



“九五”国家重点电子出版物规划项目·希望计算机知识普及系列

Web应用

编程

丛书 5



网站开发新动力
用

XML

轻松开发 Web 网站

北京希望电子出版社 总策划
灯芯工作室 编写

本光盘内容包括：



1. 本版书中部分实例程序的源文件
2. 本版电子书



北京希望电子出版社
Beijing Hope Electronic Press
www.bhpe.com.cn



“九五”国家重点电子出版物规划项目·希望计算机知识普及系列

Web应用 **编程** 丛书5



网站开发新动力
用

XML

轻松开发 Web 网站

北京希望电子出版社 总策划
灯芯工作室 编写



本光盘内容包括:

1. 本版书中部分实例程序的源文件
2. 本版电子书



北京希望电子出版社

Beijing Hope Electronic Press
www.bhpe.com.cn

内 容 简 介

本书是一本专门介绍如何用 XML 开发 Web 网站的书。XML (eXtensible Markup Language) 作为网络发展新一代可扩展标识语言, 被誉为“继 Java 之后最激动人心的技术”。在 HTML 语言已为广大网络编程者熟悉的现在, 随着网络的成长, 需要更方便和完美的技术来满足个人设计的需要。与 HTML 相比较, XML 具有更强的兼容性和扩展性。用 XML 书写的文档更简洁也更易于执行, 使网络时代又走向了新的境界——全面包容、开放的程序设计。

全书由六部分(二十三章)组成。第一部分(XML 基础), 内容包括: 什么是 XML 语言, 新一代网络编程语言 XML, XML 的应用与发展前景, XML 工具; 第二部分(XML 的语言基础), 内容包括: XML 语法, XML 链接语言, XML 指针语言; 第三部分(XML 的实践), 内容包括: 创建 XML 文档, 创建属于自己的 DTD, 内容与形式的结合——XSL, XML DOM 技术, 同步多媒体合成语言 SMIL, ASP 与 XML 的联合开发; 第四部分(XML 高阶), 内容包括: XML 在数据库中的应用, XML 中的矢量图形处理技术, WML——无线接入的 XML, XML 与 Java; 第五部分(基于 XML 的 Web 站点应用与开发), 内容包括: WIDL 自动控制 Web 站点, 频道定义格式推送 Web 站点, Web 站点的设计实现; 第六部分(相关协议与标准), 内容包括: 可扩展标识语言 1.0(第二版)规范, XML 术语与词汇参考和 XML 技术动态等。

本书具有技术内涵高、指导性强, 内容新颖、丰富, 涉及面广, 范例实用性和可操作性强的特点。本书不但是从事网站开发与设计的广大从业人员重要的指导书, 同时也是高校相关专业师生教学、自学参考书和社会相关培训班推荐教材、各科研院所图书馆馆藏读物。

本光盘内容包括本书中部分实例程序的源文件和本版电子书。

系 列 盘 书: “九五”国家重点电子出版物——Web 应用编程丛书(5)

盘 书 名: 网站开发新动力 用 XML 轻松开发 Web 网站

总 策 划: 北京希望电子出版社

文 本 著 者: 灯芯工作室 编著

C D 制 作 者: 希望多媒体开发中心

C D 测 试 者: 希望多媒体测试部

责 任 编 辑: 马红华

出版、发行者: 北京希望电子出版社

地 址: 北京中关村大街 26 号, 100080

网址: www.bhp.com.cn

E-mail: lwm@hope.com.cn

电话: 010-62562329, 62541992, 62637101, 62637102, 62633308, 62633309

(发行和技术支持)

010-62613322-215 (门市) 010-62531267 (编辑部)

经 销: 各地新华书店、软件连锁店

排 版: 希望图书输出中心

CD 生 产 者: 北京中新联光盘有限责任公司

文 本 印 刷 者: 北京双青印刷厂

开 本 / 规 格: 787 毫米×1092 毫米 1/16 开本 28.25 印张 657 千字

版 次 / 印 次: 2001 年 2 月第 1 版 2001 年 2 月第 1 次印刷

印 数: 0001—5000 册

本 版 号: ISBN 7-900056-60-2/TP·59

定 价: 48.00 元(1CD, 含配套书)

