

第2章 XML概述

1. 什么是XML?
2. 为什么需要XML?
3. XML的发展历史
4. XML的主要应用
5. XML的主要开发工具

1 什么是XML? (1)

XML，既可扩展标记语言，是一种你可以按一定的规则来创建和组织自己的标记和数据的标记语言。

XML = eXtensible Markup Language

具体地讲：

- ✱ XML是可扩展的标记语言；
- ✱ XML在风格上类似HTML语言；
- ✱ XML是用来对数据进行描述的语言；
- ✱ XML中的标记需要你自己去定义，而不是预先定义的；
- ✱ XML中用 DTD 或 XML Schema 来对数据进行描述；
- ✱ 带 DTD 或 XML Schema 的 XML 具有自描述性。

1 什么是XML? (2)

一个简单的 XML 的例子:

例1

```
<address>
  <name>
    <title>Mrs.</title>
    <first-name> Mary </first-name>
    <last-name> McGoon </last-name>
  </name>
  <street> 1401 Main Street </street>
  <city state="NC" >Anytown</city>
  <postal-code> 34829 </postal-code>
</address>
```

1 什么是XML? (3)

有三个通用术语用来描述 XML 文档的组成部分：标记、元素和属性。
前面的例子说明了这些术语：

- ✱ 标记是左尖括号（<）和右尖括号（>）之间的文本。有开始标记（例如 <name>）和结束标记（例如 </name>）
- ✱ 元素是开始标记、结束标记以及位于二者之间的所有内容。在上面的样本中，<name> 元素包含三个子元素：<title>、<first-name> 和 <last-name>。
- ✱ 属性是一个元素的开始标记中的名称-值对。在该示例中，state 是 <city> 元素的属性。

<联系人列表>

<联系人>

<姓名>张三</姓名>

<ID>001</ID>

<公司>A公司</公司>

<EMAIL>zhang@aaa.com</EMAIL>

<电话>(010)62345678</电话>

<地址>

<街道>五街1234号</街道>

<城市>北京市</城市>

<省份>北京</省份>

<ZIP>100001</ZIP>

</地址>

</联系人>

XML

<联系人>

<姓名>李四</姓名>

<ID>002</ID>

<公司>B公司</公司>

<EMAIL>li@bbb.org</EMAIL>

<电话>(021)87654321</电话>

<地址>

<街道>南京路9876号</街道>

<城市>上海</城市>

<省份>上海</省份>

<ZIP>200002</ZIP>

</地址>

</联系人>

</联系人列表>

2023/9/15

张三

用户ID: 001

公司: A公司

EMAIL:

zhang@aaa.com

电话: (010)62345678

地址: 五街1234号

城市: 北京市

省份: 北京

ZIP: 100001

HTML

李四

ID: 002

公司: B公司

EMAIL: li@bbb.org

电话: (021)87654321

地址: 南京路9876号

城市: 上海市

省份: 上海

ZIP: 200002

XML 语言

1 什么是XML? (4)

仅仅将数据标记还是不够的，还需包括标记的语法、结构、每个标记的含义。在 XML 中，标记的语法是通过 DTD 或 XML Schema 来描述的。各标记的含义是由与之相关的 XSL 来向应用程序提供如何处理或显示的指示说明。(XSL=eXtensible Stylesheet Language)

XML Terminology

DTD (Document Type Definition)

- ▣ Describes structure of XML files (old)

XSD (XML Schema)

