# XML概述

**XML**—extensible Markup Language(可扩展标记语言)

XML最基本的三个概念

1)XML语言---描述事物本身(可扩展)

2)XSL语言---展现事物表现形式

3)DTD(Schema)---定义XML语言的语法

XML—extensible Markup Language(可扩展标记语言)

**XML的优点**

1)异质信息互通

2)机器语言(netCravler、自动搜索、更精确思索)

3)简单

4)可以自定义语言(可扩展)

**XML Vs  HTML**

|  |  |  |
| --- | --- | --- |
| 比较内容 | HTML | XML |
| 可扩展性 | 不具有可扩展性 | 是元标记语言，可以定义新的标记语言，标记  由用户定义 |
| 侧重点 | 侧重于信息的表现形式，  数据显示为什么格式为HTML所关注 | 侧重于结构化的描述信息，数据是什么为XML所关注 |
| 语法 | 不严格(嵌套、配对) | 严格要求嵌套、配对严格按照DTD的要求 |
| 可读、可维护 | 难于阅读，难于维护 | 结构清晰，便于阅读与进行维护 |
| 数据本身显示 | 数据与显示合在一起 | 数据与显示分离 |
| 可重用 | 差 | 可重用性很高 |

**XML语法**

验证一个.XML文档是否正确

1)格式正确

   --编程验证

   --XMLParser(比如用浏览器验证)

2)与DTD(Schema)文档相符

　--编程

   --XML Parser

基本语法

1)空格

  --<name></name>

  --<name> </name>(有空格)

  --<poem xml:space="preserve">

                 xml

           学习xml

      ysw在学习xml

      </poem>

   空格包括:

   --Space

   --Tab

   --CR(Mac) LF(Unix) CR/LF(Microsoft)

2)PI(Process Instruction)

   格式：<?...........?>

   传递指令给下游程序

3)NameSpace 命名空间

   有下面两种形式：

　<policeman>                                              <criminal>

　　  <name>007</name>                                   <name>008</name>

  </policeman>                                              <criminal>

**XML文档结构**

--通常从XML声明开始;

 --通过XML元素来组织数据

 --为了使数据更加清晰并且具有自身的格式可以引入CDATA(保留标签里面特殊的格式:如描述信息中有尖括号<>,容易引发错误)区数据块

 --可以在文档中加入注释

 --如果需要给XML处理程序提供一些信息，可以包含处理指示

**XML特殊字符**

|  |  |
| --- | --- |
| > | &gt |
| < | &lt |
| & | &amp |
| " | &quot |
| ' | &apos |

**XML编码**

UTF-8、UTF-16

**XSL概念**

XSL:extensible stylesheet Language,包含XSLT和FO

FO：Formatting Object(已经淡出视野了)

XSLT：T--Transformation(就是现在XSL)

XSL的用途

1)XML -->HTML,因为HTML是XML的一种，所以也可以理解为XML -->XML

2)XML-->XML,用于企业间数据传输

3)XSL-->XSL,因为XSL也是一种XML，所以也可以理解为XML -->XML