CSI 3140 – Laboratory 7

XML

As seen in class, an RSS document is an XML document that can be used to publish some short information about a blog, for example. The following links will provide more information and one example:

http://www.rssboard.org/rss-specification

http://www.rssboard.org/files/sample-rss-2.xml

http://www.rssboard.org/rss-profile

In this lab, we are going to create and update an RSS document, and produce an html page to display our current RSS document in a web browser. We focus on a subset of RSS, namely the three mandatory child elements of channel: description, link and title, and one optional element: item. A channel may contain any number of items (or no items at all).
For items, we focus on the description, link, pubDate, title and categories elements. We assume that items have exactly one description, pubDate and title child elements, zero or one link and any number of categories. See an example below.

Question 1: creating the RSS document

We first need to be able to populate the RSS document. We assume that we already have saved in a file on the server an existing RSS document that contains at least the channel elements and its child elements (except for items, maybe).
Create an xhtml form that can be used to enter all the relevant elements of an item: description, link, title and categories. Link and title are inputs, description is a textarea, and categories are selected from a choice of predefined categories (using checkboxes).
Create a Servlet that will process your form by reading the current RSS file, using a DOM parser, update the DOM document with the newly entered entry (remember, link might by empty, several categories can be selected, and pubDate should be added automatically by your servlet), and then saves the file back.

Question 2: displaying the RSS document using SAX

Using a SAX parser, create an xhtml file out of your RSS file, using the following formatting:

```
<html>
    <head>
        <title> My RSS Output </title>
    </head>
    <body>
    <h1>channel's title</h1>
    <h2>
        <a href="item's link" title="item's description">
```

```
                       item's title"
            </a>
      </h2>
      <div class="postDate">Posted: item's pubDate</div>
            <div class="itemDescription">
            item's description
            <div class="category">Categories:</div>
            <ul>
                  <li> item's category 1</li>
                  <li> item's category 2</li>
                  <li> item's category 3</li>
                  …
            </ul>
      </div>
      </body>
</html>
```

Where the items part is repeated for each item, the category part lists all of the current item's categories (and the entire <div class ="category"> … </div> is removed if the item has no categories), and if the link is missing the item's title becomes simply:

```
      <h2>
                  item's title
      </h2>
```

Create a servlets that uses SAX to parse the RSS file it finds on the server, and returns the hml above to be displayed in the client's browser.

Question 3: displaying the RSS document using XSL

Do the same thing but with an XSL transformation instead.

Sample RSS file :

```
<?xml version="1.0" encoding="UTF-8"?>
<rss version="2.0">
   <channel>
      <title>My Wonderful Syndicated Column</title>
      <description>Some stuff goes here.</description>
      <item>
         <title>My first post</title>
         <link>http://somewhere.example.com?p=1</link>
         <description>My oh my, we are on-line</description>
         <pubDate>Tue, 03 Jun 2003 09:39:21 GMT</pubDate>
         <category>not interesting</category>
         <category>not funny</category>
      </item>
      <item>
      <title>My second post</title>
         <link>http://somewhere.example.com?p=2</link>
          <description>Sky  watchers  in  Europe,  Asia,  and  parts  of
Alaska      and      Canada      will      experience      a      &lt;a
href="http://science.nasa.gov/headlines/y2003/30may_solareclipse.htm"&g
t;partial    eclipse    of    the    Sun&lt;/a&gt;    on    Saturday,    May
31st.</description>
         <pubDate>Fri, 30 May 2003 11:06:42 GMT</pubDate>
         <category>not funny</category>
      </item>
      <item>
         <title>My third post</title>
         <description>Whats up? nothing much, really</description>
         <pubDate>Tue, 27 May 2003 08:37:32 GMT</pubDate>
      </item>
      <item>
         <title>Am I running out of ideas already?</title>
         <link>http://somewhere.example.com?p=4</link>
         <description>Compared to earlier spacecraft, the International
Space Station has many luxuries, but laundry facilities are not one of
them.  Instead, astronauts have other options.</description>
         <pubDate>Tue, 20 May 2003 08:56:02 GMT</pubDate>
         <category>not interesting</category>
         <category>not funny</category>
         <category>not new</category>
         <category>not wanted</category>
      </item>
   </channel>
</rss>
```