

XML 原理与应用

Summer X

RUC

CH1 XML 引论



目录

- XML 的起源
- XML 的设计目标与特点
- XML 技术体系
- XML 的应用与发展
- XML 相关工具

1.1 XML 的起源

- 1 标记起源
- 2 过程标记：RTF
- 3 通用编码：TEX
- 4 SGML
- 5 XML

标记起源

- 标记语言 (Markup Language) 起源于传统印刷
- 计算机中的电子标记 (WPS、 OpenOffice.....)
- 定义：

标记语言就是一种用来给文本添加“格式标注”以指明文档中文本编排格式的语言, 一般由定义文档格式的一些规定代码和控制标记组成。

过程标记

以微软开发的富文本格式 RTF: Rich Text Format 为典型代表：

例子

用 Word 输入如下文字，并保存为 rtf 类型：

XML and RTF!

用记事本打开保存的文件，查看其文本内容

以`TeX`为典型代表，例如以下代码片段：

TeX 示例

```
1 \noindent Tian Xia\par
2 \noindent Information Resource Management\par
3 \noindent Renmin University of China\par
4 \smallskip
5 This is a book about XML. It you have some problems, you can concat us
   directly.\par
6 \bye
7
```

SGML(Standard Generalized Markup Language), 即标准通用标记语言, 是一种定义电子文档结构和描述其内容的国际标准语言, 早在 Web 发明之前 SGML 就已存在。由 IBM 的 Goldfarb、Mosher 和 Lorie 创造。

是 HTML、DocBook、XML 等新标记语言的基础

HTML I

1989 年由欧洲量子实验室的研究人员 Tim Berners Lee 在 SGML 的基础上开发的一个简化子集。

HTML 示例

```
1 <html>
2   <head>
3     <title>HTML网页测试</title>
4   </head>
5   <body>
6     <h1>简单的HTML</h1>
7     <p>我的<font color="red">测试网页！ </font></p>
8   </body>
9 </html>
10
```

HTML II

HTML 本质上是通过标记的方式, 以纯文本形式对网页进行描述, 而对标记的解释则由浏览器执行, 如现在流行的 IE (Internet Explorer) 浏览器、FireFox 浏览器、Opera 浏览器等。浏览器读取网页源代码, 即 HTML 标记文本, 并通过渲染呈现给用户, 就形成了用户最终看到的网页。

HTML 缺点

- 标记固定
- 标记侧重于如何显示信息, 缺乏对数据内容含义的表达能力
- 缺乏严格的结构要求

W3C 设计，删除了 SGML 中所有不必要的组件，保留了 SGML 的基本原理：标记用于描述文档结构；模型必须与文档相关联。并注重简单性原则。

XML 示例文档内容：

```
1 <?xml version="1.0" encoding="gb2312"?>
2 <?xml-stylesheet type="text/xsl" href="1-5.xsl"?>
3 <books>
4   <book isbn="7-302-02368-9">
5     <title>数据结构</title>
6     <author>严蔚敏，吴伟民</author>
7     <publisher>清华大学出版社</publisher>
8     <price>22.0</price>
9   </book>
10 </books>
```

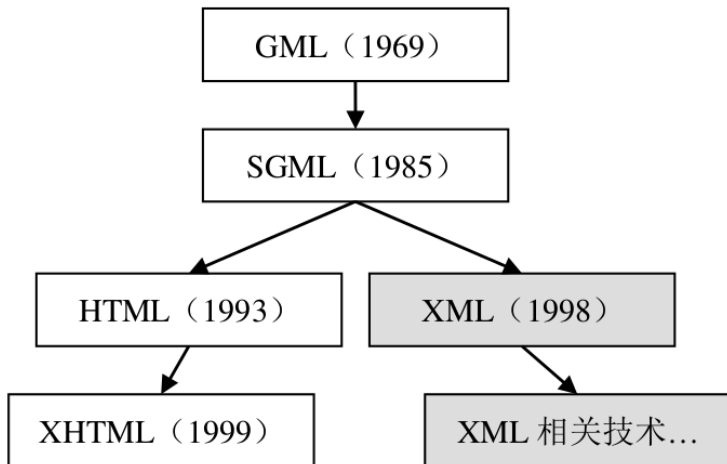
XSLT 示例文档内容：

```
1 <?xml version="1.0" encoding="gb2312"?>
2 <xsl:stylesheet version="1.0"
   xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
3 <xsl:template match="/">
4   <html>
5     <head><title>这是图书的呈现样式结果</title></head>
6     <body>
7       <table border="1">
8         <tr>
9           <th>书名 </th>
10          <th>作者 </th>
11          <th>出版社 </th>
12          <th>价格 </th>
13        </tr>
14        <tr>
```

```
15      <td><xsl:value-of select="books/book/title"/></td>
16      <td><xsl:value-of select="books/book/author"/></td>
17      <td><xsl:value-of select="books/book/publisher"/></td>
18      <td><xsl:value-of select="books/book/price"/></td>
19  </tr>
20  </table>
21  </body>
22  </html>
23  </xsl:template>
24  </xsl:stylesheet>
```

用浏览器打开 XML 文档，查看效果

SGML、HTML 与 XML 的关系



XML 的设计目标

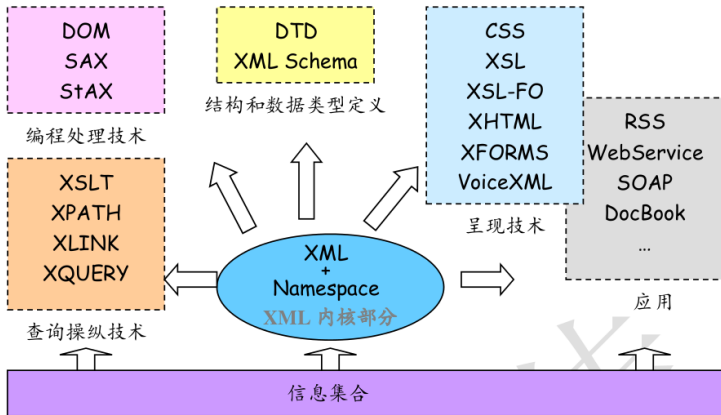
- 可以直接应用于因特网
- 支持各类不同的应用程序
- 与 SGML 兼容
- 处理 XML 文件的程序容易编写
- 选择性功能的数量尽可能少
- 清晰明了, 可读性强
- 设计应该合乎格式并且简洁
- 容易创建
- 标记必须保证其可读性, 不能因为过于简化而导致含义模糊

XML 的主要特点

- 具有良好的格式
- 具有验证机制
- 增强了 Web 应用的灵活性
- 具有丰富的显示样式
- 是电子数据交换 EDI 的通用格式
- 支持复杂的数据关系和快捷的数据处理
- 具有面向对象的特性
- 是一种开放的标准
- 技术体系性强

思考 XML 的不足之处

XML 技术体系



XML 的应用与发展

- 行业标记语言设计领域
- 电子文件的长期保存领域
- 电子数据交换领域
- Web 应用领域

XML 相关工具

- 编辑工具：文本编辑工具、oXygen XML Editor、XML Spy ...
- 浏览工具：浏览器
- 验证工具：浏览器、专用 XML 工具和软件包
- 解析器：Apache Xerces

oXygen XML Editor



推荐使用，比 XML Spy 对标准的支持更好

END

