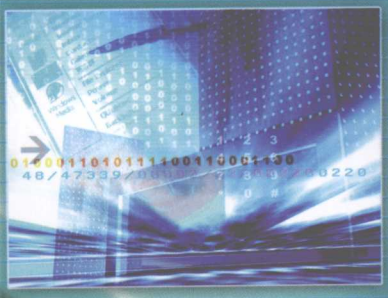
普通高等教育“十一五”规划教材

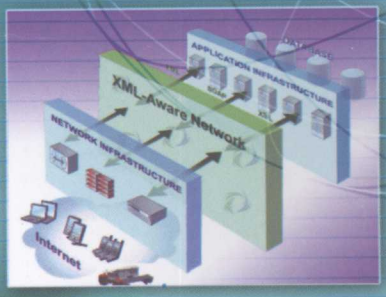
高等院校计算机技术系列教材

XML编程原理与实例教程

刘怀亮 主编

蔡 沂 编著





冶金工业出版社

普通高等教育“十一五”规划教材
高等院校计算机技术系列教材

XML 编程原理与实例教程

刘怀亮　主编

蔡　沂　编著

北　京

冶金工业出版社

内 容 简 介

本书是根据普通高等教育“十一五”规划教材的指导精神而编写的。

XML 是由 W3C 定义的一种可扩展标记语言，其使用越来越普及，越来越多的领域和环境下都采用 XML 来实现需要的功能。本书结合实例，详细地描述了 XML 的基本概念、与 XML 相关的各种标准和实现的技术以及 XML 数据库的相关知识，最后是通过一个实际的例子来介绍 XML 的开发。本书强调通过实践来掌握 XML 的基本概念、相关知识和具体应用，每一个知识点都相应地利用具体的例子来阐释，给出例子的运行结果，以期给读者一个清晰的展示。

本书可作为高等院校本、专科计算机及其相关专业的教材，也可作为从事计算机软件开发人员的参考书。

图书在版编目（CIP）数据

XML 编程原理与实例教程 / 刘怀亮主编；蔡沂编著.
北京：冶金工业出版社，2007.4
普通高等教育“十一五”规划教材
ISBN 978-7-5024-4256-9

I. X… II. ①刘…②蔡… III. 可扩充语言，XML—程序设计—高等学校—教材 IV. TP312

中国版本图书馆 CIP 数据核字（2007）第 044738 号

出版人 曹胜利（北京沙滩嵩祝院北巷 39 号，邮编 100009）
责任编辑 肖放
ISBN 978-7-5024-4256-9
广州锦昌印务有限公司印刷；冶金工业出版社发行；各地新华书店经销
2007 年 4 月第 1 版第 1 次印刷
787mm × 1092mm 1/16； 18 印张； 415 千字； 280 页
28.00 元
冶金工业出版社发行部 电话：（010）64044283 **传真：**（010）64027893
冶金书店 地址：北京东四西大街 46 号（100711） 电话：（010）65289081
（本社图书如有印装质量问题，本社发行部负责退换）

前　言

一、关于本书

本书是根据普通高等教育“十一五”规划教材的指导精神而编写的。

XML 是由 W3C 定义的一种可扩展标记语言，其出现使得在 Internet 上进行数据交换变得更加方便。对一个文档进行标注，使得文档的内容变成一种结构化数据，结构化是 XML 的最大特点，同时也能够实现对不同数据源的数据的无缝集成，提供对同一数据的多种处理方法。XML 的使用正越来越普及，越来越多的领域和环境下都采用 XML 来实现需要的功能。

本书作者从事计算机研究工作多年，在 XML 的应用上具有丰富的经验。本书结合实例，详细地描述了 XML 的基本概念、与 XML 相关的各种标准和实现的技术，最后还讲述了一些 XML 数据库的相关知识。每一个知识点都相应地利用具体的例子来阐释，给出例子的运行结果，以期给读者一个清晰的展示。