说明: 凡我社光盘配套图书若有缺页、倒页、脱页、自然破损, 本社负责调换。

前 言

今天，网络的成长速度远远超过了单个人的发展，所以需要比 HTML 更完整的语言来表达网络内容的内涵。人们刚刚看到和学会使用 HTML 来表达自己的“超链接”式的想法，就发现这种语言已经远远不能满足网络潜在应用能力的需要。“扩展标记语言”（Extensible Markup Language, XML）的出现把网络表达的语言集合推进了一大步。

XML 是继 Java 以后 Internet 上最激动人心的技术。XML 的应用弥补了许多 HTML 的缺陷，我们把它在网上的应用总结为下列四点：

- 当网络客户必须在不同的数据库之间传递信息时的应用。
- 当需要把大部分从网络服务器下载的数据在用户端处理时的应用。
- 当相同的数据对于不同的用户需要有不同的界面时的应用。
- 当网络情报供货商要把发现的信息精心裁减，并发送给不同的个人用户时的应用。

利用 XML，我们可以实现动态内容生成、企业集成和应用开发；有利于信息的表达和结构性组织，从而使数据搜索更有效；有利于 Web 站点的维护与安全保密；有利于信息的共享与重用。

本书侧重于 XML 的编程，适合 XML 的中级开发人员，但对于 XML 的初学者也有很大的参考价值。我们希望任何一位对 Web 技术感兴趣的读者，都能够从本书中获取有用的信息，紧跟 Web 技术发展的潮流。

本书由六部分组成，主要内容包括 XML 的发展历程、XML 语法，并以大量的实例，介绍了 XML 程序的开发方法。在本书的第六部分，介绍了 XML 的词汇表、标准、规范、发展动态等相关的资源信息以便于读者在实际应用中参考。

本书由北京希望电子出版社总策划，灯芯工作室负责组织实施。参与本书编写校订工作的有唐宇、刘军辉、孙兆林、冯晨华、齐占杰、王世元、张成海、邱中等。由于时间仓促，水平有限，书中错误与疏漏之处在所难免，恳请广大读者批评指正。

灯芯工作室
2000.12



目 录

第一部分 XML 基础

第一章 什么是 XML 语言	3	2.6 小结	23
1.1 XML 的产生	3	第三章 XML 的应用与发展前景	24
1.2 XML 的基本概念	7	3.1 XML 行业的未来	24
1.3 XML 的结构	10	3.2 XML 带给网络应用的技术革命	24
1.4 关于 XML 的常见问题	13	3.3 XML 能够带给我们什么	25
1.5 小结	17	3.4 XML 在各个领域中的运用	27
第二章 新一代网络编程语言 XML	18	3.5 小结	30
2.1 HTML 现状	18	第四章 XML 工具	31
2.2 XML 与 HTML 的比较	19	4.1 XML 开发工具	31
2.3 XML 与 SGML 的比较	19	4.2 XML 浏览器	33
2.4 XML 的优势及其特点	19	4.3 小结	34
2.5 XML 的应用	22		

第二部分 XML 的语言基础

第五章 XML 语法	37	5.11 数据岛	61
5.1 XML 文档	37	5.12 XML 的相关标准	62
5.2 XML 声明	41	5.13 小结	65
5.3 注释	42	第六章 XML 链接语言	66
5.4 属性与标记	45	6.1 概述	66
5.5 实体参考	48	6.2 深入 XLink	72
5.6 CDATA 节	49	6.3 小结	78
5.7 文档类型定义 DTD	51	第七章 XML 指针语言	79
5.8 XML Schema	53	7.1 “第一次亲密接触” XPointer	79
5.9 名字空间	56	7.2 XPointer 规范及应用	80
5.10 正规有效的 XML 文档	58	7.3 小结	84

第三部分 XML 的实践

第八章 创建 XML 文档	87	8.3 XML 文档的格式良好性与有效性	95
8.1 建立一个简单的 XML 文档	87	8.4 一个 XML 文档实例	96
8.2 文档的组成要素	90		



