w3school的xml参考手册学习：

XML可以用来描述数据，重点是‘数据是什么’。

HTML则是用来显示数据，重点是‘如何显示数据’。

**什么是XML?**

* XML是E**X**tensible **M**arkup **L**anguage的缩写
* XML是一种类似于HTML的标记语言
* XML是用来描述数据的
* XML的标记不是在XML中预定义的，你必须定义自己的标记
* XML使用文档类型定义（DTD）或者模式（Schema）来描述数据
* XML使用DTD或者Schema后就是自描述的语言

HTML是与显示信息相关的, XML则是与描述信息相关的。

XML允许你定义自己的标记以及文档结构。标签不是固定的

在将来的网页开发中，XML将被用来描述、存储数据，而HTML则是用来格式化和显示数据的。

首先要明确的是：XML是被设计用来存储数据、携带数据和交换数据的，他不是为了显示数据而设计的。

XML数据同样可以以“数据岛”的形式存储在HTML页面中。你仍然可以集中精力到使用HTML格式化和显示数据上去。

XML数据是以纯文本格式存储的，那么XML提供了一种与软件和硬件无关的共享数据方法。

XML的语法规则既简单又严格，非常容易学习和使用。

**XML文档使用了自描述的和简单的语法。**

|  |
| --- |
| <?xml version="1.0" encoding="ISO-8859-1"?>  <note>  <to>Lin</to>  <from>Ordm</from>  <heading>Reminder</heading>  <body>Don't forget me this weekend!</body>  </note> |

文档的第1行:XML声明——定义此文档所遵循的XML标准的版本（实际这个xml文件头声明也可以不加），在这个例子里是1.0版本的标准，使用的是ISO-8859-1 (Latin-1/West European)字符集。

文档的第2行是根元素（就像是说“这篇文档是一个便条”）：

|  |
| --- |
| <note> |

文档的第3--6行描述了根元素的四个子节点（to, from, heading,和 body）：

|  |
| --- |
| <to>Lin</to>  <from>Ordm</from>  <heading>Reminder</heading>  <body>Don't forget me this weekend!</body> |

文档的最后一行是根元素的结束：

|  |
| --- |
| </note> |

所有的xml文档必须有一个结束的标记，xml结构语法严格（第一行xml文档的声明不需要结束标记，因为这不是xml文档的一部分），html相对不严格

xml标记是大小写敏感的

因此在XML文档中开始标记和结束标记的大小写必须保持一致。

|  |
| --- |
| <Message>This is incorrect</message> //错误的  <message>This is correct</message> //正确的 |

所有的xml文档必须有一个根元素，xml文档的第一个元素就是根元素；所有的元素都可以有子元素，子元素必须正确的嵌套在父元素中

属性值必须使用双引号””””，没有双引号的属性值是不符合规定的

如同HTML一样，XML元素同样也可以拥有属性。XML元素的属性以名字/值成对的出现。XML语法规范要求XML元素属性值必须用引号引着。请看下面的两个例子，第一个是错误的，第二个是正确的。

|  |
| --- |
| <?xml version="1.0" encoding="ISO-8859-1"?>  <note date=12/11/99>  <to>Lin</to>  <from>Ordm</from>  <heading>Reminder</heading>  <body>Don't forget me this weekend!</body>  </note> |

|  |
| --- |
| <?xml version="1.0" encoding="ISO-8859-1"?>  <note date="12/11/99">  <to>Tove</to>  <from>Jani</from>  <heading>Reminder</heading>  <body>Don't forget me this weekend!</body>  </note> |

第一个文档的错误之处是属性值没有用引号引着。

正确的写法是: date="12/11/99". 不正确的写法: date=12/11/99.

使用xml文档，空白将被保留，即空白部分不会被解析器自动删除，这一点与html不同：

html中会忽略空白，如下：

<a href="http://download.csdn.net/">

首页</a>

xml中空白被保留，做为值的一部分：

<kongbai>hhhh ccc </kongbai>

使用XML, CR –回车/ LF-换行 被转换为 LF（line feed，新行）

xml的注释和html的注释一样：

<!--这是一个注释 -->

XML文档可以被扩展一边携带更多的信息,对于同一个解析器，必须要的信息会被自动忽略