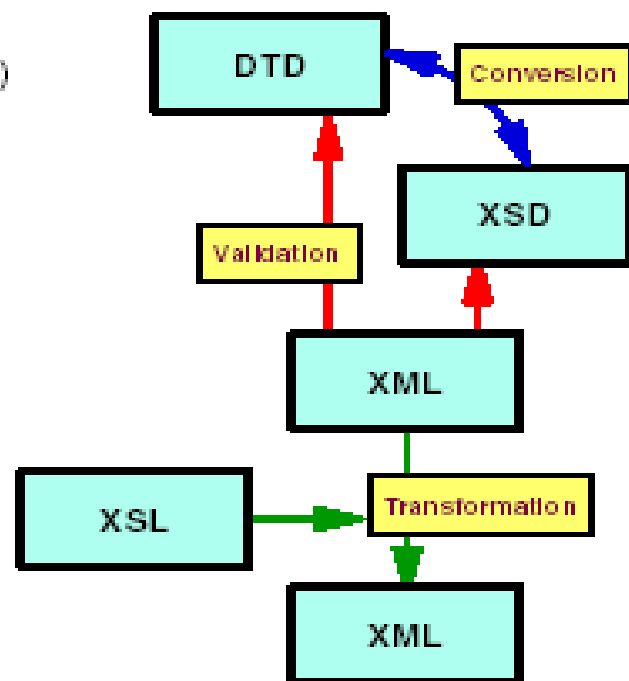
- ▣ Describes structure of XML files and data types (new)

XML file

- ▣ Well structured
 - ▶ Every tag has an end tag
 - ▶ Properly nested
- ▣ May point to DTD or XSD
 - ▶ Validation

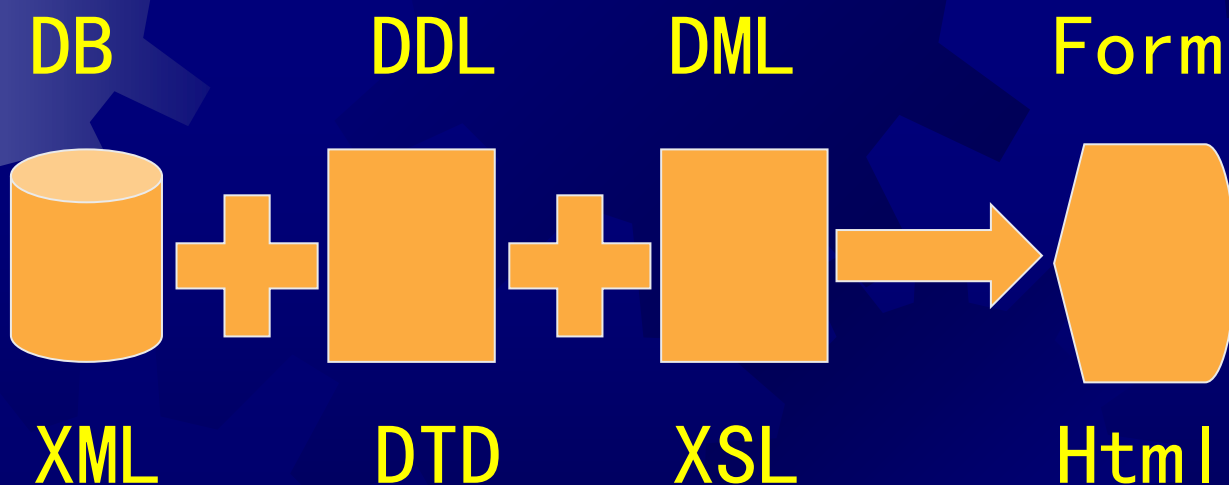
XSL (XML Stylesheet Language)

- ▣ Transformation of XML files



1 什么是XML? (5)

XML可以看作是一个文本数据库，可以借助数据库的概念来理解XML、DTD、XSL三者的关系



1 什么是XML? (6)

下面以一个XML实例来说明XML、DTD、XSL三者的关系。

例2 客户关系列表的DTD文件: [fc1ml.dtd](#)

例3 客户关系列表的信息的XML文件:
[client.xml](#)

例4 客户关系列表到HTML的转换XSLT文件:
[mystyle.xsl](#)

DTD + XML + XSL

fclml.dtd:

```
<?xml version="1.0" encoding="GB2312"?>

<!ELEMENT 联系人列表 (联系人)*>
<!ELEMENT 联系人 (姓名,ID,公司,EMAIL,电话,地址)>
<!ELEMENT 地址 (街道,城市,省份)>
<!ELEMENT 姓名 (#PCDATA)>
<!ELEMENT ID (#PCDATA)>
<!ELEMENT 公司 (#PCDATA)>
<!ELEMENT EMAIL (#PCDATA)>
<!ELEMENT 电话 (#PCDATA)>
<!ELEMENT 街道 (#PCDATA)>
<!ELEMENT 城市 (#PCDATA)>
<!ELEMENT 省份 (#PCDATA)>
```

