



全国信息化计算机应用技术资格认证推荐教材



XML 应用案例教程

高寿柏 主编

山东友谊出版社



全国信息化计算机应用技术资格认证推荐教材

HandsOn
翰子昂实训体系第三阶段

XML 应用案例教程

高寿柏 主编

山东友谊出版社



图书在版编目 (CIP) 数据

翰子昂实训体系第三阶段 / 高寿柏主编. — 济南:

山东友谊出版社, 2008.5

ISBN 978-7-80737-322-3

I. 翰子昂… II. 高… III. 软件开发—人才—培养—教材
IV. TP311.52

中国版本图书馆 CIP 数据核字 (2008) 第 060138 号

翰子昂实训体系第三阶段

XML 应用案例教程

高寿柏 主编

主 管: 山东出版集团
集团网址: www.sdpress.com.cn
出版发行: 山东友谊出版社
地 址: 济南市胜利大街 39 号 邮政编码: 250001
电 话: 总编室 (0531) 82098756 82098142
发行部 (0531) 82098035 (传真)
印 刷: 济南希尔康印务有限公司
版 次: 2008 年 5 月第 1 版
印 次: 2008 年 5 月第 1 次印刷
规 格: 889mm×1194mm 16 开本
印 张: 6.25
字 数: 125 千字 (本册)
定 价: 180.00 元 (全三册)

(如印装质量有问题, 请与出版社总编室联系调换)

翰子昂实训体系软件理论教材丛书序

本套丛书是翰子昂实训体系在软件项目实训前的技术准备教材，与项目实训教材丛书一起组成了一个较完善的实训教材体系。丛书采用案例教学的形式，带领学员以解决项目问题为目标去学习知识，掌握技术细节。从而使学员在学习知识的同时，掌握整合技术点，为软件开发寻找解决方案的思路与途径。

软件开发要求从业人员熟悉软件开发的行业规范，具有良好的实践基础，有自主学习意识、独立思考的能力。案例教学法就是把软件开发对知识技术的要求带入理论知识点的学习过程，让学员带着要解决的问题去学习知识，逐步形成自己对问题的理解，建立问题与知识点间的联系，形成有效解决方案，从而提高自己分析和解决问题的能力。

案例教学法是一种推动式的教学方法，能创建良好宽松的学习氛围。传统的教学方法是拉动式的，使学员始终处于被动吸收的学习之中，不能培养学员主动思考及独立解决问题的能力。案例教学引进项目实际要求，学员解决问题的过程就是按照自己的理解去寻找相关知识与技术点的过程。在这一过程中，项目的规范与要求是学员要达到的目标，教师指导及相关资料的辅助是学员学习的推动力。项目完成后，学员掌握了问题所涉及的所有知识，形成了自己的解决方案，培养了自己主动思考的学习和工作习惯。

丛书以专题的形式呈现一系列知识结构，在每个专题中又以案例的形式描述了一个完整的实现过程，案例之后是所用知识点的讲解，最后是这个案例的扩展实现。学员在使用丛书时，应首先按照书中的步骤，逐步实现案例，遇到不了解的知识点时，再去参考案例后面的知识点讲解。通过这样主动探求，逐步深入的过程，来实现自己对知识点从感性认识到理性掌握的转变。案例完成后，学员还可以参考案例后的扩展内容进一步深化学习。由此，学员在完成一个案例时，不仅掌握了项目包含的技术知识，还可以了解与之相关的知识与技术。

软件可能是人类做出来的最复杂的东西了，软件开发也是一个复杂的过程，软件开发的讲授就更是复杂的工程了，笔者很高兴看到这套丛书将这—个复杂的过程变得清晰易懂，简单明了。这套丛书是通向正确方向的一条路径。

在丛书成书的思路整理、高校与企业调研、结构设计与编写成书的过程中，得到了许多高校研究与教学人员的大力支持，尤其是华东师范大学职业与成人教育研究所徐国庆博士的指导与建议，使我们能更加清晰地把握丛书定位与方向；

