CSP2103-4102: Markup Languages

Lecture 7: RSS Data Feeds

Learning Outcomes

- Learn what RSS feeds are
- What they do
- In what context they are used
- How they are built
- How they are presented
- Elements of RSS
- The role of XSLT with RSS

Definition of RSS

- Opinion varies, some say;
- Really Simple Syndication
- Others says;
- Rich Site Summary
- Whatever the acronym stands for, RSS is basically an xml language for describing and presenting information within a ‘channel’

What Does RSS Do?

- There are numerous definitions about the purpose of RSS, but typically it is for presenting and distributing written content, such as news and blog content, auction sites, airline timetables and job vacancies
- In many ways, it is the equivalent of the old news groups that were popular in the early days of the web
- It is also like the news Agents software that was briefly popular in the late 90's, where you could select news content on a specific topic and be notified of updates

http://www.w3schools.com/rss/spam.gif

Choose your news
With RSS you can choose to view the news you want, the news that is interesting to you, and relevant to your work.

Remove unwanted information
With RSS you can (finally) separate wanted information from unwanted information (the spam), and remove unwanted news, remove disturbing noise.

Increase your site traffic
With RSS you can create your own news channel, and publish it to the Internet and make your site more accessible for more people.

The Development of RSS

- News groups, agents and subscription email systems tended to lose popularity, due largely to two things
- One, SPAM. Any time you signed up for a service to receive news emails, you could expect your email to be sold to spammers around the world
- Two, news groups based on the nntp system tended to lose popularity as they were harder to configure and use than web sites
- RSS has been around in various forms since xml itself, and is a defined standard based on xml itself
- Has become particularly popular since about 2002 onwards (with the coming of RSS 2.0)

RSS Technology

- RSS is essentially a push-pull technology, in that viewers can subscribe to a RSS feed (whatever its content)
- When the content changes, the RSS reader software (or content aggregator) downloads the latest content
- The reader subscribes to the content they want (given an RSS channel exists for it)
- No email need be exchanged or received
- When a user is no longer interest in a specific feed, they simply unsubscribe using their RSS reader