8.5 小结	99	11.2 DOM 的核心结构	147
第九章 创建属于自己的 DTD	100	11.3 节点接口	149
9.1 什么是文档类型定义	100	11.4 使用 XML 解析器	150
9.2 文档类型声明	101	11.5 装载一个 XML 文档到 解析器中	150
9.3 判定 DTD 是否有效	102	11.6 XML 错误	151
9.4 元素声明	103	11.7 ParseError 对象与属性	151
9.5 DTD 中的注释	109	11.8 节点树	151
9.6 文档共享相同的 DTD	109	11.9 装载 XML 进入解析器	152
9.7 MATHML 的 DTD 文档	110	11.10 遍历 XML 节点树	152
9.8 小结	124	11.11 小结	152
第十章 内容与形式的结合——XSL 的应用	125	第十二章 同步多媒体合成语言 SMIL	153
10.1 XSL 概述	125	12.1 SMIL 是什么	153
10.2 理解 XSL	126	12.2 SMIL1.0 规范简介	154
10.3 构造结果树	129	12.3 SMIL DTD	154
10.4 样式表结构	130	12.4 SMIL 主要结构细节	157
10.5 模板规则与模式	131	12.5 SMIL 支持工具	159
10.6 模板	132	12.6 SMIL 实例	159
10.7 联合样式表	135	12.7 小结	160
10.8 XSLT 概述	137	第十三章 ASP 与 XML 的联合开发	161
10.9 对象格式化	144	13.1 三层 Web 应用程序简介	161
10.10 XSL 完整实例	145	13.2 Server-Side XML in ASP	166
10.11 小结	146	13.3 小结	170
第十一章 XML DOM 技术	147		
11.1 DOM 规范简单介绍	147		

第四部分 XML 高阶

第十四章 XML 在数据库中的应用	173	16.1 无线应用协议——WAP	254
14.1 XML=数据库	173	16.2 WAP 网页设计——WML 编程	262
14.2 来自业界的支持	196	16.3 小结	295
14.3 XML 查询语言——XQL	207	第十七章 XML 与 Java	296
14.4 小结	229	17.1 Web 技术双子座—— XML&Java	296
第十五章 XML 中的矢量图形处理技术	230	17.2 用 Java 创建 XML 文档	304
15.1 可伸缩的矢量图形 SVG	230	17.3 JSP+XML 平台	316
15.2 矢量标记语言 VML	247	17.4 小结	323
15.3 小结	253		
第十六章 WML——无线接入的 XML	254		



第五部分 基于 XML 的 Web 站点应用与开发

第十八章 WIDL 自动控制 Web 站点.....	327	19.2 CDF 文档规范及频道创建.....	335
18.1 概述	327	19.3 CDF 高级应用.....	341
18.2 规范与元素说明.....	329	19.4 小结	353
18.3 应用实例	333	第二十章 Web 站点的设计实现.....	354
18.4 小结	333	20.1 站点整体设计	354
第十九章 频道定义格式推送 Web 站点 .	334	20.2 站点创建	355
19.1 认识 CDF.....	334	20.3 小结	375

第六部分 相关协议及标准

第二十一章 可扩展标记语言 1.0 (第二版)		21.6 符号 (Notation)	405
规范.....	379	附录	407
21.1 绪论	380	第二十二章 XML 术语及词汇参考.....	416
21.2 文档	382	第二十三章 XML 技术动态.....	421
21.3 逻辑结构	389	23.1 XML 1999 技术动态.....	421
21.4 物理结构	397	23.2 XML 2000 技术动态.....	434
21.5 一致性 (Conformance)	405		

第一部分 XML 基础

第一章 什么是 XML 语言

第二章 新一代网络编程语言 XML

第三章 XML 的发展前景

第四章 XML 工具

原书空白页

第一章 什么是 XML 语言

本章将向读者介绍 XML 的产生背景、基本概念以及 XML 文档的结构，引导读者进入 XML 的神奇世界。同时还就一些关于 XML 的常见问题作了分析和回答。

本章包括以下内容：

- XML 的产生
- XML 的基本概念
- XML 的结构
- 关于 XML 的常见问题

1.1 XML 的产生

人类思想的交流和智慧的继承与语言文字标记是分不开的。当人类历史上文字符号产生的时候，就不只是一个单一的符号记录，而是由许许多多的记号组成的字符集。当文字产生时，人们在骨头、石头和木器上做标记，后来产生了纸和笔当成作记号的工具。当人们用笔和纸记录语言文字时，由于各种不同的应用需求产生了各种各样的文档格式，如书信、法律文件、各种公文、通知等等。不同的文档必须遵守各自的格式要求，文档中不同的地方有不同的标记来规定文档的格式，使不同的人阅读文档时不会把各种文档格式混淆。进入计算机时代，人们用字处理软件来完成文档的存储、格式编排和出版等工作，这些工具软件一般都是“所见即所得”的工具，掩盖了格式编排任务的复杂性。具体怎样进行格式编排、什么指令规定了标题的大小和字体，用户是不知道的。由于软件的这种标记过程注重的是视觉效果，常常依靠于计算机屏幕和打印机性能。SGML、HTML、XML 都是标志语言，利用各种标记使文档规范化格式化，以便于信息的整理与交流。这些标志语言的发展经历了一个发展过程。首先产生了 SGML，正是由于 SGML 的应用产生了 WEB 和 HTML 语言。由于 HTML 语言满足不了日益复杂的 WEB 信息交流的需要，因此，发展到了 XML 语言。

