CITS1231 Web Technologies

Introduction to XML, Formatting XML using CSS
Introducing XML

- XML stands for Extensible Markup Language. A markup language specifies the structure and content of a document.

- Because it is extensible, XML can be used to create a wide variety of document types.

- XML is a subset of the Standard Generalized Markup Language (SGML).

- SGML was introduced in the 1980s and is very complex and can be costly.

- These reasons led to the creation of Hypertext Markup Language (HTML), a more easily used markup language.

- XML can be seen as sitting between SGML and HTML – easier to learn than SGML, but more robust and flexible than HTML.
## XML Vocabularies

<table>
<thead>
<tr>
<th>XML Vocabulary</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel Definition Format (CDF)</td>
<td>Automatic delivery of information from Web publishers to PCs, PDAs, cell phones, and other information devices</td>
</tr>
<tr>
<td>Chemical Markup Language (CML)</td>
<td>Coding of molecular and chemical information</td>
</tr>
<tr>
<td>Extensible Hypertext Markup Language (XHTML)</td>
<td>HTML written as an XML application</td>
</tr>
<tr>
<td>Mathematical Markup Language (MathML)</td>
<td>Presentation and evaluation of mathematical equations and operations</td>
</tr>
<tr>
<td>Musical Markup Language (MML)</td>
<td>Display and organization of music notation and lyrics</td>
</tr>
<tr>
<td>Open Financial Exchange (OFX)</td>
<td>Exchange of financial data between financial institutions, businesses, and consumers via the Internet</td>
</tr>
<tr>
<td>Real Simple Syndication (RSS)</td>
<td>Distribution of news headlines and syndicated columns</td>
</tr>
<tr>
<td>Synchronized Multimedia Integration Language (SMIL)</td>
<td>Editing of interactive audiovisual presentations involving streaming audio, video, text, and any other media type</td>
</tr>
<tr>
<td>Voice Markup Language (VoiceXML)</td>
<td>Creation of audio dialogues that feature synthesized speech, digitized audio, and speech recognition</td>
</tr>
</tbody>
</table>
Well-Formed and Valid XML Documents

- An XML document is *well-formed* if it contains no syntax errors and fulfills all of the specifications for XML code as defined by the W3C.

- An XML document is *valid* if it is well-formed and also satisfies the rules laid out in the DTD or schema attached to the document.
The Structure of an XML Document

• XML documents consist of three parts
  – The prolog
  – The document body
  – The epilog

• The prolog is optional and provides information about the document itself

• The document body contains the document’s content in a hierarchical tree structure.

• The epilog is also optional and contains any final comments or processing instructions.
Creating the Prolog

• The prolog consists of four parts in the following order:
  – XML declaration
  – Miscellaneous statements or comments
  – Processing instructions
  – Document type declaration
The XML Declaration

- The XML declaration is always the first line of code in an XML document. It tells the processor what follows is written using XML. It can also provide any information about how the parser should interpret the code.

- The complete syntax is:
  
  `<?xml version="version number" encoding="encoding type" standalone="yes / no" ?>`

- A sample declaration might look like this:
  
  `<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>`
Inserting Comments

• Comments or miscellaneous statements go after the declaration. Comments may appear anywhere after the declaration.

• The syntax for comments is:

  <!- - comment text - ->

• This is the same syntax for HTML comments
Elements

- Elements are the basic building blocks of XML files.
- Elements contain an opening tag and a closing tag
  - Content is stored between tags
- A **closed** element, has the following syntax:
  
  `<element_name>Content</element_name>`
  
  - Example:
    
    `<Artist>Miles Davis</Artist>`

- An **open** or **empty** element is an element that contains no content. They can be used to mark sections of the document for the XML parser.
  
  - `<element_name />`
Element

- Element names are case sensitive

- Elements can be nested,
  - Nested elements are called child elements.
  - All elements must be nested within a single document or root element. There can be only one root element.
  - Elements must be nested correctly. Child elements must be enclosed within their parent elements.
  - For example,
    <tracks>Kind of Blue
    <track>So What (9:22)</track>
    <track>Blue in Green (5:37)</track>
    </tracks>
Working with Attributes

- An **attribute** is a feature or characteristic of an element. Attributes are text strings and must be placed in single or double quotes. The syntax is:

  `<element_name attribute="value"> ... </element_name>`
Adding elements to the Jazz.XML File