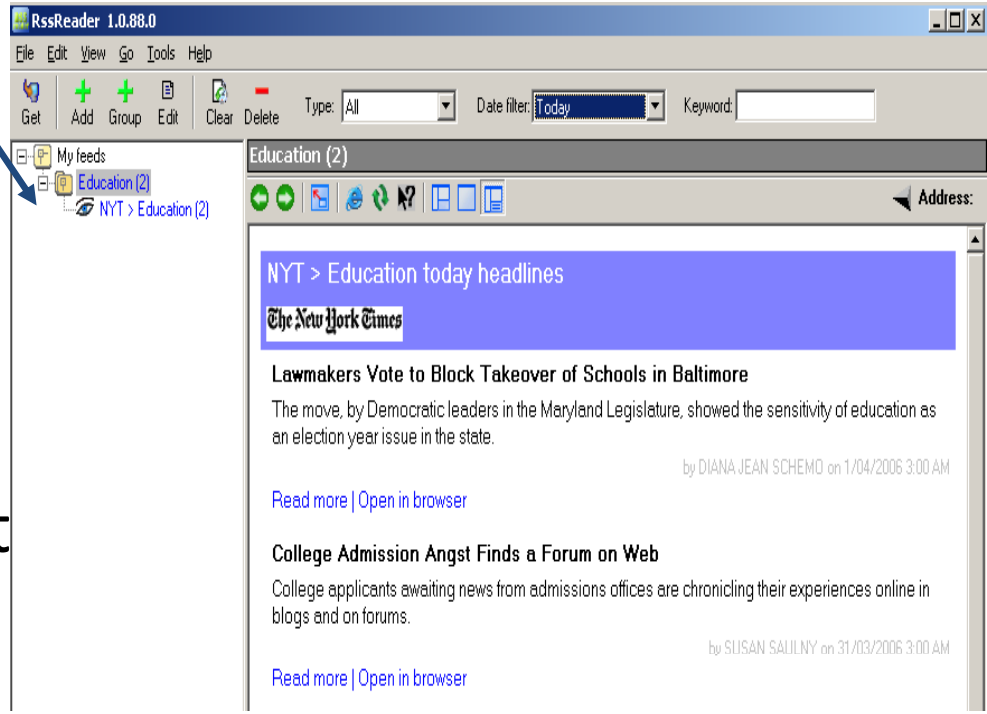
RSS Versions

- Dec. 1997 - Dave Winer developed scriptingNews
- Mar. 1999 - RSS 0.90 was developed by Netscape, which supported the scriptingNews format. This was simply XML with a RDF Header
- Jun. 1999 - scriptingNews 2.0b1 was developed by Dave Winer at UserLand. This included Netscape's RSS 0.90 features
- Jul. 1999 - RSS 0.91 was an attempt by Netscape to move towards a more standard format and included most features from scriptingNews 2.0b1 but they removed the RDF header
- Jul. 1999 - UserLand uses RSS 0.91 and gets rid of scriptingNews
- Netscape discontinues their RSS development
- Jun. 2000 - Official RSS 0.91 specification release from UserLand
- Aug. 2000 - RSS 1.0 developed by a group lead by Rael Dornfest at O'Reilly. This format uses RDF and namespaces. Because of its name it is often confused as being a new version of 0.91, but this is a completely new format with no ties to RSS 0.91
- Dec. 2000 - RSS 0.92 developed by Dave Winer at UserLand and includes optional elements
- Sep. 2002 - RSS 2.0 is designed by Dave Winer after leaving Userland
- Jul. 2003 - RSS 2.0 specification released through Harvard under a Creative Commons license

http://www.w3schools.com/rss/rss_history.asp

RSS Example

- Here is an RSS reader which is being used to subscribe to the Education channel of an online newspaper
- This example shows that I have two unread items in the Education channel
- For each item I can either read it in the RSS reader or open it to a Web browser (taking me to a full html version of the story)
- In many ways, RSS is a content abstract system



RSS Example

- The code RSS code for the headlines shown in the previous slide looks like this

```
<?xml version="1.0" encoding="iso-8859-1" ?>
- <rss version="2.0">
  - <channel>
      <title>NYT > Education</title>
      <link>http://www.nytimes.com/pages/education/index.html?partner=rssnyt</link>
      <description />
      <copyright>Copyright 2006 The New York Times Company</copyright>
      <language>en-us</language>
      <lastBuildDate>Sat, 1 Apr 2006 03:05:01 EST</lastBuildDate>
    - <image>
        <url>http://graphics.nytimes.com/images/section/NytSectionHeader.gif</url>
        <title>NYT > Education</title>
        <link>http://www.nytimes.com/pages/education/index.html</link>
      </image>
    - <item>
        <title>Lawmakers Vote to Block Takeover of Schools in Baltimore</title>
        <link>http://www.nytimes.com/2006/04/01/us/01educ.html?
          ex=1301547600&en=37bdfd4d4e642d17&ei=5088&partner=rssnyt&emc=rss</link>
        <description>The move, by Democratic leaders in the Maryland Legislature on Friday, showed the
          sensitivity of education as an election year issue in the state.</description>
        <author>DIANA JEAN SCHEMO</author>
        <pubDate>Sat, 01 Apr 2006 00:00:00 EDT</pubDate>
        <guid isPermaLink="false">http://www.nytimes.com/2006/04/01/us/01educ.html</guid>
      </item>
    - <item>
        <title>College Admission Angst Finds a Forum on Web</title>
        <link>http://www.nytimes.com/2006/03/31/education/31chatter.html?
          ex=1301461200&en=68102a55bf704112&ei=5088&partner=rssnyt&emc=rss</link>
        <description>College applicants awaiting news from admissions offices are chronicling their
          experiences online in blogs and on forums.</description>
        <author>SUSAN SAULNY</author>
        <pubDate>Fri, 31 Mar 2006 00:00:00 EDT</pubDate>
        <guid isPermaLink="false">http://www.nytimes.com/2006/03/31/education/31chatter.html</guid>
      </item>
```

RSS Example

- Notice the Channel definition code
- Notice the Item definition code
- Notice that this;

Lawmakers Vote to Block Takeover of Schools in Baltimore

The move, by Democratic leaders in the Maryland Legislature, showed the sensitivity of education as an election year issue in the state.