1.1.1 SGML 的产生

六十年代末，IBM 公司的 Charles Goldfarb、Ed Mosher 和 Ray Lorie 为了解决由不同专用格式创建的法律文件不能在公司不同部门间相互移植的问题，提出了任何可互操作系统的三个主要要求：

- 文件处理程序需要支持一个公共文档格式。
- 这个公共格式对特定的领域是专用的。
- 为了达到高度可靠性文档格式必须遵守特殊规则。

这个文档格式编排系统称为通用标记语言（Generalized Markup Language, GML）。在以后的八年时间中，Goldfarb 继续研究 Generalized Markup Language，并担任了一个国际开发人员组织的技术领导，终于建立了所谓的标准通用标记语言（Standard Generalized Markup Language, SGML），而且国际标准化组织 ISO 于 1986 年最终批准采用 SGML。

SGML 提供描述文档和创建新的一致性衡量准则所必需的公共框架。几乎所有处理文档的语言，都可以至少部分追溯到 SGML。诚然，如果不实行这种描述标准，Web 技术就不存在。SGML 可以用下列三条来描述其特点：

- 不提倡一种特殊的文档结构。
- 不存在必须使用的有限标注集。
- 不限制创建新文档的标准的潜力。

SGML 避免了其它文档格式共同的表示特征，而集中于信息的结构。它不提倡一种特殊的文档结构，而是要求定制数据。只通过结构而不是表现形式约束数据，就能以任何方式迅速编辑、检索、分类、转换或汇集数据，并传递给任何媒体、打印机或屏幕等。正是由于 SGML 的产生导致了超文本标志语言 HTML 的诞生。

1.1.2 Web 和 HTML 语言的诞生

Tim Berners-Lee 是设在瑞士日内瓦的欧洲离子物理实验室的一位英国研究人员。1989 年，他对该研究中心的大量研究资料的不可移植和不可兼容深感失望，并构想出这样一个世界：在其中存取数据是一项简单的任务，以一种一致的方式进行，不考虑使用什么终端和程序，包含任何地点的任何计算机应能只是用一个简单的和常用的程序存取数据。他意识到传统的线性思考不能达到这种跨越系统表达信息的要求，于是导致了超级链接概念的产生。随着 Berners-Lee 建立的超文本标志语言（Hypertext Markup Language, HTML）的开发成功，形成了今天的 Web 信息存储和交流技术。

1.1.3 从 HTML 到 XML

今天，网络的成长速度远远超过了单个个人的发展，所以需要比 HTML 更完整的语言来表达网络内容的内涵。人们刚刚看到和学会使用 HTML 来表达自己的“超链接”式的想法，就发现这种语言已经远远不能满足网络潜在应用能力的需要。“扩展标记语言”（Extensible Markup Language, XML）的出现把网络表达的语言集合推进了一大步。HTML 只是一种表达的技术，它并不一定能揭示 HTML 标记中的含义。举一个最简单的例子：<h2>Pine</h2>这句话在网络浏览器中有特定的表现。但是 HTML 却并没有告诉我们它到底是什么。Pine 只是一个英文单词罢了，它在不同的环境之下可能会有不同的意义。是一个计算机公司，一棵松树，还是一个人？HTML 并没有告诉我们 Pine 的具体内容。HTML 中有一个大问题就是它的标记的集合是固定的，用户不能新增有意义的可以让其他人使用的标记。网络浏览器是一个应用平台，以 HTML 作为数据标准，网络上的应用程序依赖于服务器上的 CGI 脚本或 ASP 程序等来处理网页上的数据。这样的结构是非常清晰的，但 HTML 却使在服务器一端要处理的数据量过大。这使得网络速度低下，效率不高。

SGML 是 HTML 技术的前身。它是文件和文件中信息的构成主体。SGML 与 HTML 不同，它允许用户扩展标记集合，允许用户建立一定的规则。SGML 所产生的标记集合是用来描叙信息段特征的，而 HTML 仅仅只是一个标记集合，所以我们可以说，实际上 HTML 只是 SGML 的一个子集。