1

DTD

client.xml

```
<?xml version = "1.0" encoding="GB2312"
standalone = "no"?>
<!DOCTYPE 联系人列表
SYSTEM "fclml.dtd">
<?xml-stylesheet type="text/xsl"
href="mystyle.xsl"?>
```

<联系人列表>

<联系人>

<姓名>张三</姓名>

<ID>001</ID>

<公司>A公司</公司>

<EMAIL>zhang@aaa.com</EMAIL>

<电话>(010)62345678</电话>

<地址>

<街道>五街1234号</街道>

<城市>北京市</城市>

<省份>北京</省份>

</地址>

</联系人>

<联系人>

<姓名>李四</姓名>

<ID>002</ID>

<公司>B公司</公司>

<EMAIL>li@bbb.org</EMAIL>

<电话>(021)87654321</电话>

<地址>

<街道>南京路9876号</街道>

<城市>上海</城市>

<省份>上海</省份>

</地址>

</联系人>

</联系人列表>

2

XML

mystyle.xml

```
<?xml version="1.0" encoding="GB2312"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/TR/WD-xsl"
  xmlns="http://www.w3.org/TR/REC-html40"
  result-ns="">
<xsl:template><xsl:apply-templates/></xsl:template>

<xsl:template match = ""/>
  <HTML>
  <HEAD>
    <TITLE>F公司的客户联系信息</TITLE>
  </HEAD>
  <BODY>
    <xsl:apply-templates select="联系人列表"/>
  </BODY>
</HTML>
</xsl:template>

<xsl:template match = "联系人列表">
  <xsl:for-each select="联系人">
    <UL>
      <LI><xsl:value-of select="姓名"/></LI>
    </UL>
    <LI>用户ID:<xsl:value-of select="ID"/></LI>
    <LI>公司: <xsl:value-of select="公司"/></LI>
    <LI>EMAIL: <xsl:value-of select="EMAIL"/></LI>
    <LI>电话: <xsl:value-of select="电话"/></LI>
    <LI>街道: <xsl:value-of select="地址/街道"/></LI>
    <LI>城市: <xsl:value-of select="地址/城市"/></LI>
    <LI>省份: <xsl:value-of select="地址/省份"/></LI>
  </UL>
</xsl:for-each>
</xsl:template>
</xsl:stylesheet>
```



XSL

XML语言

•张三

- o用户ID:001
- o公司: A公司
- oEMAIL: zhang@aaa. com
- o电话: (010) 62345678
- o街道: 五街1234号
- o城市: 北京市
- o省份: 北京

•李四

- o用户ID:002
- o公司: B公司
- oEMAIL: li@bbb. org
- o电话: (021) 87654321
- o街道: 南京路9876号
- o城市: 上海
- o省份: 上海

显示

2 为什么需要XML? (1)

HTML 始终是最成功的标记语言，你几乎可以在任何设备上查看最简单的 HTML 标记。但是，HTML有许多固有的弱点：

- ✱ HTML是专门为描述主页的表现形式而设计的，它疏于对信息语义及其内部结构的描述。
- ✱ HTML对表现形式的描述能力实际上也还是非常不够的，无法描述矢量图形、科技符号和其他的一些特殊显示效果。
- ✱ HTML的标记集日益臃肿，而其松散的语法要求使得文档结构混乱而缺乏条理性。

2 为什么需要XML? (2)

相对于HTML的缺陷, XML具有如下的一些优点:

- XML具有良好的可扩展性: XML提供了一些基本语法, 但没有定义确切的标记。标记集可以根据不同的目的进行扩充, 制定符合各种需要的新标记。
- XML的内容与形式分离: XML是数据描述语言, 和数据的表示无关。数据的合法性和正确性可以用 DTD 或 XML Schema 来描述。它的表现形式则从数据信息中抽取出来, 用 XSL 来描述。
- 遵循严格的语法要求: 不但要求标记配对、嵌套不交叉, 而且还要求严格遵循 DTD 或 XML Schema的规定。清晰可读、自我描述性强。
- 便于不同的系统之间信息的传输: 各种不同的系统之间可以采用XML作为交流媒介, 使得XML成为一种非常理想的网际语言。
- 具有良好的保值性: XML不但能够长期作为一个通用的标准, 而且很容易向其它格式的文件转化。

2 为什么需要XML? (3)

请观察下面这个HTML文档:

例5

```
<p><b>Mrs. Mary McGoon</b>  
<br>  
1401 Main Street  
<br>  
Anytown, NC 34829</p>
```

考虑从该地址抽取邮政编码的算法:

*如果找到有两个
 标记的段落, 那么邮政编码就是
第二个
 标记下面第一个逗号之后的第二个词。*

尽管该算法对于这个示例起作用, 但对于全世界许多完全有效的地址, 该算法根本不起作用。即使你可以编写算法来找出任何用 HTML 编写的地址的邮政编码, 但许多具有两个换行标记的段落根本不包含地址。

2 为什么需要XML？（4）

现在让我们再来看一看例1。

```
<address>
  <name>
    <title>Mrs.</title>
    <first-name> Mary </first-name>
    <last-name> McGoon </last-name>
  </name>
  <street> 1401 Main Street </street>
  <city state="NC" >Anytown</city>
  <postal-code> 34829 </postal-code>
</address>
```

使用 XML，你可以给文档中的标记赋予某种含意。更重要的是，机器也容易处理这样的信息。你只需通过找到 `<postal-code>` 和 `</postal-code>` 标记之间的内容，就可以从该文档抽取邮政编码。

2 为什么需要XML？（5）

XML与HTML的比较表

比较内容	HTML	XML
可扩展性	不具有扩展性	是元标记语言，可用于定义新的标记语言
侧重点	侧重于如何表现信息	侧重于如何结构化地描述信息
语法要求	不要求标记的嵌套、配对等，不要求标记之间具有一定的顺序	严格要求嵌套、配对，和遵循DTD的树形结构
可读性及可维护性	难于阅读、维护	结构清晰，便于阅读、维护
数据和显示的关系	内容描述与显示方式整合为一体	内容描述与显示方式相分离
保值性	不具有保值性	具有保值性
编辑及浏览工具	已有大量的编辑、浏览工具	编辑、浏览工具尚不成熟

3 XML的发展历史（1）

XML有两个先驱——SGML和HTML，这两个语言都是非常成功的标记语言，但是他们都在某些方面存在着一些缺陷。XML正是为了解决它们的不足而诞生的。

让我们首先从SGML说起：

SGML的全称是标准通用化标记语言，它从上世纪80年代初开始使用。

和 XML一样，SGML也可用于创建成千上万的标记语言，它为语法标记提供了异常强大的工具，同时具有极好的扩展性，因此在分类和索引数据中非常有用。目前，SGML多用于科技文献和政府办公文件中。

但是，SGML非常之复杂，其复杂程度对于网络上的日常应用简直不可思议。不仅如此，SGML非常昂贵。还有最关键的一点，几个主要的浏览器厂商都明确拒绝支持SGML，这无疑是SGML在网上传播遇到的最大障碍。

3 XML的发展历史（2）

我们再来说说HTML：

HTML免费、简单，而且它获得了广泛的支持。HTML最初于1990年由CERN设计，它是一个非常简单的SGML语言，可以方便普通人的使用。而正如设计之初所构想的那样，HTML现在在世界范围内得到了广泛的应用。

不幸的是，HTML有许多致命的弱点：

- ✱ HTML是专门为描述主页的表现形式而设计的，它疏于对信息语义及其内部结构的描述。
- ✱ HTML对表现形式的描述能力实际上也还是非常不够的，无法描述矢量图形、科技符号和其他的一些特殊显示效果。
- ✱ HTML的标记集日益臃肿，而其松散的语法要求使得文档结构混乱而缺乏条理性。

