元素的属性	38
1.2.2 XML 的出现	5	3.2.5 注释	39
1.2.3 XML 的好处	6	3.2.6 CDATA	41
1.2.4 XML 的发展前景	8	3.3 用 XML 存储数据	43
1.3 本章小结	9	3.4 本章小结	44
1.4 问题与提示	10	3.5 问题与提示	44
第 2 章 XML 准备	11	第 4 章 用 DTD 实现数据结定义	46
2.1 超文本语言 HTML	12	4.1 DTD 的声明和引用	47
2.2 XML 工具	19	4.1.1 DTD 的声明	47
2.2.1 编辑工具	19	4.1.2 外部 DTD 的引用	48
2.2.2 浏览工具	22	4.1.3 内外 DTD 的结合	50
2.2.3 转换工具	24	4.1.4 标签的使用方法	52
2.2.4 解析工具	24	4.2 数据类型	53
2.3 一个 XML 实例	25	4.2.1 #PCDATA	53
2.4 本章小结	30	4.2.2 空元素	53
2.5 问题与提示	30	4.2.3 ANY	53
开始篇	31	4.2.4 嵌套的选择结构	54
第 3 章 XML 的基本语法	32	4.2.5 混合类型的应用	55
3.1 XML 文档	33	4.3 ENTITY 的声明和引用	56
3.1.1 文档的结构	33	4.3.1 内部实体的设定与使用	56
3.1.2 文档类型声明	33	4.3.2 外部实体的设定与使用	58
3.1.3 文档的数据结构定义	34	4.3.3 参数型实体的设定与使用	59
3.1.4 文档的主体	34	4.3.4 实体的引用规则	60
3.1.5 文档的一致性和有效性	35	4.4 ATTRIBUTE	61
3.1.6 文档中的特殊符号	35	4.4.1 ATTRIBUTE 的设定	61

4.4.2	ATTRIBUTE 的类型.....	63	5.4.2	facets.....	93
4.4.3	ATTRIBUTE 的内定值.....	65	5.4.3	派生数据类型.....	95
4.5	本章小结.....	66	5.5	schema 的引用.....	102
4.6	问题与提示.....	66	5.6	本章小结.....	103
第 5 章	用 schema 实现元素声明 ..	68	5.7	问题与提示.....	103
5.1	XML schema 简介.....	69	第 6 章	名称空间	106
5.1.1	XML schema 是什么.....	69	6.1	为什么引入名称空间.....	107
5.1.2	XML schema 的优点.....	69	6.2	名称空间简介.....	108
5.1.3	一个 schema 的预览.....	71	6.3	名称空间的使用.....	110
5.2	XML schema 声明元素.....	73	6.3.1	声明名称空间.....	110
5.2.1	schema 根元素.....	73	6.3.2	名称空间的范围.....	111
5.2.2	用 element 声明简单的元素....	74	6.3.3	名称空间的缺省.....	111
5.2.3	声明包含子元素的元素.....	75	6.4	schema 中的名称空间.....	112
5.2.4	通过元素的引用的方法声明 元素.....	77	6.4.1	有效性的要求.....	112
5.2.5	声明空元素.....	78	6.4.2	schema 与名称空间结合 应用.....	113
5.2.6	声明混合内容的元素.....	79	6.5	本章小结.....	118
5.2.7	声明 any 内容的元素.....	80	6.6	问题与提示.....	119
5.2.8	元素出现的限制.....	81	第 7 章	用 CSS 实现 XML 的显示 ..	121
5.2.9	子元素的顺序.....	82	7.1	CSS 简介.....	122
5.2.10	选择型的子元素.....	83	7.1.1	什么是 CSS.....	122
5.2.11	声明元素组.....	84	7.1.2	CSS 结构和规则.....	123
5.3	XML schema 声明元素的 属性.....	85	7.1.3	一些常用的 CSS 属性.....	125
5.3.1	声明元素的属性.....	85	7.2	样式表套用.....	146
5.3.2	通过引用声明属性.....	87	7.2.1	套用的方法.....	146
5.3.3	声明 any 内容的属性.....	88	7.2.2	样式的继承性.....	148
5.3.4	声明属性分组.....	88	7.3	CSS 显示 XML 的优点和 局限.....	150
5.3.5	属性出现的限制.....	89	7.4	本章小结.....	151
5.4	XML schema 的数据类型.....	90	7.5	问题与提示.....	151
5.4.1	内建数据类型.....	90	第 8 章	用 XSL 实现 XML 的显示 ..	153

