XML

Lecture 32

Robb T. Koether

Hampden-Sydney College

Mon, Apr 23, 2018

Outline

What is XML?

- XML = Extensible Markup Language.
- XML is similar to HTML except that the programmer may create whatever tags he wants.
- For example

```
<name>Hampden-Sydney College</name>
```

- XML allows the data to be organized as a structured set of elements.
- The tags are meant to convey the semantics of the element.
- It is meant to be processed by software that can parse the tags.

Outline

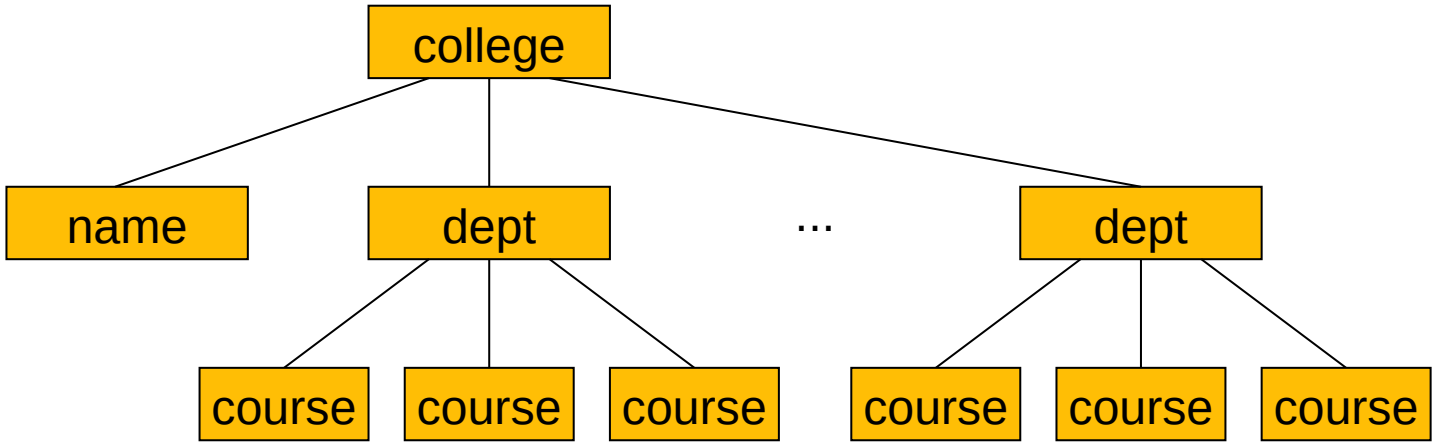
Structured Data in XML

XML Code

```
<college>
    <name>Hampden-Sydney College</name>
    <dept>
        <name>Math/CS</name>
        <course>COMS 331</course>
        <course>COMS 480</course>
    </dept>
    <dept>
        <name>Chemistry</name>
                :
    </dept>
        :
</college>
```

- The structure must contain a single root element.
- In this example, `<college>` is the root element.
- Every opening tag must be matched by a corresponding closing tag.