二、本书结构

全书分为 11 个章节，从 XML 的发展历程与应用前景开始，详细介绍了 XML 的基本概念与构成，接着介绍了 XML 中的一些高级应用，最后通过一个实际的例子来介绍 XML 的开发。本书由浅入深，一步一步将 XML 知识介绍给读者，并通过最后一章的例子让读者融会贯通，全面地掌握 XML 的原理，并获得必要的实践经验。

第 1 章：XML 概述。讲述 XML 产生的背景及发展过程，还有 XML 的基本概念和 XML 的相关应用。

第 2 章：XML 基本概念。解释了 XML 文档的结构和各个主要组成部分。

第 3 章：XML 文档类型定义。介绍了 XML 文档的类型定义及其基本内容、基本方法和具体结构，包括文档类型定义的基本格式、元素声明、属性声明和实体声明。

第 4 章：XML Schema。本章介绍了 XML 的另外一种文档类型声明方式 Schema，包括 Schema 声明元素、属性、数据类型的一般方法以及 XML 文档与 Schema 文档的结合方式。

第 5 章：XML 名称空间。引入名称空间的概念，介绍了其说明、使用、作用范围。

第 6 章：使用 CSS 显示 XML 文档。解释了如何使用 CSS 来解决一个 XML 文档在浏览器中的显示问题。

第 7 章：使用 XSL 对 XML 进行格式转换。介绍了 XSL（Extensible Stylesheet Language，可扩展的样式表语言）的基础知识以及它如何与 XML 进行结合从而将一个 XML 文档按照具体要求进行显示。

第 8 章：数据岛。讲述了如何通过数据岛技术在 HTML 页面中嵌入一段 XML 数据，根据页面的实际需求提供相应的数据。

第 9 章：XML 相关协议和规范。介绍了用于链接资源的 XLink、XInclude，还有用于 XML 编程的程序接口规范 DOM 和 SAX，以及新兴的 AJAX。

第 10 章：XML 数据库。主要介绍了 XML 数据库的主要特点和相关知识。

第 11 章：综合例子。综合运用了前面章节里学习 XML 的相关知识来完成一个基于 XML 的简单客户关系管理系统模型。

三、本书特点

本书内容丰富，结构合理，系统性强，强调通过实践来掌握 XML 的基本概念、相关知识和具体应用。

（1）内容丰富。

本书覆盖了 XML 技术的几个主要方面：XML 文档组成、文档类型说明、XML 显示、名称空间、数据岛、XML 数据库和 XML 程序接口等等。既包括了基本的概念，也包括了深层的理论，还有 XML 的应用，具有丰富的内容。

（2）结构合理，系统性强。

本书先由 XML 的基本概念、基础知识开始，层层深入地讲解了 XML 的相关技术。先介绍了 XML 文档的组成，接着介绍定义 XML 文档的两种方式，然后是显示 XML 的方式，最后是 XML 的应用。读者可以一步一步地、由浅入深地学习 XML 技术。

（3）实用性与理论性并重。

本书无论在技术介绍上，还是所举的例子，或者是编排的实验，都是来源于对实际的抽象，具有与实际紧密结合的特点。而且在每一章的介绍中，既包括了基本的技术，也包括了其深层次的概念，理论与实践并重。

（4）实践性强。

本书强调实践，通过精心编排的实验让读者紧跟课程讲授的知识，并将其应用于实际，从而能够有效地理解和掌握课程知识。

（5）紧贴最新技术。

本书除了对基础知识、基本概念的介绍以外，也选择了一些当前热点技术，如 AJAX 等进行介绍，力求让读者能够紧跟前沿技术。

四、本书适用对象

本书可作为高等院校本、专科计算机及其相关专业的教材，也可作为从事计算机软件开发人员的参考书。

由于作者水平有限，加之编写时间仓促，书中如有疏漏及不足之处在所难免，希望广大读者及专家批评指正。联系方法如下：

电子邮箱：service@cnbook.net　　作者邮箱：great_liu@126.com

网址：www.cnbook.net

本书习题参考答案、电子教案和源程序可从该网站下载，此外，该网站还有一些其他相关书籍介绍，可以方便读者选购参考。

编　者

2007 年 3 月

目　录

第 1 章　XML 概述

XML 是由 W3C 组织提出的一种可扩展标记语言。通过本章的学习，读者可以对 XML 有个初步的认识，了解 XML 的发展过程，明白 XML 的主要概念和应用。