8.1 XSL 简介	154	9.2.3 元素属性值筛选	205
8.1.1 XSL 声明与套用	154	9.2.4 综合筛选	207
8.1.2 XSL 中应用 HTML	156	9.2.5 样板数据选择	209
8.1.3 在 XSL 中应用 CSS	158	9.3 内建函数	213
8.2 XSL 的数据处理元素	161	9.3.1 index	213
8.2.1 xsl:stylesheet	161	9.3.2 formatindex	215
8.2.2 xsl:value-of	162	9.4 本章小结	218
8.2.3 xsl:for-each	164	9.5 问题与提示	218
8.2.4 xsl:copy	168	第 10 章 用数据岛实现数据存储	
8.2.5 xsl:attribute	169	管理	219
8.2.6 xsl:element	171	10.1 HTML 与 XML 的结合	220
8.3 样版(TEMPLATE)的应用	173	10.1.1 HTML 中自定义标签	220
8.3.1 xsl:template	173	10.1.2 HTML 中使用数据岛	224
8.3.2 xsl:apply-templates	174	10.2 数据岛 (Data Island)	226
8.4 XSL 的其他元素	179	10.2.1 数据岛的对象和属性	226
8.4.1 xsl:comment	179	10.2.2 documentelement	228
8.4.2 xsl:node-name	182	10.2.3 childnodes	231
8.4.3 xsl:script	184	10.3 数据岛数据显示	234
8.4.4 xsl:eval	185	10.3.1 数据绑定	234
8.4.5 xsl:variable	187	10.3.2 多条数据的显示	236
8.4.6 xsl:sort	188	10.3.3 应用表格显示数据岛	239
8.5 本章小结	188	10.4 本章小结	243
8.6 问题与提示	189	10.5 问题与提示	244
第 9 章 用 XSL 实现数据处理	190	加速篇	245
9.1 标签、属性的引用	191	第 11 章 综合实例	246
9.1.1 标签和属性的一般引用	191	11.1 一个图书订单的例子	247
9.1.2 “*” 和 “//” 的使用	192	11.1.1 功能概述	247
9.1.3 标签集合的处理	196	11.1.2 编程思路与技术要点	247
9.2 数据的筛选	199	11.1.3 实现步骤及讲解	247
9.2.1 元素名称筛选	199	11.1.4 最终结果	251
9.2.2 元素内容筛选	202	11.2 数据岛	256

11.2.1 功能概述.....	256	11.3.2 编程思路与技术要点	274
11.2.2 编程思路与技术要点	256	11.3.3 实现步骤及讲解	274
11.2.3 实现步骤及讲解	257	11.4 本章小结.....	279
11.2.4 最终结果.....	263	11.5 问题与提示.....	279
11.3 回到最初的例子.....	274	附录 各章问题参考答案.....	280
11.3.1 功能概述.....	274		

预备篇

本篇包括第 1 章和第 2 章，简介开发工具、操作系统、应用领域和使用特点；完整实例的演示让读者初步了解开发工具的操作和编程基本规则。

本章导读

这一章主要是为学习 XML 做一些理论上的准备, 通过本章的学习, 读者将了解如下的知识:

- Internet 的发展历史
- Web 语言的发展和 XML 的出现
- XML 的应用和发展前景

相信细心的读者可以从本章学到不止上面列举的知识, 由于篇幅和作者能力有限, 感兴趣的读者可以通过其他各种途径获得更多的相关知识。

1.1 Internet 的历史

Internet 是在 ARPANET (美国国防部高级研究工程局网络) 的基础上发展起来的, 20 世纪 60 年代后期, 美国的建立的一个高级研究工程中心 (简称 ARPA) 构造了一个实验性的计算机网络, 叫做 ARPANET。这个网络使用叫做“包交换”的传输信息的新技术, 后来成为 Internet 的标准。在 1975 年, 美国军方加强了对网络的控制, 把它限制在军事用途的范围内, 这就驱动使用 ARPANET 的非军事机构决定组成他们自己的网络。