by DIANA JEAN SCHEMO on 1/04/2006 3:00 AM

Read more | Open in browser

- Is represented in RSS as this;

```
- <item>
    <title>Lawmakers Vote to Block Takeover of Schools in Baltimore</title>
    <link>http://www.nytimes.com/2006/04/01/us/01educ.html?
      ex=1301547600&en=37bdfd4d4e642d17&ei=5088&partner=rssnyt&emc=rss</link>
    <description>The move, by Democratic leaders in the Maryland Legislature on Friday, showed the
      sensitivity of education as an election year issue in the state.</description>
    <author>DIANA JEAN SCHEMO</author>
    <pubDate>Sat, 01 Apr 2006 00:00:00 EDT</pubDate>
    <guid isPermaLink="false">http://www.nytimes.com/2006/04/01/us/01educ.html</guid>
  </item>
```

RSS Example

- The terms Read more | Open in browser do not appear in the RSS code
- One can assume that as part of the XSLT translation that would be associated with the RSS file, a <xsl:condition> or <xsl:if> element is being used to conduct the transform of the link into these hyperlinks
- The <link> element from the RSS file would then be encoded in the *Open in browser* link

RSS Example

- If the Open in browser option is clicked, the following page loads
- So in this case, the RSS is being used to essentially present an abstract to the reader
- The reader then follows the link and the content is pushed down as a html page

RSS Example

- Notice some of the elements used in RSS, including these;

```
<pubDate>Sat, 01 Apr 2006 00:00:00 EDT</pubDate>
<guid isPermaLink="false">http://www.nytimes.com/2006/04/01/us/01educ.html</guid>
```

- pubDate can be used in conjunction with XSLT conditional statements to decide when an item of content is no longer to be displayed
- Guid isPermaLink="false" indicates that at some time the link to the story will no longer be available
- Again, the XSLT code can use these values, particularly for removing 'old' content from the RSS feed
- Also, the thought may have already occurred to you that while sites hand out their RSS feed data, they do not typically hand out their XSLT documents
- In many cases it may not be XSLT working in the background with the RSS file, but given the work you have done thus far, you should be able to see how easy it would be to write an XSLT translator to interpret RSS files

Components of the RSS Document

- The main components of an RSS document are;
- <rss>
- <channel> root element
- <item>
- <title>
- <link>
- <description>
- </rss> closer

RSS Channel

- The RSS channel is essentially just the topic, such as Education as seen in the previous example
- The channel is considered the top level component of an RSS document (root element), and forms a one to many relationship with RSS <item> elements
- For example, an Education channel (topic) can have a number of associated news items
- A channel is describe by the contents of its associated <title>, <link> and <description> elements (these are required elements)

RSS Item

- RSS Items exist within RSS channels
- In the previous example, an RSS item was an article within an Education channel (topic)
- As with channels, an Item is describe by the contents of its associated <title>, <link> and <description> elements
- All <item> elements exist between the <channel> opening and closing root elements

RSS Example

- In this example from W3Schools.com, we can see a single channel which in turn contains a single item
- As you can see, it is simply a formalised way of using XML to represent content that is to be changed or updated
- It is a basic repeating node/element structure

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<rss version="2.0">

<channel>
<title>W3Schools</title>
<link>http://www.w3schools.com</link>
<description>W3Schools Web Tutorials </description>

<item>
<title>RSS Tutorial</title>
<link>http://www.w3schools.com/rss</link>
<description>Check out the RSS tutorial
on W3Schools.com</description>
</item>