本章教学目标：

（1）了解 XML 产生的背景以及经历的发展过程。

（2）掌握 XML 的基本概念。

（3）了解 XML 的相关应用。

1.1　XML 发展历程

随着网络技术的发展，世界上不同地区的联系变得越来越密切，人们通过网络传输各种信息，信息交换变得越来越频繁，对数据传输的要求也越来越高，如何不断提高数据传输的效率与质量已成为网络技术发展关注的一大问题。在实际过程中，由于这些数据信息交换时会发生数据格式不同的问题，从而给信息交换带来困扰，提供一个便捷有效的数据格式以应用于网络传输成为解决这一问题的重要途径。XML（eXtensible Markup Language，可扩展标记语言）提出的初衷之一就是为了解决这一问题。XML 是一种可以描述任意数据逻辑关系的语言，它提供一个统一的数据说明方式，是一种可以自解释的标记语言，它的出现给数据交换领域带来了一场革命，同时它也促进了下一代网络的发展。

XML 是基于 SGML 上而发展起来的，从某种意义上，XML 是 SGML 的精简版。XML 与 HTML 也有一定的联系，但它们是着眼点不同的两类标记语言，人们在使用 HTML 的过程中发现了一些问题，而这些问题恰恰是 XML 所致力解决的问题。

1.1.1　SGML

SGML（Standard Generalized Markup Language，标准通用标记语言），最初是 IBM 在解决一个项目的过程中提出的一个雏形，称为 GML，1986 年国际标准化组织（ISO）对其进行整理并命名为 SGML 或者称为 ISO8879。

SGML 是一种元语言（Meta Language），具有良好的扩展性，可以用于定义新的语言。一个 SGML 语言文档由三部分组成：语法定义、文档类型定义和文档实例。语法定义部分规定文档必须遵循的语法；文档类型定义部分规定文档的结构、组成文档的各个部分和这些部分的内部结构和关系；文档实例则是 SGML 文档的主要载体。SGML 是非常严谨的一门标记语言，它要求文档具备良好的结构性。在 SGML 的实际使用中，每一个特定的 DTD 都定义了一类文件。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，比如 HTML 语言。这样 SGML 就成为这些派生语言的元语言。

SGML 可移植性强，因为 SGML 在设计之初并没有针对特定的应用程序而专门设计，因此可以跨平台使用 SGML。SGML 具有极强的完整性和稳定性，在 ISO 对其进行标准化以后几乎就没有再修改，并且它适用的应用领域也非常之广，因为制定 SGML 时就考虑到它的适用性问题。由于 SGML 功能的完整性，导致它是一门非常强大而复杂的语言，因此

目前并没有足够多的支持 SGML 的应用软件。

在与 XML 的比较方面，由于 SGML 可扩展性高、功能完整并且应用广泛，从而导致它的复杂度高，较难学习和进行编写，支持 SGML 的处理器价格昂贵，因此人们又在 SGML 的基础上制定了一种新的标记语言——XML，它的规模相对较小，并且易于学习和掌握，它将 SGML 中不经常使用的和不适应于 Web 应用的部分去掉。在 XML 中，用户可以根据需要来选择是否编写相应的 DTD，而在 SGML 中是必须包含 DTD 的。XML 也可以用于编写其他的标记语言，它保留了 SGML 高度的灵活性，是 SGML 的一个子集。

