B336 Advanced Internet Computing

Namespaces, Infosets and Other Key XML Standards

Learning Objectives

- Understand where the standards for Namespaces, infosets, XLink and XML Schema fit in the XML technologies.

Learning Objectives

- In the scheme of what we are doing in this unit:
    - We are studying how to use XML as an important set of Internet technologies to use as solutions in different areas (for example, the problem defined in Assignment 2).
    - Although we will not have time to cover all available XML-related standards, you should at least be aware of some key ones, so you can decide to go deeper into them if you need to.

Lecture Outline

- Namespaces in XML
- Infosets and Canonical XML
- XML Schema
- XLink

Name Conflicts in XML

- Since any element names can be created in XML, very often a name conflict will occur when two different documents use the same names describing two different types of elements.

- In an environment where there are a lot of XML document being processes automatically by software, this can create many problems.

Example Name Conflict

```
<table>
   <tr>
   <td>Entry</td>
   </tr>
</table>
```

Can’t put these two sets of tags together in a single document.

```
<table>
   <name>African Coffee Table</name>
   <width>80</width>
   <length>120</length>
</table>
```

Resolving Name Conflicts using Prefixes

```
<html:table>
     <html:tr>
     <html:td>Entry</html:td>
     </html:tr>
</html:table>
```

```
<furniture:table>
   <furniture:name>African Coffee Table
   </furniture:name>
   <furniture:width>80</furniture:width>
   <furniture:length>120</furniture:length>
</furniture:table>
```

Namespaces

```
<html:table xmlns:html="http://www.w3.org/TR/html4/">
     <html:tr>
     <html:td>Entry</html:td>
     </html:tr>
</html:table>
```

```
<furniture:table xmlns:furniture="http://www.table.org/">
   <furniture:name>African Coffee Table
   </furniture:name>
   <furniture:width>80</furniture:width>
   <furniture:length>120</furniture:length>
</furniture:table>
```

Namespaces and the xmlns attribute

- Instead of using only prefixes, an **xmlns attribute** has been added to the <table> tag to give the element prefix a **qualified name** associated with a **namespace**.

Namespace for XSLT

- From last lecture:

```
<?xml version="1.0"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:template match="/">
    <HTML>
      <BODY>
        <xsl:foreach select="/poetry/anthology/poem/stanza/line">
          <P>
            <xsl:value-of select="."/>
          </P>
        </xsl:foreach>
      </BODY>
    </HTML>
  </xsl:template>
</xsl:stylesheet>
```

Mixing Namespaces

```
<my_information
 xmlns:html="http://www.w3.org/TR/html4/"
 xmlns:furniture="http://www.table.org/" >

    <html:table>
      ...
    </html:table>

    <furniture:table>
      ...
    </furniture:table>

</my_information>
```

XML Information Set (infosets)

- With all the information being created in XML documents, it would be very useful to be able to compare these documents.
    - We can’t do that by comparing line by line because two documents may have the same information even when the text is not the same. Eg.
        - » Tags are in different order.
        - » Different comment lines
        - » Use internal vs external DTDs, but the DTDs are the same.
        - » etc.

XML Information Set (infosets)

- The infoset specification attempts to define a set of **information items** about an XML document, so that we can do those comparisons.
    - Each one of the items have their own properties.
    - We can compare two document's the items and their properties rather than using the original text.

- http://www.w3.org/TR/xml-infoset

XML Information Set (infosets)

- An XML document has an information set if
  - it is well-formed, and
  - satisfies the namespace constraints.
- There is no requirement for an XML document to be valid in order to have an information set.

Canonical XML

- XML documents can also be compared without having to get their associated infosets.

- Canonical XML is a specification of XML syntax which is extremely strict.
    - Two XML documents with the same “information” will convert to the same canonical XML document form.

- http://www.w3.org/TR/xml-c14n

XLink

- XLink allows elements to be inserted into XML documents in order to create and describe links between resources.

- It uses XML syntax to create structures that can describe the simple unidirectional hyperlinks such as in HTML, as well as more sophisticated links.

- http://www.w3.org/TR/xlink/

Example XLink

```
<poem xmlns:xlink="http://www.w3.org/1999/xlink"
 xlink:type = "extended"
 xlink:title = "A Sick Rose by William Blake" >

   <title>A Sick Rose</title>

   <author xmlns:xlink="http://www.w3.org/1999/xlink"
    xlink:type = "locator"
    xlink:href = "http://www.poetry.org/Wblake/biography.xml">
      William Blake
   </author>

   <poem_text xmlns:xlink="http://www.w3.org/1999/xlink"
    xlink:type = "locator"
    xlink:title = "The text for A Sick Rose by William Blake"
    xlink:href = "http://www.poetry.org/Wblake/ASickRose.xml">
   </poem_text>

</poem>
```

XML Schema

- XML Schema does the same job as DTD, but is a lot more sophisticated:
    - Like DTD, it specify the structure of XML documents
- http://www.w3.org/XML/Schema

Why Have XML Schema?

- People are dissatisfied with DTD's:
  - It uses different syntax compared to XML documents
    - » so can't be checked by standard XML syntax validators, not easy to generate using XSLT, etc.
  - Has very limited datatype capabilities
    - » eg can't specify that a piece of data must be a number between 0 and 100.

Some Features of XML Schema

- Some advancement over DTDs:
  - Enhanced datatypes
    - over 41 versus 10 in DTDs
    - Can create your own datatypes
  - Written in the XML syntax
  - Object-oriented concepts (eg. can extend or restrict a type)
  - Can specify element content as being unique
  - Can define multiple elements with the same name but different content
  - Can define substitutable elements - e.g., the "subway" element is substitutable for the "train" element.
  - etc.

Other W3C's XML Standards

- We will not be covering the above standards in any depth beyond what is above.
- The same applies to other W3C XML standards, which no doubt will grow in number.
  - In the future, you will have to keep up with the details of these standards and how to use them yourself, if you want to operate properly in the XML environment.
  - In Assignment 2, you will be rewarded for demonstrating the capability to go out and learn beyond what is presented in this unit (lectures, labs, etc).