XML 开发者源于 SGML 的设计和应用者。他们在 SGML 上投入了大量精力，但是却发现 SGML 并没有完全发挥作用。从下面几个方面就可看出。

1. 对 EDI 的支持

EDI 就是电子数据交换,它是网络发展的一个主要目的市场。结构化信息的一个主要目的就要使数据交换成为可能。不同的行业都制定了本行业统一的模型,就像是不同的国家有着不同的语言,这便于本国文化的交流。不同的行业内部信息用统一的模型标识,就能够使信息方便、高效地共享。这样一个统一的模型就是文件类型定义(Document Type Definition, DTD)。当然 DTD 已经落伍了,它正被 XML 的模式(Schema)所替代。很明显的,网络是一个理想的电子数据的集散地。很显然,网络环境下 HTML 是一种有缺陷的数据形式,因为 HTML 不能完全表示不同行业中所需的不同的模型和它的语义。能不能有一种新的语言来解决这个问题呢?答案就是 XML。

2. 对 Java 技术的支持

Java 技术是本世纪最重要的技术发展之一。Java 使浏览器工作时就像在通用的应用平台上,而平台与平台之间却是独立的。但固定的标志集合和 HTML 语义上的贫乏使得 Java 的应用受到了极大的限制。正如前面提到的,在 HTML 中不同的语义无法表现,因此数据元中丰富的信息得不到一种统一的表示,而 XML 却能完全胜任这份工作。HTML 页面要依赖网络服务器上的 CGI 脚本来表现几乎每一个编程函数,这显然使服务器的工作量过大。有了 XML 和 Java 技术,更多的应用软件处理起来将不会占用多少网络通信量。这使得网络更加快捷,客户可以同时应用多个应用软件。XML 真正使 Java 有了用武之地。

3. 信息独立于平台之间

SGML 作为 HTML 和 XML 的前身技术,一直是一种平台之间互相独立的信息技术。这便于指定信息语义的结构。当企业正忙于开发各种各样的信息格式时(比如微软的 RTF, Adobe 的 PostScript 和 MIF, 以及 WordPerfect, Lotus, Borland 等公司的格式), SGML 已先人一步,建立了一套严格一致的、独立于平台之间表达信息的格式。但在 80 年代,正当 SGML 悄然兴起时,绝大多数工业计算机的开发者都把目光集中在新的计算机平台上。人们并没有意识到多种私有信息格式可能带来的麻烦。一直到了 90 年代,网络技术崛起以及 Internet 的迅速流行之后,人们才清醒过来,试图寻找一种解决办法。

4. 网络上的 SGML

SGML 已不适用于网络社会的需要。“如何使 SGML 能成功地运用于网络”这一问题已自然而然地提上了议程。1996 年 8 月, GCA(图形通信协会)在西雅图召集 SGML 开发者们举行了一次会议。会议由 Sun Microsystems 公司的 Jon Bosak 主持。论题集中于两大方向:

(1) 在软件应用中 HTML 不是一种理想的信息表现格式。讨论的结果是有必要把 SGML 应用于网络。

(2) SGML 的某些方面已经落伍了,并已成为它广泛传播的障碍。讨论的结果是有必要考虑怎样修补 SGML。

既然 SGML 有着多重语义的标记集合,它早就应出现在网络上了。然而,在 1996 年 8 月,在网络上人们热衷的却并不是 SGML,而是 HTML 和它的固定的标记集合。SGML 的开发者们立刻制定了一个紧急修改 SGML 标准的方案。因为 SGML 是一个严格而完整的系

统，方便软件应用并不是它的首要任务。所以在 SGML 中有许多语法语义标准。它们既不方便而且消耗昂贵。它们必须被修改或是删除。SGML 开发者们首先做的工作就是得到一个 SGML 中可移给网络的、非关键的结构信息的清单。他们可以基于这个清单对 SGML 进行修改。

5. XML 的诞生

早在西雅图会议之前，Bosak 和一些 SGML 结构信息专家就已向 W3C 提出了“网络上的 SGML”计划。W3C 支持并赞助了他们的努力，工作于 1996 年 7 月正式开始启动。工作的早期，有较大的阻力。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。工作组（原称“SGML 编辑审议委员会”）并未退缩。他们打算让 SGML 以全新的面目出现在网上，给 SGML 以全新的面貌，因此将其命名为“可扩展标识语言（eXtensible Markup Language）”，即 XML。

