

XML

November 3, 2008

Outline

1 Introduction

2 Syntax and Usage

- Syntax
- Well Formed and Valid XML
- Displaying XML on Web
- Parsing

3 XML Databases

Example

```
<message>  
  <from>Yanni</from>  
  <to>Bruce</to>  
  <subject>Reminder</subject>  
  <body>Let's play this weekend!</body>  
</message>
```

XML

- eXtensible
- Markup
- Language

XML

- **eXtensible**
- **Markup**
- **Language**
- Meta-language
- Similar to other MLs
 - $\text{\TeX}$, HTML
- Source of other languages:
 - RSS, XHTML, MathML, GraphML, SVG, MusicXML, ...

Introduction

- Store data in a structure (not display)
 - Doesn't *do* anything
- Plain text
- Separate data from
 - application
 - display
- Inter-platform
- Inter-application
 - web, desktop

Uses of XML

- Information exchange
 - single format
 - across Internet
 - across applications
- Encode documents
- Serialize objects (human-readable, multi-language)

Outline

1 Introduction

2 Syntax and Usage

- Syntax
- Well Formed and Valid XML
- Displaying XML on Web
- Parsing

3 XML Databases

XML Structure

- Tree structure

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<message>
  <from>Yanni</from>
  <to>Bruce</to>
  <subject>Reminder</subject>
  <body type="plain_text">Let's play this weekend!</body>
</message>
```

- `<message>`: root element, always
- `<to>`, `<subject>`, ...: child elements
- type attribute
- child, parent, siblings

Tree Structure

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<library>
  <book>
    <title> Gulliver's Travels </title>
    <isbn> 12342-123-32 </isbn>
    <pages> 442 </pages>
    <chapter>
      <number> 1 </number>
      <title> Introduction </title>
    </chapter>
    ...
  </book>

  <book>
    <title> War and Peace </title>
    <isbn> 7832-431-23 </isbn>
    ...
  </book>
</library>
```

Tree Structure

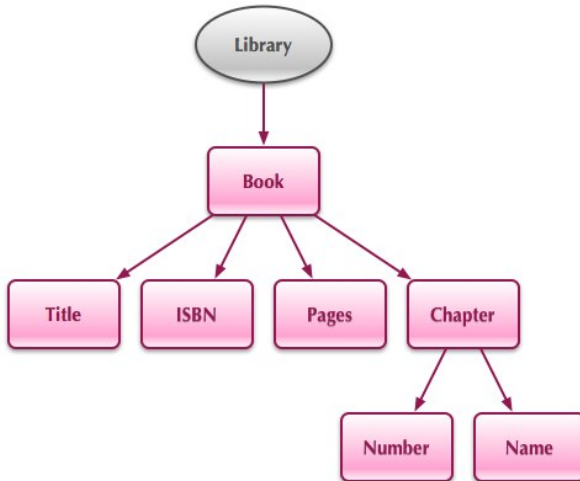


Figure: Tree Structure of an XML document

Syntax Rules

- End tag is required

```
<book>  
  ....  
  ....  
</book>
```

- Root element is required

- Case-sensitive

Incorrect:

```
<book> ... </book>
```

- Proper nesting is required

```
<book>  
  <name>  
    Othello  
  </name>  
</book>
```

- Comment:

```
<!-- Comments are good -->
```

Elements

- From start to end tag of same name
 - Text or *nested* elements
- Follow encoding
- Extensibility: new elements don't affect previous processing practice

Attributes

- Always in quotes
- Like attributes to <a> and in HTML
- Qualify an element; not actual data

- Meta-data

```
<file type="txt">  
  <name>Fall_Budget.txt</name>  
  <location>/home/raheel/Desktop</location>  
</file>
```

Attributes vs. Elements

- Type of information
 - Essential → element
 - Peripheral, supportive → attribute
- Extensibility
 - Atomic → attribute
 - Extensible, structured → element
- Readability
 - Human readable → element
 - Computer processing only → attribute