还有更多的企业中资深软件工程师和高校一线教师，亲自参加了技术知识的组织与书稿的写作。在此，对他们辛勤的劳动致以最真诚的感谢。

尽管我们始终追求尽善尽美，以便向您提供最新及最全面的知识与技术，但由于作者水平有限，疏漏之处仍在所难免，恳切希望得到您对本教材的建议。

翰子昂实训体系软件理论教材丛书编委成员，主编：高寿柏、宁东，副主编：任春梅、耿赛猛。

北京翰子昂教育科技有限公司

网址：<http://www.hands-on.com.cn>

2007 年 12 月

前 言

Internet 的兴起代表了一件事——不管是人还是电脑系统都十分需要沟通。为了实现不同系统和不同平台之间的数据交换，W3C 制定了统一的规范 XML，用于实现不同系统以及不同平台之间的数据交换。本书中主要介绍了 XML 基本语法和主要技术。主要技术包括：文档结构描述技术 DTD 和 Schema、样式表技术 XSL、利用 XML 接口技术来操作 XML 文档的 DOM。通过学习我们可以掌握 XML 的基本语法和常用技术，为今后语言的学习打下基础。

书中内容以案例驱动为主，由实践引导理论，按照由浅入深、前后照应的顺序来安排内容。本书内容有：

专题一 使用 XMLSPY 编写 XML

专题二 文档类型定义

专题三 Schema 和命名空间

专题四 XSL

专题五 文档对象模型 DOM

本书结构合理，思路清晰，语言简洁，案例丰富而实用，既可作为翰子昂实训体系 XML 实训理论教材，也可以作为学习 XML 高级特性的参考资料。

本书编委成员，主编：韩东梅、杨伟韬，副主编：马海滨、武金玲、李弓春。

北京翰子昂教育科技有限公司

2007 年 12 月

目 录

专题一 使用 XMLSPY 编写 XML.....	1
1. 教学目标.....	1
2. 工作任务.....	1
3. 相关实践知识.....	1
4. 相关理论知识.....	3
4.1 XML 的概念和发展.....	3
4.2 XML 的编写.....	5
4.3 XML 语法规则.....	6
4.4 结构完整的文档和有效文档的区别.....	8
4.5 XML 的使用范围.....	8
4.6 XML 的编译、编辑环境和设计工具.....	9
5. 实验.....	9
6. 课后作业.....	10
专题二 文档类型定义.....	11
1. 教学目标.....	11
2. 工作任务.....	11
3. 相关实践知识.....	11
4. 相关理论知识.....	16
4.1 文档类型定义 (DTD)	16
4.2 为什么要用 DTD.....	16
4.3 如何引用一个 DTD 文档.....	16
4.4 元素的定义.....	17
4.5 属性的定义.....	17
4.6 实体的定义.....	18
4.7 符号的定义.....	19
5. 实验.....	19
6. 课后作业.....	22
专题三 Schema 和命名空间.....	25
1. 教学目标.....	25
2. 工作任务.....	25
3. 相关实践知识.....	25
4. 相关理论知识.....	28
4.1 Schema 概念.....	28
4.2 数据类型.....	29
4.3 Schema 元素.....	30
4.4 element 元素.....	31
4.5 group 元素.....	31
4.6 attribute	31
4.7 attributeGroup	31
4.8 simpleType 元素.....	32

4.9 complexType 元素	33
4.10 simpleContent 元素	33
4.11 unique 元素	33
4.12 sequence 元素	34
4.13 choice 元素	34
4.14 在 XML 文档中使用 Schema 约束	34
4.15 命名空间	35
5. 实验	35
6. 课后作业	39
专题四 XSL	43
1. 教学目标	43
2. 工作任务	43
3. 相关实践知识	43
4. 相关理论知识	50
4.1 应用 CSS 和 XSL 显示 XML 文档数据	50
4.2 XSL 的解析过程	51
4.3 基本的 XSL 结构、语法和函数	51
4.4 使用 XSLT 转换 XML 文档为 HTML 文档	55
5. 实验	57
6. 课后作业	57
专题五 文档对象模型 DOM	59
Java 实现	59
1. 教学目标	59
2. 工作任务	59
3. 相关实践知识	59
3.1 使用 Java 语言中提供的 DOM 模型操作 XML 文档	59
3.2 使用 Java 语言编写自己的有效性校验器	64
4. 相关理论知识	67
4.1 DOM 模型	67
4.2 解析 XML 文档, 生成文档对象	70
4.3 Java 中的 DOM 模型的对象、接口、属性、方法	73
4.4 文档验证	76
5. 实验	77
6. 课后作业	77
.NET 实现	78
1. 教学目标	78
2. 工作任务	78
3. 相关实践知识	78
3.1 使用 C#语言中提供的 DOM 模型操作 XML 文档	78
3.2 使用 C#语言编写自己的有效性校验器	81
4. 相关理论知识	84
4.1 DOM 模型	84