1.1.2 HTML

HTML（HyperText Markup Language，超文本标记语言），1989 年，欧洲物理量子实验室（CERN）的专家蒂姆·伯纳斯·李是 HTML 的设计者，起初他为了便于在实验室内交换数据资料，提出了一种在实验室成员的文件之间建立“链接”的方法，当需要另外一个人的文件时，只要“链接”到对方的电脑上就行，而不必将其文档拷贝到自己的电脑上，采用这种方式就能够将各自的信息通过超文本传输实现网络共享。蒂姆·伯纳斯·李选择了 CERN 使用的一组 SGML 的 DTD 标记标签，在最早的 Web 浏览器和编辑器 NEXUS 中，他使用了这些标签和样式表进行排版，并增加了最重要的功能——链接，这就是 HTML 的前身。1991 年，蒂姆·伯纳斯·李定义了 HTML 语言的第一个规范，接着经历了 HTML1.0、HTML2.0、HTML3.0 和 HTML4.0 多个版本的发展，现在 HTML5.0 处于测试阶段，此外还有根据 HTML 发展起来的 DHTML、VHTML、SHTML 和 XHTML。W3C 将 HTML 规定为在互联网上发布信息的标记语言，它的 DTD 被作为标准而固定了下来，所以 HTML 不属于元语言，它不能用于定义其他标记语言。

所谓超文本，因为它可以加入图片、声音、动画、影视等内容，并且它可以从一个文件跳转到另一个文件，与世界各地主机的文件连接，通过 HTML 可以表现出丰富多彩的设计风格，它的每个标记都有特定的含义，代表一种页面的设置方法，规定数据进行显示的格式。而 XML 并不关心数据如何进行显示，相反的，XML 注重于表示数据间的逻辑关系，而 HTML 在这一点上显得力不从心，此外，由于 HTML 对超链接已支持不足，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，不能表示多种媒体之间的关系等弱点，导致了 HTML 的进一步发展，特别是在多媒体处理方面，XML 更是显示出了它卓越的数据表现能力。

尽管如此，设计 XML 的目标并不是为了取代 HTML，它们有各自的应用背景，HTML 是 SGML 的一个特例，它专门针对 WWW 上的应用，具有相当强的针对性；而 XML 具备很好的灵活性，适合于为具体的应用服务，它可以针对具体的应用构造相应的数据结构，数据根据 XML 规定的逻辑关系进行编排，甚至可以将数据提交给专门的应用程序进行处理，而不仅仅局限于浏览器，从而实现用户需要的操作。

1.1.3 XML

XML 最初是为适应出版界的要求而设计的一种描述文档的语言，它是通过使用 SGML 进行描述而派生出来的一种新的标记语言，后来人们发现在网络数据传输方面使用 XML 能够解决数据格式冲突的问题，因此它得到了迅猛的发展，因此现在它不仅仅是 SGML 定

义出来的描述文档的工具，而且被广泛地应用到网络数据交换的过程中，特别是在电子商务领域XML发挥了极大的作用。XML1.0版本是W3C于1998年2月正式发布为推荐标准，XML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。

XML简化了SGML，但是它与SGML是兼容的，实际上现有的SGML解析器只能是有针对性地实现SGML的部分功能，造成了实际上存在多种不同的SGML解析器，反而造成非标准化，而XML作为一个简化版本，软件制造商在制作XML语法解析器时能够实现其全部的功能，从而避免非标准化问题，并且随着SGML的发展，如果希望XML能够被SGML解析器所理解，则XML也必须与SGML保持兼容。

XML继承了SGML的一些特性，是一种定义严谨的标记语言，它不允许定义不明确的语法结构，不允许像HTML一样支持半结构化的数据，例如在HTML中标记可以不成对出现，而XML要求标记必须成对出现。与SGML类似，XML也是一种元语言，可以使用XML来定义新的标记语言，因此它可以成为描述电子商务数据、多媒体演示数据、数学公式等各种数据应用语言的基础语言，例如NewsML、MathML、CML、FpML、SMIL、BML等都是使用XML定义的标记语言。

XML是以文本形式来描述的一种文件格式，因此具有跨平台的能力，可以在各种不同的平台环境下实现数据交换，也适应于不同平台下的应用。但是由于文本形式下各种文字的编码问题会导致XML文档在显示和处理上出现乱码，因此XML通过指定文字使用的具体编码来告知XML处理器采用何种编码来对文字进行处理。

