TUTORIAL 1

CREATING AN XML DOCUMENT

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.

INTRODUCING XML

- XML is a subset of the Standard Generalized Markup Language (SGML) which was introduced in the 1980s. SGML 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 than HTML.

THE LIMITS OF HTML

- HTML was designed for formatting text on a Web page. It was not designed for dealing with the content of a Web page. Additional features have been added to HTML, but they do not solve data description or cataloging issues in an HTML document.

- Because HTML is not extensible, it cannot be modified to meet specific needs. Browser developers have added features making HTML more robust, but this has resulted in a confusing mix of different HTML standards.

THE LIMITS OF HTML

- HTML cannot be applied consistently. Different browsers require different standards making the final document appear differently on one browser compared with another.

THE 10 PRIMARY XML DESIGN GOALS

1. XML must be easily usable over the Internet
2. XML must support a wide variety of applications
3. XML must be compatible with SGML
4. It must be easy to write programs that process XML documents
5. The number of optional features in XML must be kept to a minimum, ideally zero

THE 10 PRIMARY XML DESIGN GOALS — CONTINUED

6. XML documents should be clear and easily understood by nonprogrammers
7. The XML design should be prepared quickly
8. The design of XML must be exact and concise
9. XML documents must be easy to create
10. Terseness in XML markup is of minimum importance

XML VOCABULARIES

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

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 STRUCTURE OF AN XML DOCUMENT

- 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.

THE STRUCTURE OF AN XML DOCUMENT: 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 STRUCTURE OF AN XML DOCUMENT: 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" ?>

THE STRUCTURE OF AN XML DOCUMENT: 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

ELEMENTS

- A closed element, has the following syntax:

  <element_name>Content</element_name>

- Example:

  <Artist>Miles Davis</Artist>

ELEMENT

- Element names are case sensitive

- Elements can be nested, as follows:

```
<tracks>Kind of Blue
    <track>So What ((:22)</track>
    <track>Blue in Green (5:37)</track>
</tracks>
```

ELEMENTS

- Nested elements are called child elements.
- Elements must be nested correctly. Child elements must be enclosed within their parent elements.