Introduction to XML Extensible Markup Language

|  |  |
| --- | --- |
| **HTML** | **XML** |
| HTML is an abbreviation for HyperText Markup Language. | XML stands for eXtensible Markup Language. |
| HTML was designed to display data with focus on how data looks. | XML was designed to be a software and hardware independent tool used to transport and store data, with focus on what data is. |
| HTML is a markup language itself. | XML provides a framework for defining markup languages. |
| HTML is a presentation language. | XML is neither a programming language nor a presentation language. |
| HTML is case insensitive. | XML is case sensitive. |
| HTML is used for designing a web-page to be rendered on the client side. | XML is used basically to transport data between the application and the database. |
| HTML has it own predefined tags. | While what makes XML flexible is that custom tags can be defined and the tags are invented by the author of the XML document. |
| HTML is not strict if the user does not use the closing tags. | XML makes it mandatory for the user the close each tag that has been used. |
| HTML does not preserve white space. | XML preserves white space. |
| HTML is about displaying data,hence static. | XML is about carrying information,hence dynamic. |

**What is XML?**

XML is the Extensible Markup Language. It improves the functionality of the Web by letting you identify your information in a more accurate, flexible, and adaptable way. It is extensible because it is not a fixed format like HTML (which is a single, predefined markup language). Instead, XML is actually a metalanguage a language for describing other languages which lets you design your own markup languages for limitless different types of documents

XML  
► User definable tags  
► Content driven  
► End tags required for well formed documents  
► Quotes required around attributes values  
► Slash required in empty tags

**Who is responsible for XML?**

**Ans.** XML is a project of the World Wide Web Consortium (W3C), and the development of the specification is supervised by an XML Working Group. A Special Interest Group of co-opted contributors and experts from various fields contributed comments and reviews by email.

##### ****Does XML replace HTML?****

**Ans.** No. XML itself does not replace HTML. Instead, it provides an alternative which allows you to define your own set of markup elements. HTML is expected to remain in common use for some time to come, and the current version of HTML is in XML syntax. XML is designed to make the writing of DTDs much simpler than with full SGML. (See the question on DTDs for what one is and why you might want one.

XML uses the same building blocks that HTML does: **elements, attributes, and values.**

An XML element is the most basic unit of your document.

An XML element can contain other elements and text.

An element has an opening tag with a name ( written between < and > ) and sometimes attributes.

The name should describe the element's purpose.

XML is self-describing

<name>

<first>Bond</first>

<last>James</last>

</name>

**Hierarchies in XML**

<name>

<first>Bond</first>

<middle>JK</middle>

<last>James</last>

</name>

<name> is parent of <first>, <middle> and <last>.

<first>, <middle> and <last> are sibling.

Text is a child of the element.

The text Bond is a child of <first>.

<name>

<first>Bond</first>

<middle>JK</middle>

<last>James</last>

</name>

<name> element has only other elements, and not text, then it is said to have element content.

<first>, <middle>, and <last> have only text as children, they are said to have simple content.

Elements can contain both text and other elements(mixed content)

<doc>

<parent>this is some <em>text</em> in my element</parent>

</doc>

<parent>has three children:

A text child containing the text 'this is some' and an <em>child.

Another text child containing the text inmyelement.

<name>

<first>Bond</first>

<middle>JK</middle>

<last>James</last>

</name>

Data from the beginning of a start-tag to the end of an end-tag is called an element.

<first>is a start-tag

</first>is an end-tag

<first>John</first>is an element

!--This is not well-formed XML!-->

<comparison>6 is < 7 & 7 > 6</comparison>

You can't put < or & characters into PCDATA.

There are two ways you can get around this: escaping characters, or enclosing text in a CDATA section.

To escape the < or &, you replace any < with &lt; and any & with &amp;.

The previous XML example could be made well formed:

<comparison>6 is &lt; 7 &amp; 7 &gt; 6 </comparison>

&lt; and &amp; are known as entity references.

The following entities are defined in XML:

&amp; - the & character

&lt; - the < character

&gt; - the > character

&apos; - the ` character

&quot; - the " character