The Structure of the Data

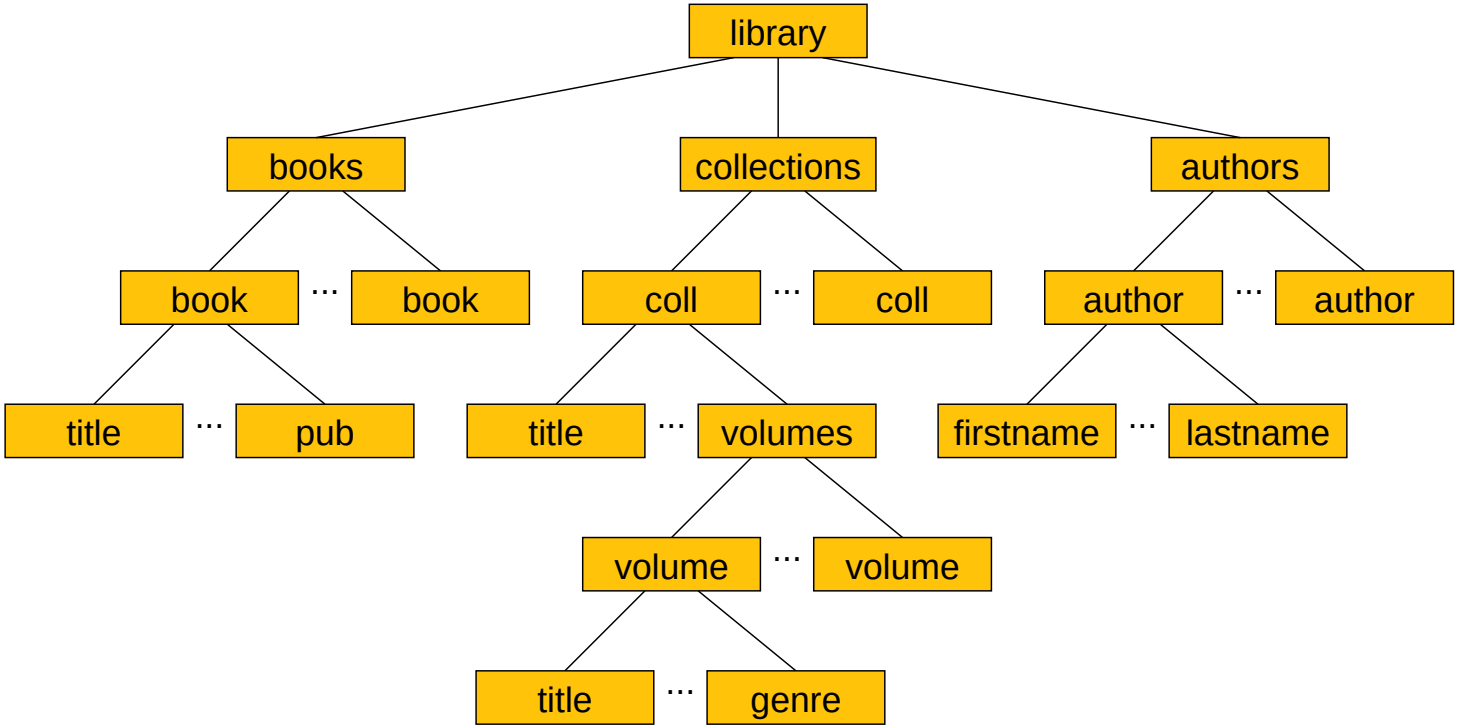


The Library Database

- Consider a library database.
- There are tables for
  - Books
  - Collections of volumes
  - Authors, etc.
- Each Book consists of
  - Title
  - Author(s)
  - Publisher
  - Date
  - Number of pages
  - Genre

The Library Database

- The library database as a tree structure.



The Library Database as XML

Library XML

```
<library>
  <books>
    <book>
      <title>Kidnapped</title>
        <authID>123</authID>
            :
    </book>
        :
  </books>
  <collections>
    <collection>
        :
    </collection>
        :
  </collections>
      :
  <authors>
    <author authID="123">
      <fname>Robert</fname>
      <mname>Louis</mname>
      <lname>Stevenson</lname>
    </author>
          :
  </authors>
  <publishers>
        :
  </publishers>
</library>
```

- We may express this by the above structure.

Untagged Elements

Untagged Elements

```
<library>
    <book>
        Kidnapped
        <authID>123</authID>
            :
    </book>
            :
```

- Although it is strongly discouraged, it is permissible to have elements that are untagged.
- Without the tags, we lose the semantics of the element.
- Is “Kidnapped” the title, or does it mean that someone stole the book and is holding it for ransom?
- Furthermore, we access elements by their tags, so accessing untagged elements is problematic.

Outline

Self-closing Tags

Self-closing Tags

```
<myselfclosingtag/>
```

- XML permits self-closing tags.
- However, without a separate closing tag, there is no content in the element.

Attributes

Attributes

```
<img src="./image/me.jpg" width="1024" height="768"/>
```

No Attributes

```
<img>
    <source>./image/me.jpg</source>
    <width>1024</width>
    <height>768</height>
</img>
```

- It is possible to include the information about an element in a sequence of attributes, as in the `<img>` tag in HTML.
- However, this is not a good idea.
- It would have been better to define the image element with `<source>`, `<width>`, and `<height>` child elements.

Attributes

Using Attributes for Data or Properties

```
<author fname="Mark" lname="Twain"/>
<title lang="en">Tom Sawyer</title>
```

- Attributes *must* have values and the values *must* be enclosed in quotes.
- The purpose of an attribute is to provide information *about* the data, not the data itself.
- Therefore, the first example above is inappropriate.
- The second example shows a more appropriate use of attributes.

Outline

XML Namespaces

XML Namespaces

```
<table>
    <tr><th>Chairs</th><th>Tables</th></tr>
</table>
<table>
    <wood>cherry</wood>
    <length>48</length>
    <width>18</width>
    <height>12</height>
</table>
```

- Suppose an XML file includes both HTML `<table>` tags and an XML tag for an ordinary table as a piece of furniture.
- This structure is ambiguous.
- How can this ambiguity be resolved?

XML Namespaces

XML Namespaces

```
<furniture xmlns:h="http://www.w3.org/TR/html4/"
    xmlns:f="http://www.hsc.edu/myfurniture/">
<h:table>
    <tr><th>Chairs</th><th>Tables</th></tr>
</h:table>
<f:table>
    <wood>cherry</wood>
    <length>48</length>
    <width>18</width>
    <height>12</height>
</f:table>
```

- Tags may be associated with a namespace.
- The namespace is a URL, but it is not used by the XML processor.
- Typically, the URL refers to a human-readable web page that explains the meaning of the tags.

XML Namespaces

XML Namespaces

```
<furniture xmlns="http://www.w3.org/TR/html4/"
    xmlns:f="http://www.hsc.edu/myfurniture/">
<table>
    <tr><th>Chairs</th><th>Tables</th></tr>
</table>
<f:table>
    <wood>cherry</wood>
    <length>48</length>
    <width>18</width>
    <height>12</height>
</f:table>
```

- If the namespace is not given a name, then it is the default namespace.

Outline

GPS Files

- GPS devices store their data in `.gpx` files, which are structured as XML files.
- Common elements are `<wpt>`, `<trkpt>`, and `<rtept>`, for waypoints, trackpoints, and route points.
- An XSD file specifies an XML Schema Definition.
- Check out some of Garmin's XSD files.

Outline

Assignment

Assignment

- Visit the W3Schools websites
  - `http://www.w3schools.com/xml/`
- Visit the sections labeled “XML HOME” through “XML Attributes.”