This figure shows the revised document

```xml
<item>
  <title>Kind of Blue</title>
  <artist>Miles Davis</artist>
  <tracks>
    <track length="9:22">So What</track>
    <track length="9:46">Freddie Freeloader</track>
    <track length="5:37">Blue in Green</track>
    <track length="11:33">All Blues</track>
    <track length="9:26">Flamenco Sketches</track>
  </tracks>
</item>

<item>
  <title>Cookin'</title>
  <artist>Miles Davis</artist>
  <tracks>
    <track length="5:57">My Funny Valentine</track>
    <track length="4:53">Blues by Five</track>
    <track length="4:22">Airegin</track>
    <track length="13:03">Turnê-Up</track>
  </tracks>
</item>

<item>
  <title>Blue Train</title>
  <artist>John Coltrane</artist>
  <tracks>
    <track length="10:39">Blue Train</track>
    <track length="9:06">Moment’s Notice</track>
    <track length="7:11">Locomotion</track>
    <track length="7:55">I’m Old Fashioned</track>
    <track length="7:03">Lazy Bird</track>
  </tracks>
</item>
```
References for Special Character

- Special characters, such as the symbol for the British pound, can be inserted into your XML document by using a character reference. The syntax is:
  
  \&\#nnn;

- *Character* is a character reference number or name from the ISO/IEC character set.

- Character references in XML are the same as in HTML.
  - \&nbsp; a blank space
  - \&\#169; ©
Commonly Used Special Characters

This figure shows commonly used character reference numbers

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Character Reference</th>
<th>Entity Reference</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>©</td>
<td>&amp;#169;</td>
<td></td>
<td>Copyright symbol</td>
</tr>
<tr>
<td>®</td>
<td>&amp;#174;</td>
<td></td>
<td>Registered trademark symbol</td>
</tr>
<tr>
<td>™</td>
<td>&amp;#153;</td>
<td></td>
<td>Trademark symbol</td>
</tr>
<tr>
<td>&lt;</td>
<td>&amp;#60;</td>
<td>&lt;</td>
<td>Less than symbol</td>
</tr>
<tr>
<td>&gt;</td>
<td>&amp;#62;</td>
<td>&gt;</td>
<td>Greater than symbol</td>
</tr>
<tr>
<td>&amp;</td>
<td>&amp;#38;</td>
<td>&amp;</td>
<td>Ampersand</td>
</tr>
<tr>
<td>&quot;</td>
<td>&amp;#22;</td>
<td>&quot;</td>
<td>Double quote</td>
</tr>
<tr>
<td>´</td>
<td>&amp;#27;</td>
<td>'</td>
<td>Apostrophe (single quote)</td>
</tr>
<tr>
<td>£</td>
<td>&amp;#163;</td>
<td></td>
<td>Pound sign</td>
</tr>
<tr>
<td>€</td>
<td>&amp;#128;</td>
<td></td>
<td>Euro sign</td>
</tr>
<tr>
<td>¥</td>
<td>&amp;#165;</td>
<td></td>
<td>Yen sign</td>
</tr>
</tbody>
</table>
Character References

This figure shows the revised Jazz.XML file

```xml
<item>
  <title>Kind of Blue</title>
  <price>US: $11.99</price>
  <price>UK: £16.31</price>
  <artist>Miles Davis</artist>
  <track length="9:22">So What</track>
  <track length="9:46">Freddie Freeloader</track>
  <track length="5:37">Blue in Green</track>
  <track length="11:33">All Blues</track>
  <track length="9:20">Flamenco Sketches</track>
</item>

<item>
  <title>cockin'</title>
  <price>US: $7.99</price>
  <price>UK: £6.29</price>
  <artist>Miles Davis</artist>
  <track length="5:37">My Funny Valentine</track>
  <track length="9:53">Blues by Five</track>
  <track length="4:22">Airegin</track>
  <track length="13:03">Tune-up</track>
</item>

<item>
  <title>Blue Train</title>
  <price>US: $8.09</price>
  <price>UK: £6.29</price>
  <artist>John Coltrane</artist>
  <track length="10:39">Blue Train</track>
  <track length="9:06">Moment's Notice</track>
  <track length="7:11">Locomotion</track>
  <track length="7:30">21st Century Schizoid Man</track>
  <track length="7:03">Lazy Bird</track>
</item>
```
Parsed Character Data

- **Parsed character data**, or pcdata consists of all those characters that XML treats as parts of the code of XML document
  
  - The XML declaration
  - The opening and closing tags of an element
  - Empty element tags
  - Character or entity references
  - Comments
CDATA Sections

- A CDATA section is a large block of text the XML processor will interpret only as text.

- The syntax to create a CDATA section is:

  ```
  <![CDATA[
  Text Block
  ]]>  
  ```
In this example, a CDATA section stores several HTML tags within an element named HTMLCODE:

```html
<htmlcode>
<![CDATA[
  <h1>The Jazz Warehouse</h1>
  <h2>Your Online Store for Jazz Music</h2>
]]>
</htmlcode>
```
CDATA Sections

This figure shows the revised Jazz.XML file

```xml
<items>
  <message>
    <![CDATA[
    Here are some of the latest specials from the Jazz Warehouse. Please note that all Miles Davis & John Coltrane CDs will be on sale for the month of March.
    ]]>.
  </message>
  <item>
    <title>Kind of Blue</title>
    <price>US: $11.99</price>
    <price>UK: £10.99</price>
    <artist>Miles Davis</artist>
    <tracks>
      <track length="9:22">So what</track>
      <track length="9:46">Freddie Freeloader</track>
      <track length="5:37">Blue in Green</track>
      <track length="11:24">All Blues</track>
      <track length="9:26">Flamenco Sketches</track>
    </tracks>
  </item>
  ...
</items>
```

Figure 1-14
Parsing an XML Document

Figure 1-15  Parsing an XML document

The author writes an XML document in an XML editor.