1.1.4 标记语言

以前，在印刷书稿时，作者会在草稿上进行各种说明，指导如何处理版面的排放，这类说明被称之为标记，久而久之，用来协调一致用于定义整套语法和文法的标记的集合就被称为语言。校对者使用手写的标记语言（Markup Language，ML）与作者进行交流。标记语言，特指用一系列约定好的标记来对电子文档进行标记，来实现对电子文档的语义、结构、格式的定义。这些标记必须能够容易地和内容相区分，易于识别。标记语言必须定义什么样的标记是允许的，什么样的标记是必须的，标记是如何与文档的内容相区分的，以及标记的含义是什么。

“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。当代的标点符号也是某种形式的标记语言，它指导读者如何对一篇文章的文字进行断句。

标记语言的典型代表就是：SGML、HTML、XML。

标记语言可以分为两大类：一般用途类与特殊用途类。一般用途类的标记语言采用开放的处理方法，不是针对具体的应用领域而专门设计的。这种语言在设计时不会定义按照何种方式来处理标记和代码，它仅仅定义文件的结构与内容的意义。特殊用途类的标记语言是专门针对具体的软件、应用领域而特别设计的标记语言，例如HTML，它是专门为互联网的应用而设计的，它的每个标记都有具体的含义，规定了具体执行的动作，并且它也是专门适用于浏览器，它对所包含的数据，并不特别强调其结构性，而是在对如何显示数

据方面下功夫。

1.2 XML 是什么

XML 是 SGML 的子集，它是为了允许普通的 SGML 在 Web 上以超文本的方式处理和传输而提出的。XML 提供一套定义标记的规则，它使用标记对一篇文档进行标识，以便于应用程序对文档进行处理。XML 是一种元标记语言，用户可以根据需要定义标记。

1.2.1 XML 的设计目标与特点

W3C 为 XML 设定了 10 个设计目标，分别是：

（1）XML 可以直接在 Internet 上使用。

（2）XML 对多种应用程序提供支持。

（3）XML 必须与 SGML 保持兼容性。

（4）可以轻松地编写能够处理 XML 文档的应用程序。

（5）XML 中的可选特征应尽量的少，理想状况为零。

（6）XML 文档应该具备可读性好，而且条理清晰。

（7）XML 设计应该快速准备好。

（8）XML 设计应该规范而且简洁。

（9）XML 文档应该易于创建。

（10）XML 标记的简洁性的重要程度最低。

XML 具有许多的优点：第一，XML 是自描述的，它不仅允许定义自己的一套标记，也可以根据其他各种规则来制定标记；第二，XML 允许对文档内容进行检验，例如文档类型定义、XML 模式等都是应用于对文档进行验证；第三，可以使用 XML 开发各种行业的专有标记语言；第四，XML 的通用性，使它成为不同应用之间交换数据的统一格式；第五，XML 是开放性的，它是 W3C 定制的开放标准，可以广泛地适用于不同的应用环境；第六，XML 规定了文档的结构，使得对文档的搜索方式和方法得到发展，提高了文档检索的效率。

XML 有基本的组成部分，一般一个 XML 文档总是由 XML 声明开始的，它声明了文档使用的 XML 的版本信息、编码信息等等。

元素是组成 XML 的基本单位，一个 XML 文档的主体就是通过若干的元素来组成的。元素拥有名称，可以有后裔，它们可能是处理指令、子元素、注释或者字符数据。一个良构的 XML 文档必须至少包含一个元素。一个元素的标识是通过开始标记“<>”与结束标记“</>”来进行的。

属性是对具体元素的一个辅助说明，提供元素一些详细的辅助信息。属性包含在元素的开始标记中，属性必须使用单引号或者双引号进行标注，属性的形式是属性值对，通过属性名称与相应的属性值挂钩。一个元素可以有任意个属性值。