工作组制定了一个雄心勃勃的计划来展示 XML 的特色，计划的实施分三部分：

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

XML1.0 版本标准由 W3C 于 1998 年 1 月 10 日正式批准公布，XML 1.0 第二版于 2000 年 10 月 6 日发布（XML 1.0 第二版不是 XML 的一个新版本，它只是为了方便读者，并入了第一版勘误表中指出的错误和修改）。XLL 和 XSL 的工作还处于进一步修订和完善的阶段。

6. XML 的应用

最初 XML 的目标是让各种结构的文件都作为统一的网络文件的一部分在网上传输。过去这些文件是用 HTML 实现的，HTML 允许指定明确的元素类型说明，比如特定的商品标号，文档标识，或是可测量的数值。和 HTML 相比，XML 允许客户定义他们自己的文件元素集合，同时也可以指示这些元素在屏幕上如何按指定的要求表现出来。在早期，为了解决怎样在固定的目标之间传输数据元，XML 定义为一种自然的编码形式。在一些方案被考虑之后，一种称为 RDF（资源描述框架）的方案受到了欢迎。RDF 为 XML 提供了数据元编码定义，这就象是一个公用的翻译器，为不同的固定目标之间的数据提供翻译。

XML 将支持更加专业的数据语言，比如说 OSD（开放软件描述）。OSD 是由 Microsoft 和 Marimba 提出的一种新的格式描述语言。在这种格式下，软件在网上能实时检查，实时刷新版本，而不是等用户自己更新，或由软件提供商提供类似的服务。当 OSD 内嵌于 XML 支持的 CDF（频道定义格式）中时，OSD 更能使支持频道的桌面自动地更新。

正如前面提到的，XML 的一个主要目标市场是电子商务。传统 EDI 机制依靠不同商业之间的强大计算机系统来实现压缩的信息传输。每一条信息在传输使用，提供给用户之前都必须编码。电子商务在网上运作时，用户端每填完一个 HTML 的表格之后，都要把表格发还给初始的服务器处理。产品交易、谈判签约、后勤管理、税收报表等等活动的数据处理都集中在了服务器一端。可以预测到，有了 XLL 所链接的行为控制机构和 XSL 所提供的客户端评价功能，将来的数据可以从屏幕上抓取，有必要的可在客户端处理。在处理数据时，传输给相关用户而不必要改换数据格式。一个类似的协议是 OTP（开放网络贸易协议），

它的草案最初是于 1998 年 1 月发布的。这个协议的制定是为了满足在网上消费者和销售者之间交易时信息的传输，它同时也允许第三方，比如说供货商、市场评估机构、消费者保护机构等来参与使用。

XML 的应用弥补了许多 HTML 的缺陷，我们把它在网上的应用总结为四点：

- (1) 当网络客户必须在不同的数据库之间传递信息时的应用。
- (2) 当需要把大部分从网络服务器下载的数据在用户端处理时的应用。
- (3) 当相同的数据对于不同的用户需要有不同的界面时的应用。
- (4) 当网络情报供货商要把发现的信息精心裁减，并发送给不同的个人用户时的应用。

1.1.4 XML 的设计目标

(1) XML 可在因特网上直接使用：这是最重要的一点，设计 XML 的出发点，就是针对 Internet 的迅速发展以及 HTML 的缺点，而寻求一种新的解决方案。

(2) XML 将支持各种各样的应用程序。

(3) XML 应与 SGML 兼容：对于 XML 和 HTML 前身的 SGML，必须提供向前兼容性。

(4) 编写处理 XML 文档的程序将十分方便。

(5) XML 中可任选的特征数目极少，理想情况为零。

(6) XML 文档应当清晰易读。

(7) XML 设计应能迅速完成。

(8) XML 设计应条理清晰、简明扼要。

(9) XML 文档应易于创建。

(10) XML 标记可以是复杂的标记。

1.2 XML 的基本概念

1.2.1 元素

元素就是文档的组件，元素可由其他元素、其他类型数据等组成。

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，相互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 文件是用于储存数据的，因此它一定要有内容，而它的内容则是在各个元素之中进行声明。在这里，内容可以是空白内容，也可以是简单的文本或单一数据，也可以是复杂的多个数据列，甚至还可以是其他的分类子元素。