4.2 解析 XML 文档，生成文档对象	85
4.3 .NET 语言中 DOM 模型的对象、接口、属性、方法	86
4.4 文档验证	90
5. 实验	90
6. 课后作业	90

专题一 使用 XMLSPY 编写 XML

1. 教学目标

- 1.1 了解 XML 的概念和发展
- 1.2 掌握 XML 的编写
- 1.3 掌握 XML 语法规范
- 1.4 掌握结构完整的文档和有效文档的区别
- 1.5 了解 XML 的使用范围

2. 工作任务

使用 XMLSPY 工具编写 XML 文档

3. 相关实践知识

XML SPY 是所有 XML 编辑器中做得非常好的一个软件，是一个集成了编辑、检验、预览等多项功能的商业性 XML 开发软件，支持 XML 文档所见即所得的编辑方式、支持 WYS-WYG、支持 UNICODE、多字符集，支持格式良好的和有效的 XML 文档，支持 Well-formed 和 Validated 两种类型的 XML 文档，支持 NewsML 等多种标准 XML 文档的所见即所得的编辑，同时提供了强有力的样式表设计。

1. XMLSPY 的图形用户界面由下列三个主要部分组成：

- Project 窗口：在这里将文件组织成工程（Project），并对这些文件进行编辑；
Info 窗口：这里显示当前编辑项的元信息。
- 主窗口（Main Window）：即显示正在编辑中的文档窗口。可用的文档视图数目与正在编辑的文档类型有关。可以根据需要在各种视图间切换。
- 各种输入助手（Entry Helper）窗口：输入助手泛指那些在文档编辑过程中提供帮助的窗口，XMLSPY 提供了多种不同的输入助手。可用的输入助手将根据正在编辑的文档类型和主窗口的文档视图类型的不同而变化。