当 Internet 有很多献身者时, 还没有万维网, 没有点击 Internet 的办法, 没有观看图片的方法, 除非下载它们, 然后用图形程序打开它们。于是很多帮助人们分类、查找和找到信息的工具 (著名的包括 Archie、Gopher 和 Veronica) 出现了, 但是最大的突破出现在 1992 年, Tim Berners-Lee, 一个在 CERN (位于瑞士的欧洲粒子物理实验室) 工作的物理学家, 发明了 World Wide Web 来辅助他的物理研究。

Berners-Lee 使用了一种叫做 SGML 的标记语言, 使在网页中包含超文本链接成为可能。不管你是否理解计算机语言, 你都可以通过浏览器在信息世界里漫游。到 20 世纪 90 年代初期, 随着 WWW 的发展, Internet 逐渐走向民用, 由于 WWW 通过良好的界面大大简化了 Internet 操作的难度, 使得用户的数量急剧增加。许多政府机构、商业公司意识到 Internet 具有巨大的潜力, 于是纷纷大量加入 Internet, 这样 Internet 上的站点数量大大增长, 网络上的信息五花八门, 十分丰富。如今 Internet 已经深入到人们生活的各个部分, 通过 WWW 浏览、电子邮件等方式, 人们可以及时地获得自己所需的信息, Internet 大大方便了信息的传播, 给人们带来了一个全新的通信方式, 可以说 Internet 是继电报、电话发明以来人类通信方式的又一次革命。

1.2 XML 的出现和发展

1.2.1 Web 语言的发展

WWW (World Wide Web) 是一种建立在 Internet 上的全球性的、交互的、动态的、多平台、分布式的图形信息系统, 是建立在 Internet 上的一种网络服务。它的开发最初是为了在科学家之间共享成果, 使科学家们可以将科研成果以图文形式放在网上进行共享。但众所周知的是, 现在网络已经渗透到人们的日常生活、

工作等多个领域,成为人们获取信息的一种重要手段。因而 WWW 已远远超出了原设想,成为 Internet 上最受欢迎的应用之一。

Web 技术利用了一种称为超文本(Hypertext)的技术,这种技术使得用户可以从一个文件中的任何一点指向另一个文件的任何一点,从而可以实现快速的信息浏览。另外超文本技术具有良好的图形用户界面,使得用户能很容易地浏览因特网中的信息,它的最基本的概念就是 Hypertext Link,亦即所谓的超文本链接。而上述功能的实现,则要归功于计算机标记语言在 World Wide Web 的应用。这其中最为人们所熟悉的应该是 HTML(Hyper Text Markup Language),即所谓的超文本标记语言,它是一种基于建立超文本/超媒体文档的标记语言,它具有通用的语义,适合于表示各系统域的信息。

HTML 的应用十分广泛,它可用于描述超文本化的新闻、邮件、文档、超媒体文档、操作菜单、数据库查询结果、嵌入图形的结构化文档等,在网上看到的大部分网页,都是通过 HTML 来显示的。其实 HTML 是 SGML(Standard Generalized Markup Language)标准通用标记语言的一种应用。SGML 是一种描述语言的语言,它定义了以电子形式表示文本的方法,具有正式的、结构化的、可扩充等特点。在 20 世纪 90 年代初,World Wide Web 的发明者选择了一组 SGML 的 DTD 标记标签,在最早的 Web 浏览器和编辑器中,使用了这些标签和样式表进行排版,并增加了重要的功能超链接,这就是 HTML 的前身,所以说 HTML 是 SGML 的应用。正是这种应用,给我们带来了各种各样的美妙的网页。

20 世纪 90 年代 HTML 随着 World Wide Web 的爆炸性增长而流行于世界,HTML 语言本身也在不断地发展着,从静态文本到静态图像,从静态图像到动态图像,随后又加入了声音、影像、三维动画等。另一方面,用户已不满足于只是坐在那里被动地浏览,他们希望服务方在网页中提供更多的交互性,以便他们也可以加入 Web 游戏中。Java 语言以及另外一些脚本语言(Script)的出现满足了服务方和用户方的这种要求,比较流行的脚本语言有 Java Script 和 VBscript。它们为 Web 内容开发者提供了一种灵活简便的 Web 页面设计途径。万维网协会(W3C)的 HTML 工作组发布的,目前统一的版本是 HTML4.0。此版本除了文本、多媒体和超级链接功能外,还支持更多的多媒体选项、脚本语言、样式表、更好的打印功能和其他功能。

