

THE AUSTRALIAN NATIONAL UNIVERSITY

XML & RSS

COMP1710 – New Media

Eric McCreath

2004

Contents

2

- XML - Extensible Markup Language
- DTD - Document Type Definitions
- XHTML
- XSLT - Extensible Stylesheet Language Transformations
- RSS - Really Simple Syndication

XML

3

- Extensible Markup Language(XML) is a standard which defines the syntax for document markup. The format used is simple and readable by humans.
- XML can be used by all sorts of documents including:
 - Web pages,
 - Vector Graphics,
 - Legal documents,
 - word processor documents, and
 - real estate listings.

XML

4

- XML is a format that is understood by a variety of platforms. Including desktop machine, mobile phones, games machines.
- XML makes it easy to communicate between different platforms.
- There is a lot of libraries for programmers to use. This simplifies reading and writing information in a XML format.

XML Format

5

- XML is a plain text with tags to mark up the content.

```
<person>
  Pascal Vuylsteker
</person>
```

Tags must match up.

XML Format

6

- XML tags can be nested.

```
<person>
  <first_name>Pascal</first_name>
  <last_name> Vuylsteker </last_name>
  <phone_number> 58192 </phone_number>
</person>
```

No white space in the tags names.

XML Format

7

- Tags can be used to mark up parts of the text within a tag.

```
<person>
  <first_name>Pascal</first_name>
  <last_name> Vuylsteker </last_name>
  <phone_number> 58192 </phone_number>
  <note>
    Pascal lectures <subject> COMP1710 </subject>.
  </note>
</person>
```

XML Format

8

- XML documents should have an XML declaration on the first line. Although this is not strictly needed.

```
<?xml version="1.0" encoding="US-ASCII" standalone="yes">
<person>
  <first_name>Pascal</first_name>
  <last_name> Vuylsteker </last_name>
  <phone_number> 58192 </phone_number>
  <!-- They may also have comments. -->
</person>
```

- XML documents should be well formed. This includes start tags matching end tags, elements not overlapping and many other checks.

XML Format

9

- Other items you may find in XML documents include:
 - tag attributes,
 - empty elements, and
 - processing instructions.

```
<?xml version="1.0" encoding="US-ASCII" standalone="yes">
<person>
  <mac_user/>
  <first_name><?php print $name; ?></first_name>
  <last_name> Vuylsteker </last_name>
  <phone_number type="ext"> 58192 </phone_number>
</person>
```

DTD

10

- A Document Type Definition(DTD) provides a way of exactly specifying the format and type of content within a XML document.

```
<!ELEMENT person (first_name,last_name?,phone_number)>
<!ELEMENT first_name (#PCDATA)>
<!ELEMENT last_name (#PCDATA)>
<!ELEMENT phone_number (#PCDATA)>
<!ATTLIST phone_number type CDATA #IMPLIED>
```

XHTML

11

- XHTML is an XML version of HTML. HTML is a lot like XML, however, it is not well formed like XML. eg in HTML it is okay not to put in end tags for <p> and tags. Also there is no empty tags in HTML.
- The latest browsers can display XHTML document without any trouble. Most legacy browsers will do an okay job as well.

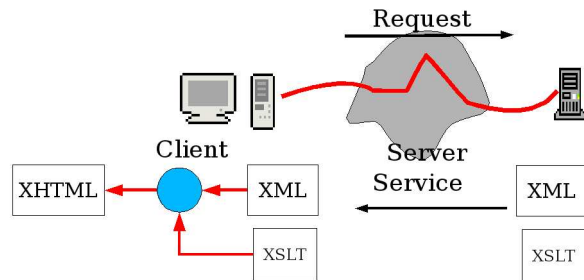
XHTML

12

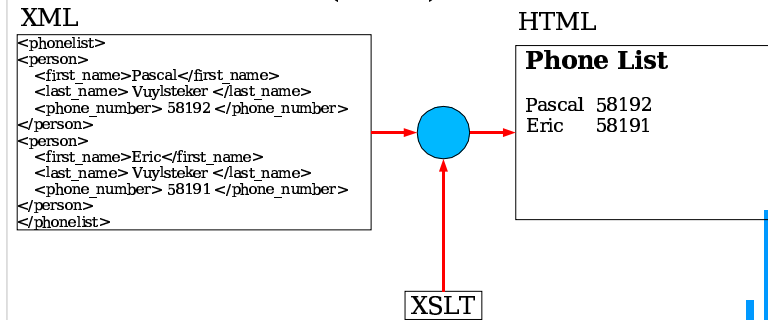
- You have to look very closely to notice the difference!