处理指令用来提供信息给处理 XML 文档的应用程序，处理指令由处理指令的名称和处理指令信息两部分组成。

XML 与 HTML 一样提供注释功能，注释可以灵活地出现在元素之前或者元素之后，也可以在元素内出现，注释采用“<!--”和“-->”符号进行标注。

1.2.2 文档类型定义

DTD（Document Type Definition，文档类型定义）最初出现在SGML中，它用于规定文档必须符合的结构：可以在文档中存在的元素、哪些元素可以具有的属性、在元素内部元素的层次结构以及元素在整个文档中出现的顺序。

DTD有三个基本用途：

（1）DTD可以对标记编制相应的文档。

（2）DTD可以加强标记参数内部的一致性。

（3）DTD提供XML语法分析器对文档进行处理的依据。

一份DTD文档包含了以下的内容：

（1）定义元素的组成部分。

（2）定义属性及其类型。

（3）对实体进行声明。

（4）定义元素的种类。

（5）定义元素排列的先后顺序。

XML还允许多个文档可以共享若干个DTD文档，它们为某一具体的应用提供一种一致的标记标准，实现将应用数据的文档化管理。DTD可以确保不同的人员或者应用程序能够相互弄清彼此不同规格的文档。DTD说明了文档的架构，它与文档的数据是相分离的，它的这一特性也是促成其实现共享的一个原因。

1.2.3 文档类型模式

Schema是一种与DTD类似的，一组描述一类给定的XML文档而预先定义好的规则。它规定可以在XML文档中出现的元素，这些元素相互之间的关系以及元素内部可以包含的属性等这类文档结构信息。

Schema提供丰富的数据类型，除了提供非常丰富的一组内置simpleType以外，XML模式还允许使用类似规则表达式的语法派生出新的simpleType，也可以通过派生数据类型来派生出新的用户自定义复杂类型。

2001年5月2日，W3C XML Schema（0，1和2部分）成为了W3C的正式标准。随着XML1.1版本的推出，2005年，Schema也相应的发展到了1.1版本。目前流行的Schema定义还有RELAX和微软的XSD等，但仅有XML Schema是W3C的标准。

XML Schema一确定下来，立刻成为全球公认的首选XML环境下的建模工具，已经基本取代了DTD在XML刚刚成为W3C推荐标准时的地位。由于XML是SGML的一个子集，因此它也继承了SGML世界中用于建模的DTD，但DTD有着不少缺陷：

（1）DTD是基于正则表达式的，描述能力有限。

（2）DTD没有数据类型的支持，在大多数应用环境下能力不足。

（3）DTD的约束定义能力不足，无法对XML实例文档作出更细致的语义限制。

（4）DTD不够结构化，重用的代价相对较高。

（5）DTD并非使用XML作为描述手段，并且DTD的构建和访问并没有标准的编程接口，无法使用标准的编程方式进行DTD维护。

XML Schema 正是针对 DTD 的缺点而设计的，XML Schema 是完全使用 XML 作为描述手段，具有很强的描述能力、扩展能力和处理维护能力的模式定义方法。

1.2.4 名称空间

制定 XML 名称空间标准的初衷是为了解决 XML 文档中名称的冲突问题。

在实际应用中，不同行业、不同领域的人都会编写各种各样的 XML 文档，不同的人根据习惯会编写出各自需要的标记名称，但是如此大量的文档必然会存在出现同名标记的情况，那么如果使用同名标记的两个不同 XML 文档刚好一起使用，则会出现名称冲突，不能区分该标记是指哪一个文档中的具体标记。如果将冲突的标记名称重新命名，则也需要修改使用这些 XML 文档的应用程序。

解决名称冲突的一个较好的方案就是给不同的文档赋以不同的名称空间，应用程序通过名称空间来区分具体的同名标记，搞清楚一个元素所来自的具体文档。XML 名称空间就是对这种方案的具体实现。

XML 名称空间机制包括几个活动的部分：本地名、名称空间 URI、前缀和声明。

名称中简洁的、只需在自身上下文中保持惟一的那一部分称为本地名。使用名称空间后，每个上下文元素及其属性都力求使用最简单的名称，这些本地名结合其所属的名称空间构成惟一的标识元素的标志。