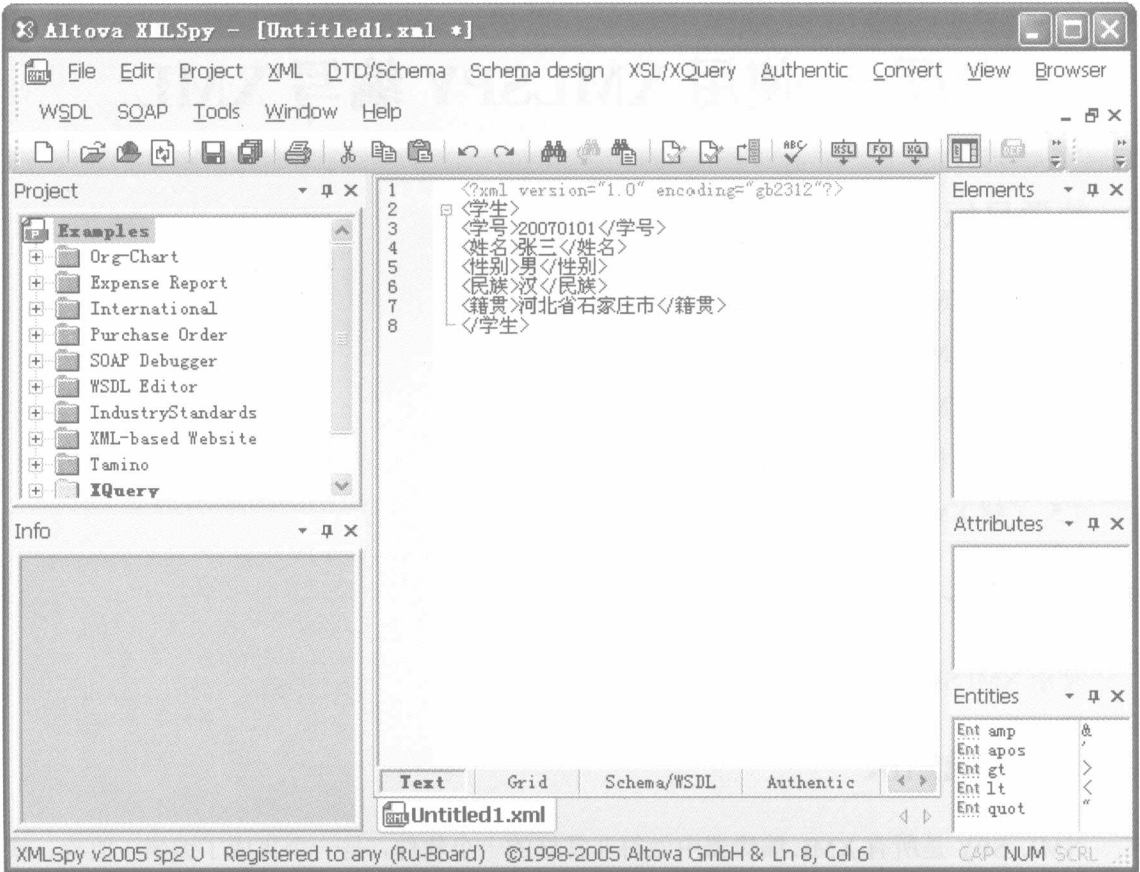


图 1-1 XML 图形用户界面

2. 使用 XMLSPY 工具新建一个 XML 文档。从主菜单的“file”选择“new”，这时将出现一个对话框：

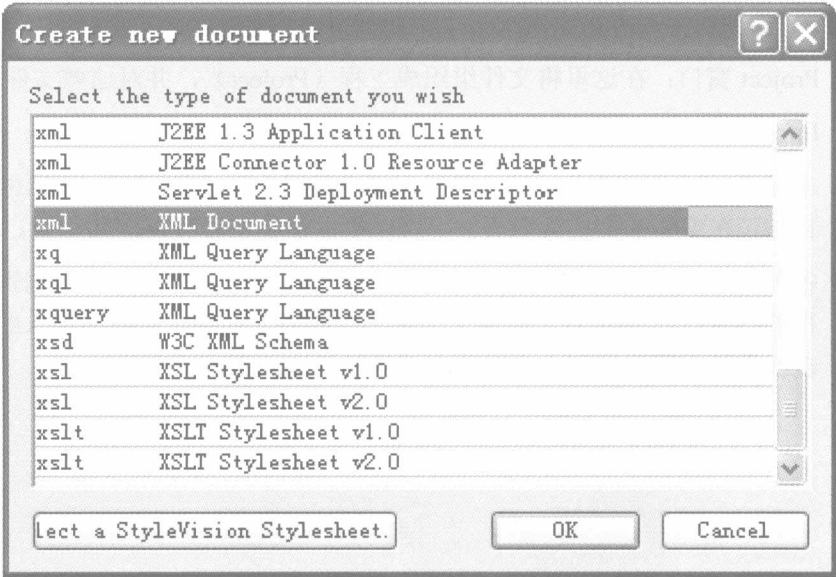


图 1-2 XMLSPY 创建新文档对话框

3. 在对话框中选择“xml XML Document”，然后点击“OK”确认。这时将出现一个提

示，让你选择该 XML 文档是否要基于某个 DTD 或 Schema。

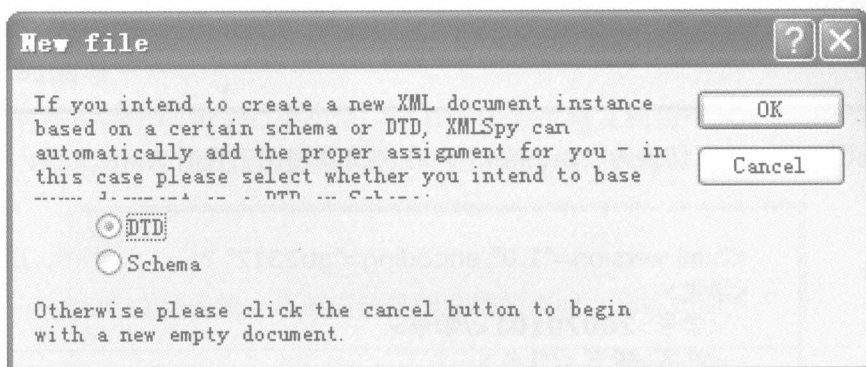


图 1-3 选择 XML 是否基于 DTD 或 Schema

4. 由于没有创建过 DTD 和 Schema，所以这里点击“Cancel”，这时主窗口中将出现新建的 XML 文档。

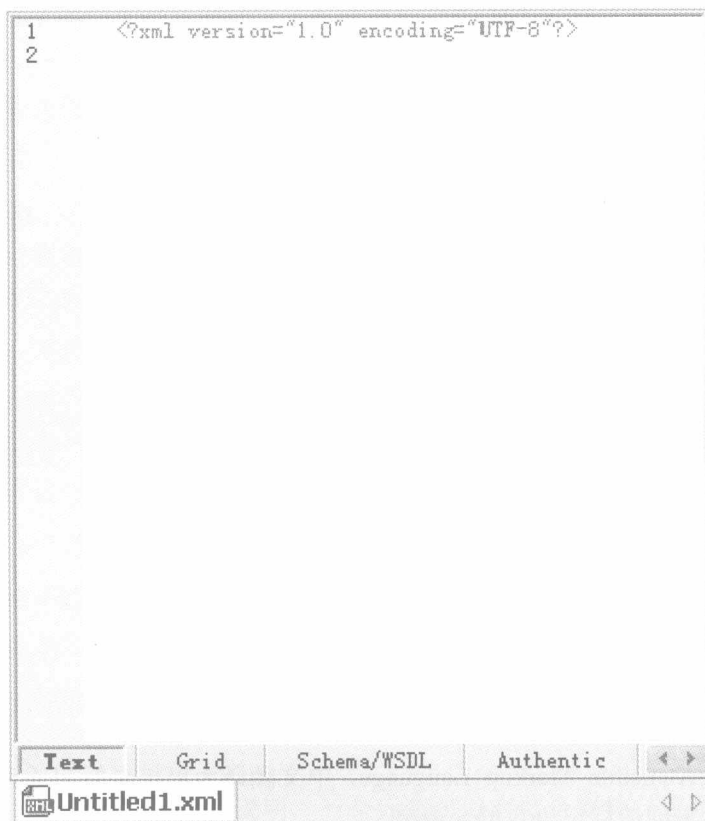


图 1-4 新建的 XML 文档

5. 在“Text”视图中输入如下代码：

```
<?xml version="1.0" encoding="gb2312"?>
<学生>
  <学号>20070101</学号>
  <姓名>张三</姓名>
```

```
<性别>男</性别>
<民族>汉</民族>
<籍贯>河北省石家庄市</籍贯>
</学生>
```