但是,目前多数人认为 HTML 还不稳定,另外,由于 HTML 对超级链接已支持不足,以及缺乏空间立体描述,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,没有时间信息,不能表示多种媒体的同步关系等,因此,这势必影响 HTML 的大规模应用以及用于复杂的多媒体数据处理,于是出现了扩展标记语言 XML、虚拟现实模型语言 VRML(Virtual Reality Modeling Language)和无线标记语言 WML(Wireless Markup Language)。关于 XML 将在

第1章 XML 概述

随后的几节中作详细介绍。

虚拟现实模型语言 VRML 是一种能在网络上使用的描述三维环境的场景描述语言,也就是说,HTML 定义二维,而 VRML 定义三维,从而使其在空间上变得非常广阔;又因其本身具有交互性,因而其对象都是处于空间中能彼此交换的。将现在的 Web 技术结合 VRML 就可将 Web 从二维提升到三维,从而可以达到一些令人难以想象的效果。从概念上讲,每个 VRML 文件都是一个基于时间的三维空间,它包含了可通过多种机制动态修改的图形对象和听觉对象,VRML 文件的一个重要特性是能够通过包含关系把文件组织在一起,并能够通过超链接把文件联系在一起。VRML 与 HTML 一样,用 ASCII 文本格式来描述境界和链接。总的来说,VRML 是 HTML 的扩展,基本的概念相类似,区别在于 HTML 文档是二维的,VRML 则带来三维空间的立体世界,这将会给 Internet 的发展带来更大的空间。

无线标记语言 WML (Wireless Markup Language) 是一种类似于 HTML 的浏览语言,可以提供浏览支持、数据输入、超级链接、文本和图像表现以及表格交互等。是应用于“无线应用软件”环境下的网页语言。用 HTML 写出的内容,我们可以在 PC 上用 IE 或是 Netscape 等浏览器进行阅读,而用 WML 写出的文件,则是专门用来在移动电话等一些无线终端的显示屏上显示,供人们阅读的。WML 同样也可以向使用者提供人机交互界面,接受使用者输入的查询等信息,然后向使用者返回他所想要获得的最终信息,它解决了 HTML Web 页不能在移动设备上显示的问题。WML 大约起源于 1998 年,现在正在慢慢普及。

随着网络的飞速发展,相信已有的 Web 语言会进一步发展,同时也会有更具有生命力的语言产生。

1.2.2 XML 的出现

近年来,随着 Web 的应用越来越广泛和深入,人们发现,HTML 过于简单且不够严密的语法阻碍了用它来表现复杂的形式。尽管 HTML 推出了一个又一个新版本,并且已经有了脚本(如常用的 VBScript, JavaScript 等)、表格、帧等表达功能,但始终满足不了不断增长的需求。另外,由于 HTML 对超级链接支持不足,多媒体能力较弱,影响了 HTML 的大规模应用以及用于复杂的多媒体数据处理,另一方面,由于近年来计算机技术的迅速发展,各种 Web 浏览器的不断产生,已经可以实现比当初发明创造 HTML 时复杂得多的 Web 浏览器。所以开发一种新的 Web 页面语言既是必要的,也是可能的。

有人建议直接使用 SGML 作为 Web 语言,这固然能解决 HTML 遇到的困难。但是 SGML 太庞大了,不利于学习;同时,开发一种可以完全实现 SGML 功能

的浏览器也比较困难。于是，Web 标准化组织 W3C，建议使用一种精简的 SGML 版本——XML 应运而生了。

W3C 组织于 1998 年 2 月发布了 XML 的标准。该组织制定 XML 标准的初衷是，定义一种互联网上交换数据的标准。W3C 采取了简化 SGML 的策略，在 SGML 基础上，去掉语法定义部分，适当简化 DTD 部分，并增加了部分互联网的特殊成分。在随后的三年时间中，W3C 组织不断推出新的相关标准，读者可以到 <http://www.w3c.org> 获得相关的标准。