名称空间是采用 URI 语法的字符串，名称空间是完整的元素或属性名称的一部分。比如本地名是 element，名称空间为：http://www.w3.org/XMLSchema，则完整名称是 http://www.w3.org/XMLSchema:element。一般有一个名称空间前缀，它是具体的一个名称空间的缩写形式，与名称空间等价，选择名称空间的 URI 是很重要的，它起到惟一标识一个名称空间的作用。

1.2.5 XML 显示

XML 的显示必须借助一些其他的工具才能实现，例如使用 HTML 来对 XML 的数据进行显示；HTML 的标记都可以在 XML 中使用，但是必须是结构严谨的，例如必须有开始标记和结束标记，标记之间不能出现重叠域等。使用 HTML 可以对 XML 包含的数据设置显示的具体格式。

此外也可使用 CSS（Cascading Style Sheets，级联样式表）来辅助 XML 的显示，CSS 在 HTML 中已经被广泛地使用。在 XML 中，CSS 同样发挥了它强大的样式表作用。在 XML 中的 CSS 和 HTML 中的 CSS 差不多，目前的版本是 CSS 2.0。CSS 文档实际上就是一些数据如何显示的规则集合，它的每个规则都给出下一规则使用的元素名称以及该规则应用哪些具体元素的样式。

XSL（eXtensible Stylesheet Language，可扩展样式语言）是 XML 的首选样式表语言。它是为了 XML 的样式问题而专门设计的。一个 XSL 文档包含了一组规则，用于在 XML 文档中抽取将需要的数据进行转换。转换后的格式是多样的，可以根据需要选择转换的格式。XSL 具有许多的优点：

（1）XSL 具有丰富的数据类型，它不仅支持简单数据类型并且也支持复杂数据类型。

（2）支持名称空间。

（3）XSL具有许多运算符，适合对XML文档编写复杂的XSL样式表。

（4）XSL可以实现对XML数据的筛选，隐蔽不需要显示的数据。

1.2.6 文档对象模型

文档对象模型（Document Object Model，DOM）是W3C为了给程序开发人员操作XML文档而提出的一系列应用程序接口（API）。它定义了一个XML文档的逻辑结构和访问、操作文档的方法。通过文档对象模型，应用程序的开发人员可以新建文档，遍历文档，增加、修改和删除文档中的元素和内容。

W3C力求为DOM提供一种独立于语言和平台的定义，它定义了组成DOM的对象，提供了标准的编程接口，却不提供特定的实现，这些定义和接口可以通用于各种程序语言、操作系统和应用程序。目前很多厂商都提供了DOM的支持，但是不同解析器的实现方法也可能有所差别。

DOM在分析XML文档时首先加载整个文档，构造出DOM树之后才允许开发人员进行操作的，这当然为开发人员在树中任意游历提供了便利，使用也比较简单。但是也带来了加载特别大的文档需要很长的时间，占用大量的内存，消耗很多的资源的问题，因此在使用过程，需要衡量，选择合适的方法进行处理。

1.3 XML应用

在实际使用中，XML发挥着巨大的作用，在很多领域和环境下都采用XML来实现需要的特定功能。

XML的应用最主要体现在四个方面：

（1）不同的应用平台上的数据交换。

在实际应用中，可能发生数据需要在不同的平台上交互，由于不同平台的数据格式也不尽相同，因此需要使用统一的标准格式来实现数据在平台间的流动，这种情况下选择XML作为描述数据的语言是最合适的。由于XML的自解释性、灵活性和可扩展性，使得它具备足够的能力来表达各种类型的数据。在这一类应用中，XML为数据处理提供统一的接口，它通过使用标记对数据进行标注，使得应用程序可以理解数据的逻辑结构和具体含义。