6. 将视图从“Text”切换为“Browser”，在浏览器视图下查看 XML 文档。



图 1-5 Browser 视图

7. 将 XML 文档以“firstxml.xml”名字保存。

4. 相关理论知识

4.1 XML 的概念和发展

XML 全称为 eXtensible Markup Language，中文翻译为可扩展标记语言。它和 HTML 是不一样的，它是用来创造标记语言的元语言。

XML 来源于 SGML（Standard Generalized Markup Language，标准通用标记语言），SGML 是一种元语言，即描述标记语言的语言。SGML 可用来建立任何类型的标记语言，如报告、乐谱，当然还有网络上最常见的 HTML。SGML 的功能非常强大，但设计周期较长，而且也非常复杂，所以其很多特性很少被使用。而由 SGML 所衍生来的 HTML 则在整个网络中发挥了重要作用。但随着网络应用越来越广泛，仅靠 HTML 单一文件类型来处理千变万化的文档和数据已经力不从心，而且 HTML 本身语法并不严密，严重影响网络信息传送和共享。为了解决 HTML 存在的这些问题，W3C 的专家们重新审视了 SGML 标准，并依照 HTML 的

发展经验，制订出一套规则严谨、使用简单的描述资料语言，这就是 XML。XML 于 1996 年成为 W3C 协会的提案，并于 1998 年 2 月 10 日成为正式的标准，即 XML 1.0。