1.2.3 XML 的好处

关于 XML 的优点有很多，通过与 HTML 的对比可以看出。在 XML 发布之前，国际互联网的发展受到 HTML 如下几个问题的束缚：

- HTML 无法描述数据内容，而这一点恰恰是数据检索、电子商务所必需的，用户希望在互联网上通过搜索引擎，快速准确地找到自己所需的信息，而这一切的实现主要依赖于对数据内容的完整准确地描述，HTML 很难做到这一点。

- HTML 对数据表现的描述能力是十分不够的，如 HTML 还不能描述矢量图形、科学符号等对象，目前仅仅是简单的通过把上述的对象转换为图像来表现，这种方法是远远不能满足需要的。

- HTML 实例标志语言的地位，完全不能适应对新标记需求的发展需要。

XML 的出现，使上述问题都得到很好的解决。

- XML 不再限于标记语言的范畴，它是一种元标记语言。XML 把 DTD 的定义权开放，不同行业可以根据自己的实际需求，定义描述内容的 DTD，以适应本行业内部的信息交流和存档需要，还可以定义专门描述形式的 DTD，丰富信息的描述效果。例如：CDF(Channel Definition Format) 是 Microsoft 开发的一种基于 XML 文件格式的描述频道信息的语言。CDF 是“PUSH”技术的一种引擎。有关它的标准可以在 <http://www.microsoft.com/xml/> 得到。IMS(Instructional Management Systems)是一种特定的支持在线教育标识元语言，它起源于广泛协作、会议请求讨论组和与开发研究相关群组。有关它的标准可以在 <http://www.imsproject.org/> 得到。

- XML 强调形式描述与内容描述的分离，一方面使 XML 文件的编写者，集中精力于数据本身，而不受显示方式的细枝末节的影响；另一方面允许为相同的数据定义不同的显示方式，从而适合于不同应用、不同媒体，使 XML 数据得到最大程度的重用性。通过后面的学习，可以看到对于一个 XML 文件，可以通过不同的方式来显示，本书将介绍常用的 CSS 和 XSL 来显示。目前，针对 XML 样式表开发的标准，被称为可扩展样式表语言 (Extensible Stylesheet Language)，

即 XSL。有几种 Web 浏览器的最新版本都能阅读 XML 文档,并采用合适的样式表,在屏幕上对信息进行分类和格式化。除了基于 XML 的网站运行速度更快、更易使用外,读者可能根本不知道他所看到的是 XML 文档而不是 HTML 文档。这种发布方式对视力有缺陷的人非常有利。样式表使他们可以将 XML 译为盲文或可听见的语音。这项功能还给想在汽车里冲浪的人带来了好处:他们会发现将页面转换为语音非常方便。

► XML 赋予标记一定的语意,加强了对内容的描述,便于网上的信息检索。用户可以自己定义有意义的标签,而不是像 HTML 那样的没有具体的意义的了。例如描述一个学生的信息,就可以一个 XML 文档来实现,在这个文档里可以定义一个<STUDENT>标签,关于学生的个人信息都放在这个标签中。在这个标签中,还可以定义另外的标签来存放具体的信息,如<NAME>存放学生的名字,<SEX>存放性别,<AGE>存放年龄,<ADDRESS>存放学生的住址,还可以定义一系列的标签来存放各种信息。只要从标签就可以看出标签中存放的内容,这么做的目的在于标示数据以便机器处理,特别强调数据的语义与元素之间的关系。对计算机来说,从 XML 文档中提取数据比从 HTML 文档来得容易。因此,当 Web 上出现更多的 XML 内容时,Web 搜索的精确程度也必定会有很大的提高。

► XML 将网络上传输的文档规范化,也方便了浏览器的设计。

► XML 具有世界通用的特性,XML 的统一能力源于它对 Unicode 新标准的依赖。这种编码标准,支持世界上所有以主要语言编写的混合文本。像大多数数字处理程序一样,在 HTML 中,文档通常是用一种特殊语言写成的,可能是英语、日语或者阿拉伯语。如果您的软件不能阅读某种语言的特殊字符,那么您就不能使用该文档。有时情况还会变得更糟:由于编码的不兼容,为中国台湾用户编写的软件,不能阅读中国大陆的文本。但是,能正确阅读 XML 的软件,就能处理这些字符集的任意组合。因此,XML 使得信息交换不仅能在不同的计算机系统之间实现,而且能跨越国界和不同的文化背景交换信息。