XML 元素命名规则：同 Java、C 等命名规则类似，可以是英文字母或中文作为元素名（也可以是其他语言，但要首先确保数据读取软件可以支持该语言），可以用数字和字母作为名字开头，需要注意的是名称中不能出现空格！可以用下划线“_”取代空格的功能来进行详细命名，而且在名称中严禁出现“:”号。

1.2.2 属性

属性是包含关于元素的额外信息的元素部分。

那么，属性值可以做什么呢？XML 中的属性值可以被用来为元素添加额外的说明信息。其实我们很早以前就已经接触过属性值，而且频率非常得高，那就是出现在 HTML 中的属性值。如：

```
<font size="7" color="red"></font>
```

标记中 size 和 color 就是 font 的属性。在 XML 中，我们也可以应用属性值到元素，但不要频繁，毕竟 XML 是用来储存和发送数据信息的，因此它的可扩展性就显得非常重要。我们可能随时需要向 XML 文件中添加数据，虽然使用属性值可以方便地为元素添加额外的信息说明，但是这样做非常不利于日后的维护和更新。更何况，这些事情使用元素来做才是正确的选择。频繁的使用属性值很难维护和更新。而且，使用属性值还有一系列的弱点：

- (1) 属性值不能包含多重数值（元素可以）。
- (2) 属性值难于扩展（考虑到长远打算）。
- (3) 属性值不能用于描述结构内容（子元素则可以）。
- (4) 属性值很难通过 DTD 来进行一个测试。

1.2.3 文档类型定义

文档类型定义（Document Type Definition, DTD）是一套关于标记符的语法规则。它告诉我们可以文档中使用哪些标记符，它们应该按什么次序出现，哪些标记符可以出现于其它标记符中，哪些标记符有属性等等。DTD 原来是为使用 SGML 开发的，它可以是 XML 文档的一部分，但是它通常是一份单独的文档或者是一系列文档。因为 XML 本身不是一种语言，而是定义语言的一个系统。它没有象 HTML 一样拥有一个通用的 DTD，相反，想使用 XML 进行数据交换的行业或组织可以定义它们自己的 DTD。如果一个组织想用 XML 来标识仅在内部使用的文档，它可以创建自己私有的 DTD。

1.2.4 XML 模式

一些供应商，包括微软，已经提议替代 DTD 的一个方法，称为模式（Schema）。他们已经将其以 XML 数据提交给了 W3C。就像 DTD，schema 提供了文档的规则，并指出用什么标记符，标记符的属性之间的联系等等。但是，不同于 DTD，schema 可以定义数据类型。例如，DTD 可能有一个标记符<PRICE>，而标记符之间的内容可以是数字或字符串。Schema 可以规定只输入数字。这个方法显然有其优点，唯一的问题是它将成为 DTD 规范，还是 XML 的一个扩展。

1.2.5 XML 名字空间

在 XML 中，用户可以自己定义标记和元素。但我们可以很容易想到，如果把多个 XML 文件合并为一个时，就很可能出现冲突。名字空间（Namespaces）就是为了解决这个问题的。对于 XML，Namespace 严格的定义是：Namespace 是用 URI 加以区别的、在 XML 文件的元素和属性中出现的所有名称的集合。有了 Namespace，用户就可以保证在他的文件中使用的名称是独一无二的。在没有 Namespace 的 XML1.0 文件中，元素和属性中出现的名称无异于一族没有结构的字符。我们称它们为 local names（本地名称）。本地名称在网络上是不合

适的。可以想象，网上会有成千上万的人使用同一个名称，而它们却代表了不同的意义。Namespace 通过 URI 区别同名的标识，那么我们完全可以相信不会出现冲突了，因为 URI 是独立的。Namespace 的声明要用到前缀 xmlns。声明的 Namespace 位于指定的 URI，同样的，它也有个名字，我们称这个名字为：Namespace 名，Namespace 名必须独有而且一致。

1.2.6 层叠样式表 CSS 与可扩展样式语言 XSL

在 XML 中内容与表现形式是分开的。这就使得不同的用户可以根据他们自己的需要来定义数据的表现形式。在一个 XML 的源文件中并没有关于表现形式的信息。这一点我们可以从下面的例子中看出：

例 1.1 在 HTML 中的一段代码：

```
<H1>test document1</H1>
<H2>test document2</H2>
<H3>test document3</H3>
<H4>test document4</H4>
<p>test document5</p>
```

