

# XML Essentials

# What Is XML?

- **XML (Extensible Markup Language)** is a general-purpose specification for creating custom markup languages
  - It is classified as an **extensible language** because it allows its users to define their own elements
  - Its primary purpose is to facilitate the sharing of **structured data** across different information systems, and it is used both to encode documents and to serialize data
  - XML is a generic framework for storing any amount of text or any data whose **structure can be represented as a tree**
  - XML is a W3C Recommendation

# Well-formed documents: XML syntax

- A XML document has exactly one **root element**
  - This means that the text must be enclosed between a root start-tag and a corresponding end-tag

- The following is a "well-formed" XML document:

```
<book>This is a book.... </book>
```

- **Comments** can be placed anywhere outside of a tag

```
<!-- This is a comment. -->
```

- The root element can be preceded by an optional **XML declaration**

```
<?xml version="1.0" encoding="UTF-8"?>
```

# The basic syntax

- The basic syntax for one **element** is:

```
<name attribute="value">Content</name>
```

*Start-tag*

*Each attribute name must appear only once in an element*

*Attribute values must always be quoted, using single or double quotes*

*Some text which may again contain XML elements. So, a generic XML document contains a tree-based data structure.*

*End-tag*

# Special Syntax For Empty Content

- XML provides special syntax for representing an element with empty content
- The following three examples are equivalent in XML:

```
<foo></foo>  
<foo />  
<foo/>
```

- An empty-element may contain attributes:

```
<info author="John Smith"  
       genre="science-fiction"  
       date="2009-Jan-01" />
```

# Not Well-formed XML

- XML requires that elements are properly nested
  - **elements may never overlap!**

*<!-- WRONG! NOT WELL-FORMED XML! -->*

```
<title>Book on Logic<author>Aristotle<title>Another Book on Logic  
<author>Boole</title></author></title></author>
```

*<!-- Correct: well-formed XML. -->*

```
<books>  
<book><title>Book on Logic</title>  
<author>Aristotle</author></book>  
<book>  
<title>Another Book on Logic <author>Boole</author> </title>  
</book>  
</books>
```

# References

- XML overview:  
<http://en.wikipedia.org/wiki/XML>
- XML definition  
<http://www.w3.org/XML/Core/#Publications>