

XML

Session 2: Using and Creating XML Documents

Using and Creating XML Documents

- XML is comprised of elements, attributes and values.
- An XML document should start with a **processing instruction** that indicates the version of XML that is used.
- Example:
`<?xml version="1.0" encoding="UTF-8"?>`

- Every XML document has a “root element” that contains every element in the document.
- Each element must have a matching closing tag
- Elements can contain elements or character data
- Example:

```
<?xml version="1.0" encoding="UTF-8"?>
<money>
  <currency>
    <code>USD</code>
    <exchangerate>1.77659</exchangerate>
    <description>US Dollars</description>
  </currency>
  <currency>
    <code>EUR</code>
    <exchangerate>1.47045</exchangerate>
    <description>Euro</description>
  </currency>
</money>
```

- XML tags are case sensitive.
- It is a good programming practice to always use lower case to describe your elements.
- Comments may be expressed in XML and the syntax is similar to HTML comments.

Example:

```
<?xml version="1.0" encoding="UTF-8"?>
<software>
  <!-- The license attribute indicates the number of
  Client licenses -->
  <cd id="Win7" license="3">Microsoft Windows 7</cd>
  <cd id="Office2010" license="5">Microsoft Office 2010</
  cd>
  <cd id="Sublime" license="1">Sublime Text 3</cd>
</software>
```

Filename: SoftwareVersion.xml

- Elements can have attributes
- Attributes are added to the start tag of an element
- Attributes must have a value
- Cannot have duplicate attribute names on an element
- Attributes can contain significant data, be reference markers or “describe” the data in the element (meta-data)

Example of XML with attributes

```
<?xml version="1.0" encoding="UTF-8"?>
<currencyforecast title="Currency Forecast"
  year="2005">
  <index id="EUR/USD">
    <month id="January">
      <exchangerate>1.2500</exchangerate>
    </month>
    <month id="March">
      <exchangerate>1.2500</exchangerate>
    </month>
    <month id="December">
      <exchangerate>1.2600</exchangerate>
    </month>
  </index>
</currencyforecast>
```

Filename: CurrencyForecast.xml

- XML Documents **must** be **well-formed**
 - Matching start and end tags for elements
 - Elements must not overlap
 - Example of an **invalid** XML document:
`<?xml version="1.0" encoding="UTF-8"?>`
`<example>`
`<incorrect><element></incorrect></element>`
`</example>`
 - One root element
 - Example of an **invalid** XML document:
`<?xml version="1.0" encoding="UTF-8"?>`
`<example>`
`<incorrect></incorrect>`
`</example>`
`<example>`
`<incorrect></incorrect>`
`</example>`

Attribute values must be quoted

```
<?xml version="1.0" encoding="UTF-8"?>
<stafflist>
  <name id="124899">
    <firstname>Johann</firstname>
    <surname>Strauss</surname>
  </name>
</stafflist>
```

Filename: AttributeExample.xml

An element may not have two attributes with the same name.

Example of an **invalid** xml document:

```
<?xml version="1.0" encoding="UTF-8"?>
<stafflist>
  <name id="124899" id="NI7877DW">
    <firstname>George</firstname>
    <surname>Gershwin</surname>
  </name>
</stafflist>
```

Filename: InvalidXMLExample.xml

- Comments must not appear inside tags
- “<” and “&” cannot appear “unescaped”.
- Escape characters:
 - & &
 - ' '
 - > >
 - < <
 - " "

Make use of Elements or Attributes?

- Elements can describe structure as attributes can not.
- It is not possible for attributes to contain multiple values. Elements can contain child elements which can contain multiple values.
- Elements can be extended as attributes may not.
- In general, it is easier to manipulate XML elements rather than attributes in most programming languages.

Validating XML documents

- An XML document with correct syntax is called a **WELL FORMED** XML document.
- XML documents may be validated against Document Type Definitions (DTD) or XML Schemas. DTDs and schemas will be covered in more details in the next sessions.

There are various XML validators

- Browsers have a built-in XML parser.
- W3Schools on-line XML validator
 - www.w3schools.com/dom/dom_validate.asp
- Oxygen XML is already installed on the lab machines and can be used to create and validate XML files.
 - www.oxygenxml.com
- Netbeans is already installed on the lab machines and can be used to create and validate XML files. It is also free and open source and can be downloaded from:
 - <http://netbeans.org/>
- Netbeans and Oxygen XML are tools that run on Mac OS X and Linux.
- Altova's XMLSPY. This software can be installed on your machine and can assist you in creating and validating XML files.

References

Useful links:

- http://webdesign.about.com/od/xml/a/xml_content.htm
- www.w3schools.com/xml/xml_elements.asp
- www.w3schools.com/xml/xml_attributes.asp