The document is submitted to an XML processor, which evaluates and parses the document (checking for errors in syntax or document structure).

The processed document is then displayed to the user in whatever format is used by the XML processor.
Displaying an XML Document in a Web Browser

- XML documents can be opened in Internet Explorer or Firefox.

- If there are no syntax errors, IE will display the document’s contents in an expandable/collapsible outline format including all markup tags.

- Netscape will display the contents but neither the tags nor the nested elements.
Displaying an XML Document in a Web Browser
Linking to a Style Sheet

- Link the XML document to a style sheet to format the document. The XML processor will combine the style sheet with the XML document and apply any formatting codes defined in the style sheet to display a formatted document.

- There are two main style sheet languages used with XML:
  - Cascading Style Sheets (CSS) and Extensible Style Sheets (XSL)
Linking to a Style Sheet

• There are some important benefits to using style sheets:
  – By separating content from format, you can concentrate on the appearance of the document
  – Different style sheets can be applied to the same XML document
  – Any style sheet changes will be automatically reflected in any Web page based upon the style sheet
Applying a Style to an Element

- To apply a style sheet to a document, use the following syntax:
  
  \[\text{selector \{attribute1:value1; attribute2:value2; \ldots\}}\]

  - selector is an element (or set of elements) from the XML document.
  - attribute and value are the style attributes and attribute values to be applied to the document.

- For example:
  
  \[\text{artist \{color:red; font-weight:bold\}}\]

  will display the text of the artist element in a red boldface type.
Applying a Style to an Element

• For example:

    artist {color:red; font-weight:bold}

• will display the text of the artist element in a red boldface type.
Creating Processing Instructions

- The link from the XML document to a style sheet is created using a processing statement.

- A **processing instruction** is a command that gives instructions to the XML parser.

- For example:

  ```xml
  <?xml-stylesheet type="style" href="sheet" ?>
  ```

  *style* is the type of style sheet to access and *sheet* is the name and location of the style sheet.
The jw.css Style Sheet

This figure shows the cascading style sheet stored in the jw.css file.

```css
message {
    display: block; width: 400px; color: blue; text-align: center;
    font-size: 10pt; font-family: Arial, Helvetica, sans-serif;
    border: 3px solid blue; background-color: ivory;
    margin: 10px; padding: 15px
}
item {
    display: block; font-size: 14pt; color: red;
    font-family: Arial, Helvetica, sans-serif;
    margin: 20px
}
title {
    display: block; font-size: 16pt; color: blue;
    font-weight: bold;
    font-family: Arial, Helvetica, sans-serif
}
price, priceuk {
    color: black; font-size: 12pt; font-weight: bold;
    font-family: Times New Roman, Times, Serif;
    margin-left: 20px
}
artist {
    display: block; font-size: 12pt; color: black;
    font-style: italic; font-weight: bold;
    font-family: Times New Roman, Times, Serif;
    margin-left: 70px
}
track {
    display: list-item; font-size: 9pt; color: black;
    list-style-type: circle;
    font-family: Arial, Helvetica, sans-serif;
    margin-left: 35px
}
```
Linking to the jw.css Style Sheet

This figure shows how to link the JW.css style sheet to the Jazz.xml file

```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!-- This document contains data on Jazz Warehouse special offers -->
<?xml-stylesheet type="text/css" href="jw.css" ?>
<items>
  <message>
    <![CDATA[
      Here are some of the latest specials from the Jazz Warehouse. Please note that all Miles Davis & John Coltrane CDs will be on sale for the month of March.
    ]]>  
  </message>
</items>
```

processing instruction to access the jw.css style sheet
The jazz.xml formatted with the jw.css Style Sheet

Here are some of the latest specials from the Jazz Warehouse. Please note that all Miles Davis & John Coltrane CDs will be on sale for the month of March.

**Kind of Blue**
- US: $11.99  UK: £8.39
- Miles Davis
  - So What
  - Freddie Freeloader
  - Blue in Green
  - All Blues
  - Flamenco Skies

**Cookin'**
- US: $7.99  UK: £5.59
- Miles Davis
  - My Funny Valentine
  - Blues by Five
  - Airegin
  - Tune-Up

**Blue Train**
- John Coltrane
  - Blue Train
  - Moment's Notice
  - Locomotion
  - I'm Old Fashioned
  - Lazy Bird

This figure shows the formatted jazz.xml file