<pre><HTML> <HEAD> <TITLE> Example </TITLE> </HEAD> <BODY> <H1> Example </H1> <HR SIZE=1 NOSHADE> <p> This is a simple example. </P> <p> It show the different between: HTML, and XML. </BODY> </HTML></pre>	HTML	<pre><!DOCTYPE html PUBLIC> <html xmlns="http://www....."> <head> <title> Example </title> </head> <body> <h1> Example </h1> <hr style="height: 1"/> <p> This is a simple example.</p> <p> It show the different between: HTML, and XML. </p> </body> </html></pre>	XHTML
--	------	--	-------

- XML may be served up directly from a web server along with either (CSS or XSLT) style sheets.
- The browser can then use the style sheet to transform XML into HTML(or XHTML).



- Extensible Stylesheet Language Transformations(XSLT)

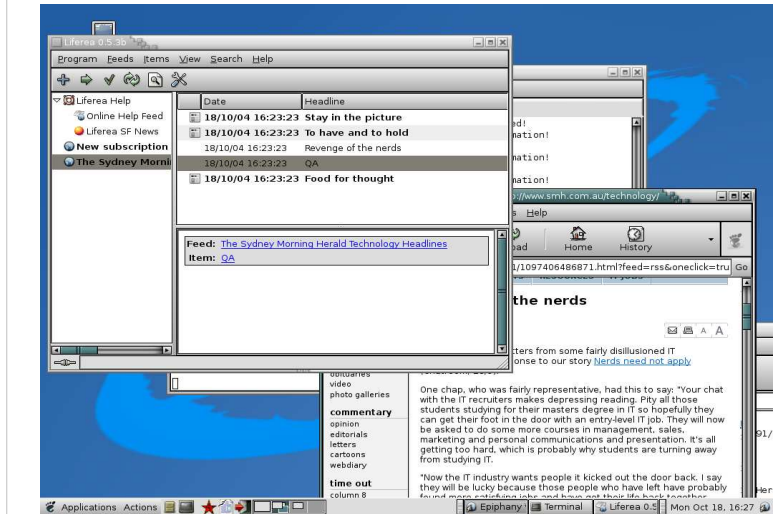


```
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="phonelist">
<html>
<head><title>Phone List</title></head>
<body>
<h1>Phone List</h1>
<table>
<xsl:apply-templates select="person"/>
</table>
</body>
</html>
</xsl:template>
<xsl:template match="person">
<tr>
<td> <xsl:value-of select="first_name"/> </td>
<td> <xsl:value-of select="phone_number"/> </td>
</tr>
</xsl:template>
</xsl:stylesheet>
```

This template generates the HTML document.

This template generates each row in the phone list.

- Really Simple Syndication(RSS) is an XML format for publishing short news-like articles.
- RSS is used in a variety of different ways including:
 - News papers(eg SMH)
 - Blogs
 - Groups working on projects
- They are a number newsreaders which enable you to organize a number of different RSS feeds. They high light unread articles and give you a uniform organized interface into this information.



- RSS is a very simple XML format.

```
<rss version="0.91">
  <channel>
    <title> COMP1710 News </title>
    <link>http://cs.anu.edu.au/students/comp1710</link>
    <description> Features some new that relates to comp1710. </description>
    <language>en-us</language>

    <item>
      <title>Last Week of Lectures! </title>
      <link>http://escience.anu.edu.au/lecture/comp1710/index.en.html</link>
      <description> It is the last week of new content in comp1710.
        Next week there will be a review of the subject.
      </description>
      <pubDate>October 18, 2004</pubDate>
    </item>
    <item>
      <title>Labs : check your marks on streams</title>
      <link>http://cs.anu.edu.au/streams/</link>
      <description> If there is a zero, it means that you haven't properly
        submitted your work as specified on the lab sheet.
      </description>
      <pubDate>October 18, 2004</pubDate>
    </item>
  </channel>
</rss>
```

- Harold and Means, XML in a nutshell, O'Reilly, 2001.
- Mark Pilgrim, What is RSS?, <http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html> 2002