4.2 XML 的编写

XML 的基本语法和 HTML 很相似，但 HTML 只允许文档制作人员使用固定的标记，而 XML 允许文档制作人员通过创建新标记，来更准确地描述数据。下面我们来学习如何编写 XML 文档。

1. XML 声明

任何一个 XML 文档的第 1 行都应该是该文档的声明，其格式如下：

```
<?xml version="1.0" standalone="yes/no" encoding="UTF-8"?>
```

声明的作用是告诉浏览器或者其他处理程序：这个文档是 XML 文档。声明语句中的 version 表示文档遵守的 XML 规范的版本；standalone 表示此文档是否为独立文档（文档如果包含 DTD 文件则不是独立文档），如果是，参数为 yes；encoding 表示文档所用的字符集，默认是 UTF-8，若字符集编码为简体中文就要改为 GB2312。

XML 声明也是一个 XML 处理指令。处理指令以“<?”开始，而以“?”结束。在“<?”后的第一个单词是处理指令名，这里是 xml，代表 XML 声明。

2. 定义根元素

任何一个 XML 文档都有且只有一个根元素，而其他元素则都要包含在这个根元素之内。根元素的起始标记要放在所有其他元素的起始标记之前，而根元素的结束标记要放在所有其他元素的结束标记之后。

3. 定义 XML 元素

XML 元素由三个部分组成：起始标记、内容、结束标记。起始标记以“<”开始，以“>”结束；而结束标记以“</”开始，以“>”结束。XML 中不存在预先定义的元素标记，所以元素标记的名称可自行定义，但需要遵循以下规则：

- 元素名称中可以包含字母、数字以及其他字母；
- 元素名称不能以数字或下划线开头；
- 元素名称中不能包含空格和冒号；
- 元素名称能够描述元素内容。

4. 属性的定义

属性是对元素进一步的描述和说明，一个元素可以有多个属性，每个属性按照由等号分开的名称与值的形式写在元素的开始标记中，属性之间用空格分开。属性值用单引号或双引号括起来。

如：

```
<学生 学号="20070101" 性别="男">张三</学生>
```

属性虽然可以扩充元素的信息，但建议尽量不使用属性，而将属性改成子元素。可将上面的代码改成：

```
<学生>
  张三
  <学号>20070101</学号>
  <性别>男</性别>
</学生>
```

原因是属性不易扩充和被程序操作。

5. 注释

注释是为了便于阅读和理解在 XML 文档添加的附加信息，将不会被程序解释。在 XML 中注释以 “<!--” 开始，以 “-->” 结束。

如：

```
<!-- 注释 -->
```

注释不是元素，它可以放在根元素之前或者之后，但是注释不能位于 XML 声明之前，也不能放在标签中，而且也不能嵌套。

6. CDATA 节

CDATA 全称为 character data，译为字符数据。CDATA 段是用来包含文本的方法，它内部的所有内容都会被 XML 解析器忽略，所以任何符号都不会被认为是标记符。

如：

```
<![CDATA [
<学生>
  张三
  <学号>20070101</学号>
  <性别>男</性别>
</学生>
]]>
```

其中的<学生>、<学号>、<性别>都不再被 XML 解析器解释为标记。需要注意的是，CDATA 段中不能再包含 CDATA 段，也就是说不能嵌套。

4.3 XML 语法规范

XML 的语法虽然和 HTML 很相似，但 XML 拥有严格的规则，XML 解析器不允许 XML 文档有任何错误存在，只有完全符合规则的文档才能被正确显示。XML 文档必须遵守的规则有以下几点：

1. 必须有 XML 声明。

如果没有声明该文档就不能被称为是 XML 文档。

2. 是否有 DTD 文件

如果文档是一个“有效的 XML 文档”，那么文档一定要有相应的 DTD 文件，并且严格遵守 DTD 文件制定的规范。DTD 文件的声明语句紧跟在 XML 声明语句后面，格式如下：

```
<!DOCTYPE 根元素 SYSTEM/PUBLIC "dtd-name">
```

其中：“!DOCTYPE”是指你要定义一个 DOCTYPE；“SYSTEM/PUBLIC”这两个参数只用其一，SYSTEM 是指文档使用的私有 DTD 文件的网址，而 PUBLIC 则指文档调用一个公用的 DTD 文件的网址；“dtd-name”就是 DTD 文件的网址和名称，所有 DTD 文件的后缀名为“.dtd”。