</channel>
</rss>
```

http://www.w3schools.com/rss/rss_syntax.asp

RSS Optional Elements

- As RSS is considered a formalised, standard language of XML, it must have a number of elements which must be included if a document is to be correctly interpreted by a RSS aggregator
- While <title>, <link> and <description> elements are required for both the <channel> and <item> elements of an RSS document, the following are the 'optional' elements which provide more descriptive power when needed

Channel Optional Elements

Optional Elements in <channel>

Element	Description
<category>	Defines one or more categories the channel belongs to
<cloud>	Allows notification of updates.
<copyright>	Notifies about copyrighted material
<docs>	An URL to documentation on the RSS version the channel is using
<generator>	If the channel is created using an automatic generator, this is defined here
<image>	Inserts a picture to the channel.
<language>	Describes what language the channel uses. By using this tag it is possible for RSS aggregators to group sites based on language.
<lastBuildDate>	Defines the last date the channel was modified
<managingEditor>	Defines an e-mail address for the editor of the site
<pubDate>	Defines the last publication date for the channel
<rating>	Parental control rating of the page
<skipDays>	Defines days where it is unnecessary for RSS aggregators to update the feed
<skipHours>	Defines hours where it is unnecessary for RSS aggregators to update the feed
<textInput>	Creates a text input for the channel
<ttl>	(ttl = time to live) Defines how many minutes the channel can stay cached before refreshing
<webMaster>	Defines an e-mail address for the webmaster of the site

http://www.w3schools.com/rss/rss_channel.asp

Item Optional Elements

Optional Elements in <item>

Tag	Description
<author>	Defines the author of the item.
<category>	Places the item in one or more of the channel categories.
<comments>	An URL to a comment's page for the item.
<enclosure>	Describes a media object related to the item
<guid>	GUID = Globally Unique Identifier. Defines a unique identifier to the item.
<pubDate>	The publication date for the item.
<source>	Is used to define a third party source.

http://www.w3schools.com/rss/rss_item.asp

RSS and XSLT

- As stated, as with any XML source, RSS needs an external processor like XSLT to transform the data into something that end-users can appreciate
- The transformations can be relatively simple through to quite complex depending upon the number of optional elements added to the required <channel> and <item> elements
- Most of the conditional formatting would be used to determine if a particular channel was out of date or item was in a specific category
- While for end users an RSS reader can be used for this job, XSLT could be used to embed RSS content into a web page

RSS and XSLT

- The XSLT skills your have covered thus far the semester are more than adequate to cover most of the RSS elements
- Given the popularity of podcasts, blogs, forums, auctions and almost any other type of online service that can represented in RSS format, time spent looking at these two technologies should be time well spent
- Given that much of the RSS material usually describes and links to highly stylized content, it may be worth further enhancing your CSS skills in conjunction with XML and XSLT skills

RSS Aggregator

- Software
  - Client Based
  - Server Based
- Subscribe to sites by adding feeds
- Different aggregators that do basically the same thing
- Content is syndicated into one place
- Different sites and different contents, almost like a portal
- Aggregators essentially manage the adding of new feeds and the distribution of new feeds

Publishing an RSS Feed

Publish Your RSS Feed

Publishing and getting people to notice your RSS feed is as important as making one.

First, put your rss.xml file on the internet. You can place it on the same server as your site.

Next, let people know you have a RSS feed. Register your feed with an aggregator.

- Syndic8: The largest RSS directory. Syndicate has over 300,000 feeds listed. Register your feed here.
- Daypop: A large news oriented RSS aggregator. Over 50,000 news oriented feeds. Register your feed here.
- Newsisfree: A news oriented RSS aggregator. Over 18,000 feeds. Register your feed here.

Then, place this RSS and this XML with URL's to a page that explains briefly how other people can view your RSS feed.

http://www.w3schools.com/rss/rss_publishing.asp

RSS Where and When

- RSS feeds should be added to any site where content changes frequently or a ‘’news’ service exists
- For example, the SCIS student site has an RSS feed for student related news;
  http://www.scis.ecu.edu.au/rss/general.xml
- For those of you who will go on to work as web developers, you will be asked about RSS technology
- Your job will be to design the xml structure of the RSS file, place a link to that file in any number of RSS aggregators for wide distribution and perhaps even to develop an in-site RSS reader (using XSLT)

RSS Popularity

- RSS is popular due to its inherent simplicity for subscribers (using a reader)
- No spam
- For content producers, aggregators offer an easy way to distribute links to content
- A certain number of users will eventually end up at the web site of the content producer
- RSS content is highly ephemeral, here today, gone tomorrow, which keeps people coming back
- Thus RSS works at all levels, from in-house information distribution to world wide content subscription

RSS and DTD's

- Interestingly, though RSS is essentially an xml syntax, DTD's are not often used
- In RSS 2.0 (which you should be using) DTD's are not used
- In older versions, such as version .91 DTD's were required
- In one example, my.netscape.com changed its server and thousands of .91 RSS readers failed because they could not access the DTD which they required in order to operate;

  http://blog.netscape.com/2007/01/16/to-dtd-or-not-to-dtd/

- Modern readers that work with RSS 2.0 have no need of an RSS, but backwards compatibility might be an issue for future clients you develop for

Conclusion

- RSS is a technology that, as web developers, you will have to deal with, most likely sooner rather than later
- RSS allows for the integration of your xml, xslt, html and css skills