XML TAGS

Start Tag

The beginning of every non-empty XML element is marked by a start-tag. Following is an exampled of start-tag:

<college>

End Tag

Every element that has a start tag should end with an end-tag. Following is an example of end-tag:

</college>

Note, that the end tags include a solidus ("/") before the name of an element.

Empty Tag

The text that appears between start-tag and end-tag is called content. An element which has no content is termed as empty. An empty element can be represented in two ways as follows:

A start-tag immediately followed by an end-tag as shown below:

<course></course>

A complete empty-element tag is as shown below:

<course />

Empty-element tags may be used for any element which has no content.

XML Naming Rules

XML elements must follow these naming rules:

- Names can contain letters, numbers, and other characters
- Names cannot start with a number or punctuation character
- Names cannot start with the letters xml (or XML, or Xml, etc)
- Names cannot contain spaces

Any name can be used, no words are reserved.

Entity References

Some characters have a special meaning in XML.

If you place a character such as '<' within an XML element, an error is created because it is interpreted by the parser as the start of a new element. This will generate an XML error:

<marks>if marks < 90 then</marks>

To avoid this error, replace the "<" character with an entity reference:

<marks>if marks &lt; 90 then</marks>

There are 5 predefined entity references in XML:

- &lt; < less than
- &gt; > greater than
- &amp; & ampersand
- &apos; ' apostrophe
- &quot; " quotation mark

Note: Only the characters "<" and "&" are strictly illegal in XML. The greater than character is legal, but it is a good habit to replace it.

Comments in XML

The syntax for writing comments in XML is similar to that of HTML.

<!-- This is a comment -->

White-space is Preserved in XML

HTML truncates multiple white-space characters to one single white-space:

HTML: university of   lucknow

Output: university of lucknow.

With XML, **the white-space in a document is not truncated.**

DTD- DOCUMENT TYPE DEFINITION

A Document Type Definition (DTD) defines the legal building blocks of an XML document. It defines the document structure with a list of legal elements and attributes.

A DTD can be declared inline inside an XML document, or as an external reference.

Internal DTD Declaration

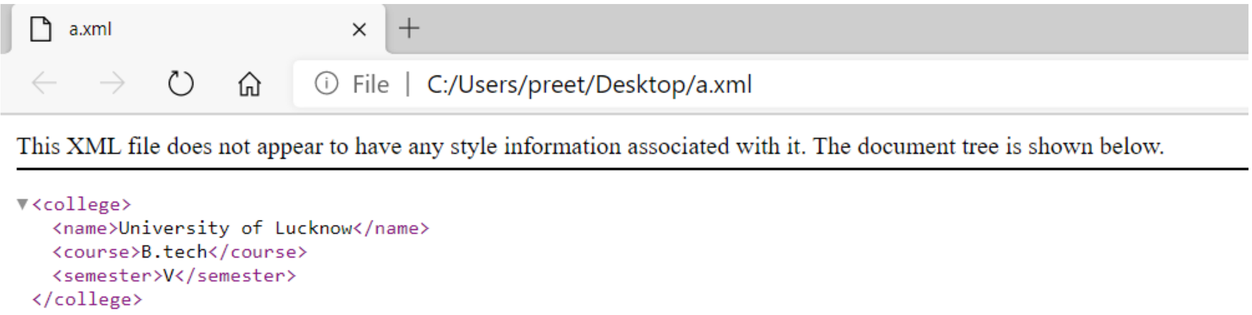
If the DTD is declared inside the XML file, it should be wrapped in a DOCTYPE definition with the following syntax:

<!DOCTYPE root-element [element-declarations]>

Example code:

```
<?xml version="1.0"?>
<!DOCTYPE college [
<!ELEMENT college (name,course,semester)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT course (#PCDATA)>
<!ELEMENT semester (#PCDATA)>
]>
<college>
<name>University of Lucknow</name>
<course>B.tech</course>
<semester>V</semester>
</college>
```

Output:



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<college>
   <name>University of Lucknow</name>
   <course>B.tech</course>
   <semester>V</semester>
 </college>
```

The DTD above is interpreted like this:

- !DOCTYPE college defines that the root element of this document is college
- !ELEMENT college defines that the college element contains elements:

"name, course, semester''.

- !ELEMENT name defines the name element to be of type "#PCDATA"
- !ELEMENT course defines the course element to be of type "#PCDATA"
- !ELEMENT semester defines the semester element to be of type "#PCDATA"

External DTD Declaration

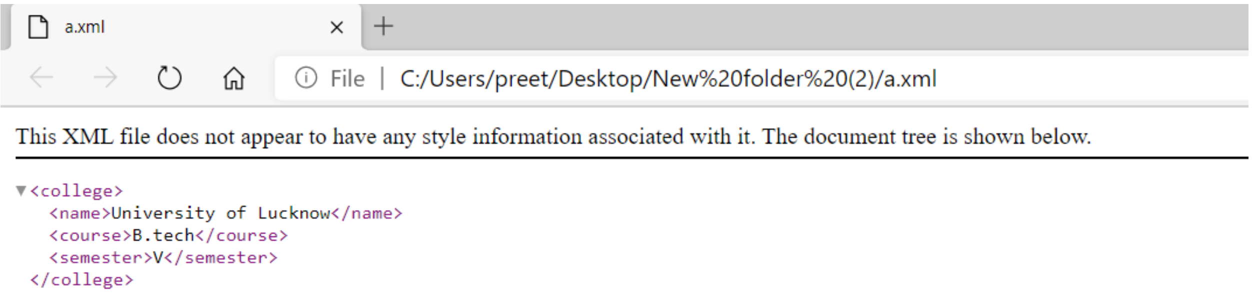
If the DTD is declared in an external file, it should be wrapped in a DOCTYPE definition with the following syntax:

<!DOCTYPE root-element SYSTEM "filename">

Example code:

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

<!DOCTYPE college SYSTEM "college.dtd">

<college>

<name>University of Lucknow</name>

<course>B.tech</course>

<semester>V</semester>

</college>
```

Output:



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<college>
   <name>University of Lucknow</name>
   <course>B.tech</course>
   <semester>V</semester>
 </college>
```