Introduction to XML

SSE 3200
Web-based Services

Michigan Technological University
Nilufer Onder
What is XML?

• Acronym for: eXtensible Markup Language.
• It is a text-based format for representing structured information.
• Markup refers to the fact that different parts of the text are identified by tags.
• These tags designate information.
• It was designed for program-to-program communication but it is human readable.
A simple example

<ssecourses>
  <course number="SSE3200">
    <name>Design and Analysis of Web-based Services</name>
    <coursedesc>The strategy behind developing web-based service systems will be the focus of the course. Topics will include flowcharting, cost estimating, performance measurement, database management, and alpha and beta testing. A semester project will illustrate the use of web based services technology.</coursedesc>
  </course>
</ssecourses>
A simple example (cont'd)

<COURSE NUMBER="SSE2300">

(NAME)Service Systems Dynamics and Design</name>

<COURSEDESC> The class introduces system dynamics principles and explores the effect of system structure and uncertain events and interactions on system behavior. Students will learn to analyze uncertainties and dynamic feedback loops defining complex system behavior and use them to design system components or manage system outcomes. Basic simulation and system modeling methods and their application to strategic decision-making will be introduced. Specific methods introduced include queueing theory, and system dynamics.

</COURSEDESC>

</COURSE>

</SSECOURSES>
The characters which make up and XML document are divided into **markup** and **content**.

- All strings that begin with the character “<” and end with a “>” are **markup** constructs called **tags**.

- All strings that begin with the special character “&” and end with a “;” are **markup** constructs called **entity references**.
Markup and content (cont’d)

• Strings of characters which are not markup are content.

• Example:
  
  `<name>` Design &amp; Analysis of Web-based Services `</name>`

`<name>` is a start tag.  
`</name>` is an end tag. 
`&amp;` is an entity reference for the ampersand character (&) 
The rest are content characters.
Tags

- **Start tags** begin a tagged region (an *element*)
  e.g., `<name>`

- **End tags** finish a tagged region
  e.g., `</name>`

- **Empty element tags** tag empty regions
  e.g., `<name/>`  (no name is specified)
Element

- An element is a logical component of a document which either begins with a start-tag and ends with a matching end-tag, or consists only of an empty element tag.
- Elements can be nested, i.e., included in other elements.
- XML requires one element that contains all others (root element).
- SSECOURSES is the root element in the example.
Element name requirements

- Must begin with a name or underscore.
- The first character may be followed by any combination of letters, numbers, or other ASCII characters.
- Elements beginning with “XML” whether capitalized or not are reserved.
Attribute

- An **attribute** is a markup construct that consists of a name/value pair.
- It is written within a start-tag or empty-element tag.
- In the course XML listing, “name” is an attribute.
- Attribute values must always be quoted, even when they are numbers.
XML declaration

- Most web documents start by declaring information.
- An XML declaration might look like this: `<?xml version="1.0" ?>`
- The declaration is optional.
Advantages of XML

- Flexible: One can define their own elements.
- Self-describing: As long as meaningful tag names are used, people can understand them easily.
- Structured: Documents are better organized into structures that allow for easier reference.
Advantages of XML (cont'd)

- Redundant: It is verbose. For example, every end tag must be supplied. It makes it easy to catch errors.

- Tools available: There are many tools for working with XML descriptions as well as many tools that work with XML descriptions. These descriptions and tools enhance each others’ value in a positive feedback loop.
XML with CSS

- CSS stands for Cascading Style Sheets
- Look at the example at:
  http://www.comptechdoc.org/independent/web/xml/guide/xmlexample.html
- A note about color:
  Colors on computer screens are typically specified using three components: red, green, blue.
  A color is a three part number where each part specified the intensity of red, green, or blue.
References

- XML information at:
  http://www.w3.org/standards/xml/core

- Wikipedia's entry on XML:

- Online xml tutorial at:
  http://www.w3schools.com/xml/default.asp

- XML document at:
  http://www.comptechdoc.org/independent/web/xml/guide