（2）在客户-服务器模式下，作为客户端存储数据的容器，在客户端上可以根据具体需求对数据进行处理，减少重复通信量，服务器只需传递相应的XML文件到客户端上即可。

在客户-服务器模式下，原本服务器担负着复杂的逻辑控制的任务，不同的客户端向服务器发送请求，服务器必须根据具体情况加以响应，比如用户权限、业务执行步骤、用户请求的具体数据等等。这些逻辑控制提高了服务器端的程序编写的复杂度，而且由于需求的多变，使得服务器端的程序也会因此受到影响。如果在客户端设置一些XML文档，则它们可以作为存储数据的容器，许多逻辑判断就可以直接在客户端完成而不必提交到服务器端。服务器只需尽可能准确地对客户端需要的数据进行封装，将数据封装为XML文档，提供给相应的客户端就可以。这样两端都可以专注于一方面的工作，也加速了开发周

期。这里使用 XML 文档正是充分地发挥了 XML 的自解释性和灵活性的特性。采用这种方式的应用也就更具备分布性和通用性。

（3）数据表达的多样性。

一个 XML 文档可以作为数据源提供给各种支持 XML 的应用程序，这些应用程序对 XML 数据的处理也是多样的，在对数据的表现形式上也是多种多样的，比如可以将数据表现为图表、图像、文本、动画等具体的形式。

（4）作为编制新语言的工具。

XML 是一种元标记语言，可以用来编制特定领域的专业标记语言，这些基于 XML 编制的标记语言也可称为 XML 应用程序。XML 在很多专业领域都被使用，例如在化学领域上使用的化学标记语言（Chemical Markup Language，CML）、在数学领域上使用的数学标记语言（Mathematical Markup Language，MathML）、在 Web 站点上的频道定义格式（Channel Definition Format，CDF）、在音频方面的标记语言（Voice Extensible Markup Language，VoiceXML）等等。

以下简单的对几类专业领域的标记语言进行介绍。

（1）Chemical Markup Language。

Chemical Markup Language 是由 Peter Murray-Rust 提出的，它也可能是最早的 XML 应用，CML 原来是要发展成 SGML 应用的，但随着 XML 标准的发展，逐步演化成了 XML。

CML 的目标是为了能够以一种直接的方式组织复杂的化学对象，便于计算机的理解，以达到对这些化学对象的显示和检索。CML 可以用于分子结构和序列、光谱分析、结晶学、出版、化学数据库和其他方面。它的词汇表包括分子、原子、化学键、晶体、分子式、序列、对称、反应和其他化学术语。CML 还使得复杂的分子数据可在 Web 上发送。由于 XML 的平台无关性而避免了跨平台时发生的数据格式冲突问题，而如果使用传统的化学软件和文档格式则会发生数据问题。

（2）Mathematical Markup Language。

数学标记语言是用于表达数学方程的一种 XML 应用，MathML 具有很大的灵活性，它可以处理的数学问题也很广泛，例如从基本的算术到微积分、微分方程等等，甚至还能有更高级的应用。但是目前也还需要进一步发展，因为在一些晦涩的记号面前它也无能为力，在一些纯数学和理论物理的高端方面还具有局限性。目前，MathML 对于工程、教育、科学、商业、经济和统计学的一般要求还是能够满足的。随着 MathML 的进一步扩展，在 Web 上进行数学语言的传播业将成为可能。

（3）Channel Definition Format。

Microsoft 定义了频道定义格式，它是用于定义频道的 XML 应用。Web 站点使用频道向用户传送信息，一改过去那种坐等用户前来浏览并获取信息的状况。这也叫做 Web 广播。CDF 首先是在 Internet Explorer 4.0 中引入的。

CDF 文档是一个 XML 文件，与被广播的站点的 HTML 文件分别存放，但是却链接到此 HTML 文件上。CDF 文档中的频道定义决定了要发送哪个页面。页面可以通过发送通知向预订者加以推送，但也可以发送整个站点，或是由阅读者在方便的时候自己来获取信息。用户可向自己的站点添加 CDF，而不用改变现存的所有内容。只要在页面上添加与 CDF