► XML 大大加快 Web 的响应时间。目前,连接在 Web 上的计算机设备,无论是功能强劲的台式机,还是小型便携机,所能做的不过是获得表单,填写后返回给 Web 服务器,直到完成一项任务。但是,随着 XML 中增加的结构化和语义化信息,这些设备将可以在现场进行大量的处理。这不仅能减轻 Web 服务器的许多负担,而且会极大地减少网络流量,从而大大加快 Web 的响应时间。

► XML 使超链接具有更多的功能。被命名为 XLink 的基于 XML 的超文本标准使您能够从目标页面列表中选择要链接的页面。其他类型的超链接可以在您点击的地方插入文字或图片,而不是迫使你离开页面。还有,XLink 所具有的更有效的处理、更准确的搜索、更灵活的链接等功能,将使 Web 结构产生革命性的变化,有可能产生全新的信息访问方式。用户将发现,这个新的 Web 比今天的

Web 更快、更强大、更有用。

总之, XML 具有易于编辑、便于管理、适于存档、容易查询等诸多优点, 成为网络发展中倍受各界瞩目的新星。

1.2.4 XML 的发展前景

从各界的广泛支持可以看出, XML 具有的上述优点决定了它的发展前景是一片光明的。

XML 目前已经包含在包括 Microsoft、Oracle 及 IBM 等公司的几乎所有软件之中。Microsoft 公司对 XML 显示了非凡的信心, 并称 XML 是知识管理系统的基石, Microsoft 公司的知识管理系统, 将作为附加工具与其客户端及服务器端软件一起发行。按照 Microsoft 公司对市场的解释, XML 可以统一多种语言、多种数据格式以及多种表现方式, 它将在所有产品中迅速普及。

Oracle 公司日前加大了对 XML 的支持力度, 着手开发从 Java 到 XML 过渡的产品, 并发布了支持 XML 的 Oracle 开发工具: 最新版本 JDeveloper 3.1 和 Oracle 开发工具包 (Oracle XML Developer Kit)。借助 JDeveloper 3.1, 开发人员能够便捷地创建、调试和部署可以处理 XML 数据和文档的 Internet 应用, Oracle 也因此成为业界第一个提供全面集成的、允许开发人员以 Java 和 XML 创建和部署电子商务应用的开发工具供应商。

网络服务软件供应商 Novell 公司, 日前发布了一个基于 XML 的全新企业应用集成技术的开放 beta 版本。Novell 的 XML 集成服务, 是 Novell “单一网络”战略的关键组成部分, 它使用开放的互联网标准提供数据中心功能, 能够有效地适应和调整网络电子商务进程。

将 XML 对象用作通用电子商务中最主要结构部分的工作也正紧锣密鼓地进行, CommerceNet 在以 XML 和对象技术为基础的电子商务体系结构 (eCoarchtectures) 方面从事的研究, 将对适用于家庭银行业务的开放式金融交易 (Open Finaccial eXchange: OFX) 和适用于 Internet 采购和销售的开放式贸易协议 (Open Trading Protocol: OTP) 等现有标准之间的数据交换产生重大影响。在计算机零售领域供应链应用的供应商 Rosettanet, 正在实施基于 XML 和对象技术的供应链过程。统一代码协会 UCCnet 也将采用 XML 和对象技术, 在供应链之间集成零售目录, 以促进分散用户之间的商业信息的交换。此外, 作为当今一大热点的 WAP 技术, 已经引起业界的关注, 而 WAP 是 XML 技术在无线移动通信上的应用, 可广泛建立在多种无线网络, (如 TDMA、CDMA、GSM 等) 上, 以及任意操作系统 (如 PalmOS、EPOC、Windows CE、FLEXOS、OS/9、JavaOS 等) 上, 它甚至提供不同设备在家庭之间的业务互通能力。最后, 面向网页制作