3. 所有 XML 标记都区分大小写

在 XML 文档中，大小写是有区别的。<P>和<p>是不同的标记。注意在写元素时，前后标记大小写要保持一样。

如：

```
<City>Beijing</City>
```

如果写成

```
<City>Beijing</city>
```

是错误的。

4. 所有标记的属性值必须用引号括起来

在 HTML 代码里，属性值可以加引号，也可以不加。但在 XML 中，属性值必须加引号（单引号和双引号都可以），否则是错误的。

如：

```
<学生 学号=20070101 性别=男>张三</学生>
```

是错误的写法。

5. 所有的标记必须有相应的结束标记

在 HTML 中，标记可能不是成对出现的。但在 XML 中，所有标记必须成对出现，而且起始标记和结束标记必须完全对应，包括字母的大小写。

6. 所有的空标记必须被关闭

空标记指的是标记对之间不包含任何内容的标记，格式如下：

```
<banner/>
```

7. 所有 XML 标记必须合理嵌套

XML 中元素嵌套指的是元素必须完全包含在另一个元素中。

例如：

```
<b><i>test</i></b>
```

虽然以下代码在浏览器中能够正确显示，但在 XML 语法中却是错误的。

```
<b><i>test</b></i>
```

4.4 结构完整的文档和有效文档的区别

一个遵守 XML 语法规则，并遵守 XML 规范的文档称之为结构完整的文档。如果所有的标记都严格遵守 XML 规范，那么你的 XML 文档就不一定需要 DTD 文件来定义它。

结构完整的文档必须以一个声明开始，例如：

```
<?xml version="1.0" standalone="yes" encoding="UTF-8"?>
```

其中 version 是必须的，说明文档遵守的 XML 版本，目前是 1.0；其次说明文档是“独立的”，它不需要 DTD 文件来验证其中的标记是否有效；第三，要说明文档所使用的语言编码，默认的是 UTF-8，如果使用中文，需要设置为 GB2312。

结构完整的 XML 文档有且只有一个根元素，即在声明的后面建立的第一个元素，其他元素都要包含在这个根元素中。

结构完整的 XML 文档的内容书写时必须遵守 XML 语法。

一个遵守 XML 语法规则，并遵守相应 DTD 文件规范的 XML 文档称为有效文档。结构完整的文档和有效文档最大的区别在于一个完全遵守 XML 规范，一个则有自己的 DTD。

有效文档也必须以一个声明开始，例如：

```
<?xml version="1.0" standalone="no" encoding="UTF-8"?>
```

这里 standalone 属性的值要设置为“no”，由于必须和相应的 DTD 一起使用，所以要在声明之后定义 DTD 文件。例如：

```
<!DOCTYPE filelist SYSTEM "filelist.dtd">
```

4.5 XML 的使用范围

XML 是一种可以用来创建标记语言的元语言，它可以用来标记任何一种所能想到的、看到的事物。数学公式、化学分子结构、音乐符号这些行业信息都在 XML 中得以结构化地表示，跨平台的信息交换也可以制定基于 XML 的通信协议。XML 不仅已广泛用于与计算机和网络有关的各个方面，还可用于数学、机械、生物、化学、物理、音乐制作等各个领域。以下是 XML 近年来最主要的应用体现。

1. 数据交换

用 XML 在应用程序和公司之间进行数据交换已不是什么秘密了，毫无疑问应被列为第一位。那么为什么 XML 在这个领域里的地位这么重要呢？原因就是 XML 使用元素和属性来描述数据。在数据传送过程中，XML 始终保留了诸如父/子关系这样的数据结构。几个应用程序可以共享和解析同一个 XML 文件，不必使用传统的字符串解析或拆解过程。相反，普通文件不对每个数据段做描述(除了在头文件中)，也不保留数据关系结构。使用 XML 做数据交换可以使应用程序更具有弹性，因为可以用位置(与普通文件一样)或用元素名(从数据库)来存取 XML 数据。

2. Web 服务

Web 服务是最令人激动的革命之一，它让使用不同系统和不同编程语言的人们能够相互交流和分享数据。其基础在于 Web 服务器用 XML 在系统之间交换数据。交换数据通常用