3 XML的发展历史（3）

XML的诞生：

正因为如此，1996年人们开始致力于描述一个标记语言，它既具有SGML的强大功能和可扩展性，同时又具有HTML的简单性。国际互联网联盟 W3C (World Wide Web Consortium) 决定专门成立一个SGML专家小组来从事此项工作。

SGML中所有非核心的、未被使用的和含义模糊的部分都被删除，剩下的就成为短小精干的标记工具—XML。对于XML的描述只有26页，而当初的SGML的描述却长达500页之多。尽管篇幅只是SGML的1/20，但SGML中所有的精华都被保留了下来。

这以后，XML不断发展演化。最后，W3C于1998年2月批准了XML的1.0版本，一个崭新而大有前途的语言诞生了。

4 XML的主要应用（1）

- 设计标记语言：作为元标记语言，XML为用户提供了定义本行业本领域的标记语言的最好工具。目前这一应用的成功例子比比皆是，例如化学领域的CML，数学领域的MathML，移动通信领域的WML等。
- 文件保值：XML良好的保值性和自描述性使它成为保存历史档案，如政府文件、公文、科学研究报告等的最佳选择。
- 数据交换：数据交换的核心问题是信息的标准化，主要解决信息的可理解性问题，包括人和机器对信息的理解。而且，更重要的是机器对信息的识别，并能根据数据进行自动处理。XML 的出现，为信息的标准化提供了有力的工具。

4 XML的主要应用（2）

目前 XML 在数据交换的应用中最为活跃的领域是电子商务(EB)：

- 替代传统的 EDI：电子数据交换（Electronic Data Interchange，简称 EDI）是 XML 出现以前在电子商务中的数据交换标准，用于在异构系统间交换数据并进行事务处理。但由于 EDI 难于使用和实施费用昂贵等缺点使得它没有成为数据交换的主流技术。XML 的出现，使得数据交换更为廉价和方便。
- 智能代理：智能代理能够理解接收到的数据，然后作出相应的反应。如果数据是结构化的 XML 数据，智能代理就能够非常容易利用已有的知识库理解数据。XML 的出现，为智能代理的发展提供了新的推动力。
- Web应用：由于 XML 是由 SGML 特别为 Web 简化的，因此 XML 文档将成为 Web 资源的重要组成部分，XML 使得搜索引擎更为智能和准确。XML 还可以用于建立多层 Web 应用。例如：SOAP、UDDI、WSDL等都是建立在XML的基础之上的。

5 XML的主要工具（1）

XML编辑工具

- **XMLwriter**
- **XML Spy**
- **Stylus Studio**
- **Visual XML**
- **Microsoft XML Notepad**
- **Xray**
- **Editplus**

5 XML的主要工具（2）

XML验证工具

- **Microsoft XML Validator**
- **Microsoft XSL Debugger**

XML解析器

- **MSXML4.0**

XML浏览工具, 几乎所有现代网页浏览器都可以浏览XML。典型的有:

- **IE (Internet Explore)**
- **Firefox**
- **Chrome**

5 XML的主要工具（3）

在Windows上安装IE XML TOOL

- 从网络上搜索并下载 **iexmltls.exe**
- 点击**iexmltix.exe**解压到一个文件夹中
- 鼠标右键点击上述文件夹中的**.inf**的文件进行安装

在ie浏览器上使用iexmltls:

- 在ie浏览器上先点击鼠标右键弹出下拉菜单，再选择其中的“**Validate XML**”或“**View XSL Output**”

5 XML的主要工具（4）

在Windows上安装MSXML4.0

- 从网络上搜索并下面的网址下载 **msxmlchs.msi**
- 点击**msxmlchs.msi**进行安装
 - 默认安装(不安装**SDK**)
 - 自定义安装(可选择安装**SDK**)

本章小结

- ✱ XML标记、元素、属性的概念
- ✱ XML与DTD和XSL的关系
- ✱ XML与HTML的区别