

XML Fundamentals



XML Elements

- XML documents consist of one or more elements.
- Elements consist of a pair of tags and (optionally) enclosed text.
`<TITLE>The XML Companion</TITLE>`
- Elements may have attributes with their values in single or double quotes.
`<TITLE type="book">The XML Companion</TITLE>`
- XML trees (nesting): Elements may contain other elements.
`<REFERENCE>
 <TITLE type="book">The XMLCompanion</TITLE>
</REFERENCE>`
- **Empty elements** may be self closing.
`<TITLE type="book" name="XML Companion"> </TITLE>
<TITLE type="book" name="XML Companion" />`



XML Elements (cont.)

- XML mixed content:

<TITLE>The XML Companion

<author>O'reilly</author>

</TITLE>

- XML grammar : Data can be represented as PCDATA or as attribute values:

- <book>

- <title>The Forty-nine Steps</title>

- </book>

- <book title="The Forty-nine Steps">

- ...

- </book>



XML Elements

Attribute & element naming:

- A-Z, a-z, 0-9, “_”, “-”, “.” Are the allowed literals.
- %, \$, /, ' are forbidden..
- Spaces within the element name are forbidden.
- Entity references: > , < , & , " , '
- XML comments:

<!-- this is a comment -->

- The XML declaration:

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

- Style sheet processing instruction:

```
<?xml-stylesheet type="text/css"
      href="cd_catalog.css"?>
```



XML Rules

1. All elements *must* be closed.
2. Elements cannot overlap.
`bold <I>bold italic italic</I>` this is illegal !
3. XML is case sensitive
`<father>` and `<FATHER>` are *different* tags
4. Attributes *must* be enclosed in inverted commas
`<TITLE type="book" />`
5. An element must not have 2 attributes with the same name.
6. There is only one root element per XML document



A Simple XML Document

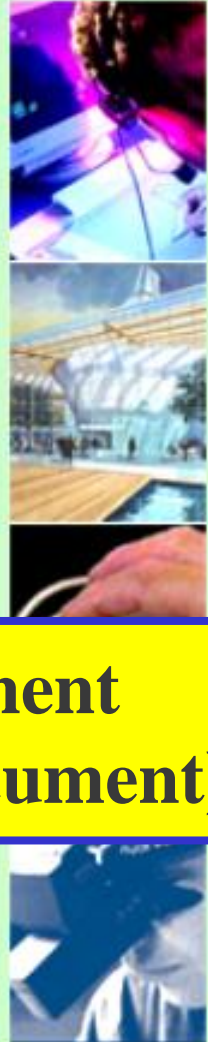
```
<?xml version="1.0" ?>
```

XML declaration

```
<booklist title="Some XML Books">
```

**Root element
(one per document)**

```
</booklist>
```



A Simple XML Document



```
<?xml version="1.0" ?>  
<!DOCTYPE booklist SYSTEM "books.dtd" >  
  
<booklist title="Some XML Books">  
.....  
.....  
</booklist>
```

A Simple XML Document

```
<?xml version="1.0" ?>  
<!DOCTYPE booklist SYSTEM "books.dtd" >  
<!-- This is a comment -->  
  
<booklist title="Some XML Books">  
  
</booklist>
```

This is a comment



A Simple XML Document

```
<?xml version="1.0" ?>  
<!DOCTYPE booklist SYSTEM "books.dtd" >  
<!-- This is a comment -->  
<?xml-stylesheet type="text/css" href= "iti-xml2.css"?>
```

```
<booklist title="Some XML Books">
```

```
</booklist>
```

This defines the stylesheet used



Adding Content

```
<booklist title="Some XML Books">
<book>
  <author>
    <name>St. Laurent</name>
    <initial>S</initial>
  </author>
  <date>1998</date>
  <title edition="Second">XML: A Primer</title>
  <publisher>MIS Press</publisher>
  <website href="http://www.simonstl.com/xmlprim/" />
  <rating stars="4"/>
</book>
</booklist>
```



Parsing XML Documents

- What is meant by parsing?.

- Types of parsers:

Validating parser :

validate XML document against DTD or Xml schema

Non validating parser :

Well formed XML document (against 6 rules of XML)



Tools for Working with XML

- Simple Text Editors (ex : Notepad)
- **Altova XMLSpy**

