**XML**

**一、什么是XML？:**

XML指可扩展标记语言

XML是一种标记语句，很类似HTML

XML标签没有被预定义，需要自行定义标签

**二、XML和HTML的差异**：

1.XML为不同的目的而设计

2.XML被设计为传输和存储数据，其焦点是数据的内容，HTML被设计用来显示数据， 其焦点是数据的外观

3.XML意在传输信息，而HTML意在显示信息

## 三、XML的特点

1.XML可以自定义标签

<note>

<to>George</to>

<from>John</from>

<heading>Reminder</heading>

<body>Don't forget the meeting!</body>

</note>

上例中的标签没有在任何XML标准中定义过，比如<to><from>这些标签都是有作者自己发明的

2.XML必须包含根元素

XML 文档必须有一个元素是所有其他元素的**父元素**。该元素称为**根元素**。

<?xml version="1.0" encoding="ISO-8859-1"?>

<note>

<to>George</to>

<from>John</from>

<heading>Reminder</heading>

<body>Don't forget the meeting!</body>

</note>

如上所示，XML文件的根元素为note,剩下的4行描述的是根的子元素

所有的XML标签都必须有关闭标签，在XML中省略标签是非法的，所有的元素都必须有关闭标签

如下所示：

<p>This is a paragraph</p>

<p>This is another paragraph</p>

3.XML标签对大小写敏感

XML 标签对大小写敏感。在 XML 中，标签 <Letter> 与标签 <letter> 是不同的。

必须使用相同的大小写来编写打开标签和关闭标签：

<Message>这是错误的。</message>

<message>这是正确的。</message>

XML必须正确的嵌套：如下所示：由于 <i> 元素是在 <b> 元素内打开的，那么它必须在 <b> 元素内关闭。

<b><i>This text is bold and italic</i></b>

4.XML 的属性值须加引号

与 HTML 类似，XML 也可拥有属性（名称/值的对）。

在 XML 中，XML 的属性值须加引号。请研究下面的两个 XML 文档。第一个是错误的，第二个是正确的：

<note date=08/08/2008>

<to>George</to>

<from>John</from>

</note>

<note date="08/08/2008">

<to>George</to>

<from>John</from>

</note>

在第一个文档中的错误是，note 元素中的 date 属性没有加引号。

5.标签可以自由命名

6.标签成对出现

7.标签可以重名

## 四、XML 中的注释

在 XML 中编写注释的语法与 HTML 的语法很相似：

<!-- This is a comment -->

## 五、什么是 XML 元素？

**XML 元素**指的是从（且包括）开始标签直到（且包括）结束标签的部分。

元素可包含其他元素、文本或者两者的混合物。元素也可以拥有属性。

<bookstore>

<book category="CHILDREN">

<title>Harry Potter</title>

<author>J K. Rowling</author>

<year>2005</year>

<price>29.99</price>

</book>

<book category="WEB">

<title>Learning XML</title>

<author>Erik T. Ray</author>

<year>2003</year>

<price>39.95</price>

</book>

</bookstore>

在上例中，<bookstore> 和 <book> 都拥有**元素内容**，因为它们包含了其他元素。<author> 只有**文本内容**，因为它仅包含文本。

在上例中，只有 <book> 元素拥有**属性** (category="CHILDREN")。

**六、XML 命名规则**

XML 元素必须遵循以下命名规则：

* 名称可以含字母、数字以及其他的字符
* 名称不能以数字或者标点符号开始
* 名称不能以字符 “xml”（或者 XML、Xml）开始
* 名称不能包含空格

七、XML读取文本

from xml.dom import minidom

#打开文件

root =minidom.parse(filepath)

firstnode =root.getElementsByTagName(onenode)[0]

secondnode=firstnodee.getElementsByTagName(twonodee)[0].firstChild.data