这是一段典型的 HTML 代码。在每一个标签中都含有表现形式的信息，而关于标签之间联系和结构的信息却没有，这事实上约束了 HTML 的发展。

例 1.2 在 XML 中的一段代码：

```
<Car Register>
<Car>
  <Registration Number>ABC123 </Registration Number>
  <Make>Saab 9000</Make>
  <Model>1995 </Model>
  <Owner>
    <Name>jone </Name>
    <Address>address</Address>
    <Zip code> Zip code </Zip code>
    <City> City </City>
  </Owner>
</Car>
</Car Register>
```

在 XML 文件之中，标签的信息是非常“纯”的，它没有表现部分，所以有人又称 XML 为传输知识的语言。那么 XML 文件是怎样进行表现的呢？XML 文件的所有表现信息多放在了样式表（stylesheet）文件当中。样式表文件全权负责 XML 源文件的表现形式。所以说如果一个 XML 源文件对应不同的样式表文件就会有不同的表现形式。有了样式表文件，我们可以对文件表现型的大小、颜色、空白做特定的规定。CSS（Cascading Style Sheets，层叠样式表）和 XSL（XML Style Language，可扩展类型语言）是两种样式表语言。它们是互相补充、各有特色的。CSS 可以展现 HTML 和 XML 文件，而 XSL 可以展现 XML 和转型语言（transformation）。

1.2.7 XML 链接

当人们想到 Web 时，可能就会想到引人入胜的超链接。但是 HTML 只能向一个方向进

行, 从用户所在的页面进入用户想去的页面。没有办法链接到页面的特殊部分, 也无法轻易地创建外部页面之间的链接。XML 的链接语言就能做到许多新功能的链接。三个 W3C 工作草案文件规范了 XML 的链接和寻址机制, 即 Xpath 语言、XLink 语言和 XPointer 语言。

Xpath: 主要目的是进行部分 (而不是整个) XML 文档的实际寻址。

XLink: 利用 XML 句法创建结构, 来描述目前的 HTML 的单向链接以及更复杂的多端和多类型的链接。XLink 的重要部分是定义两个或多个数据对象或对象部分之间的关系。

XPointer: 建立在 Xpath 的基础上, 以支持到 XML 文档的内部结构中的寻址。因此, 可以使用 XML 标记联接到另一个文档的特殊部分, 而不必提供 ID 引用。

1.2.8 文档对象模型

XML 文档对象模型 (Document Object Model, DOM) 结构将 XML 文档的内容实现为一个对象模型。W3C 的 DOM Level 1 说明定义了 DOM 结构如何实现属性、方法、事件等等。微软的 DOM 实现完全支持 W3C 标准, 并且还有许多使程序更容易访问 XML 文件的新增特点。

要使用 DOM, 需要创建一个 XML 解析器的实例。微软公司在 Msxml.dll 中创建了一系列标准的 COM 接口来使创建实例变得可能。Msxml.dll 中包含了类型库和可应用的代码, 用户可以使用其处理 XML 文件。如果用户使用可实现脚本的客户端, 比如 VBScript 和 IE, 就可以使用 CreateObject 方法来得到解析器的实例。

```
Set objParser = CreateObject("Microsoft.XMLDOM")
```

如果使用 ASP, 则使用 Server.CreateObject 方法。

```
Set objParser = Server.CreateObject("Microsoft.XMLDOM")
```

如果用户在使用 VB, 则可以创建一个对 MSXML 类型库的引用, 这样就可以访问 DOM。要在 VB6.0 中使用 MSXML, 操作如下:

(1) 打开 Project References 项。

(2) 从 COM 对象中选择 Microsoft XML, version2.0, 如果找不到此项, 则需要重新安装获取它。

这样我们就可以创建一个解析器的实例了:

```
Dim xDoc As MSXML.DOMDocument Set xDoc = New MSXML.DOMDocument
```

用户可以通过两个途径获取 Msxml.dll:

- 安装 IE5.0, MSXML 解析器是其中集成的部件。
- 或者到相关网站上下载。

一旦建立了对类型库的引用, 就可以执行解析, 调入文档; 用户从而可以对 XML 文档进行表现和处理。

1.3 XML 的结构

XML 语言, XML 名字空间和 DOM 是 W3C 建议的, 这是 W3C 发展过程中决定性的一步。由于它们已经成为正式的规范, 开发人员能够用 XML 的格式标记和交换数据。XML 在三层架构上为数据处理提供了很好的方法。