

CSci 5131 Advanced Internet Programming (Spring 2012)**Instructor: Anand Tripathi****Department of Computer Science & Engineering****University of Minnesota Twin Cities****Assignment 5: Java Servlets, XML Parsing, and RSS Processing****Due Date: April 24, 2012****This assignment can be done in a group of up to 2 students****Objective:**

The goals of this assignment are to learn the following:

- Java Servlet Programming
- XML Document Parsing, and using this to parse [RSS](#) stream documents

Here is an example of the [RSS feed from CNN](#)

Problem Description:

An [RSS](#) feed is an XML, and it contains a list of news items, where each item contains two parts: title and description. (The underlined words are XML tags in the RSS documents.)

Following is an example of an RSS-XML document :-

```
<?xml version="1.0"?>
<rss version="2.0">
  <channel>
    <title>Liftoff News</title>
    <link>http://liftoff.msfc.nasa.gov/</link>
    <description>Liftoff to Space Exploration.</description>
    <language>en-us</language>
    <pubDate>Tue, 10 Jun 2003 04:00:00 GMT</pubDate>
    <lastBuildDate>Tue, 10 Jun 2003 09:41:01 GMT</lastBuildDate>
    <docs>http://blogs.law.harvard.edu/tech/rss</docs>
    <generator>Weblog Editor 2.0</generator>
    <managingEditor>editor@example.com</managingEditor>
    <webMaster>webmaster@example.com</webMaster>
    <item>
      <title>Star City</title>
      <link>http://liftoff.msfc.nasa.gov/news/2003/news-starcity.asp</link>
      <description>How do Americans get ready to work with Russians aboard the International Space Station?
      <pubDate>Tue, 03 Jun 2003 09:39:21 GMT</pubDate>
      <guid>http://liftoff.msfc.nasa.gov/2003/06/03.html#item573</guid>
    </item>
    <item>
      <description>Sky watchers in Europe, Asia, and parts of Alaska and Canada will experience a &lt;a href=
      <pubDate>Fri, 30 May 2003 11:06:42 GMT</pubDate>
      <guid>http://liftoff.msfc.nasa.gov/2003/05/30.html#item572</guid>
    </item>
    <item>
      <title>The Engine That Does More</title>
      <link>http://liftoff.msfc.nasa.gov/news/2003/news-VASIMR.asp</link>
      <description>Before man travels to Mars, NASA hopes to design new engines that will let us fly throug
      <pubDate>Tue, 27 May 2003 08:37:32 GMT</pubDate>
      <guid>http://liftoff.msfc.nasa.gov/2003/05/27.html#item571</guid>
    </item>
    <item>
      <title>Astronauts' Dirty Laundry</title>
      <link>http://liftoff.msfc.nasa.gov/news/2003/news-laundry.asp</link>
      <description>Compared to earlier spacecraft, the International Space Station has many luxuries, but 1
      <pubDate>Tue, 20 May 2003 08:56:02 GMT</pubDate>
      <guid>http://liftoff.msfc.nasa.gov/2003/05/20.html#item570</guid>
    </item>
  </channel>
</rss>
```

(This example is available at <http://cyber.law.harvard.edu/rss/examples/rss2sample.xml>)

In this assignment you will gather RSS streams from different new outlets and pick from them those items that contains a given set of keywords for which the user is interested in finding any news items.

In this assignment you would build a CGI based system that would obtain RSS feed data from different news websites, and present to the user filtered news items based on the search criteria specified by the user.

[Click here to see working \(implemented using Perl/CGI\).](#)

An XHTML document would present a form with the following elements:

- A multiple selection drop-down menu to select one or more news sites as source of RSS feed to be filtered.
 - Use at least the following four sites:
 - CNN http://rss.cnn.com/rss/cnn_world.rss
 - ABC <http://feeds.abcnews.com/abcnews/internationalheadlines>
 - NBC <http://rss.msnbc.msn.com/id/3032091/device/rss/rss.xml>
 - BBC http://newsrss.bbc.co.uk/rss/newsonline_uk_edition/world/rss.xml
- A textfield to get a list of keywords for searching. (keywords would be separated by spaces)
- A radio button group to indicate if the news items should be selected based on occurrence of ANY or ALL of the given keywords in the title.
- A submit button to submit the form for processing which will cause this form to be **submitted to a Java servlet running on your Tomcat server.**

A Java servlet running on your Tomcat server will perform the following actions:

- It will contact each of the selected RSS newsfeed sites and get from each of them an XML-RSS document.
- It will use a Java XML Parser to parse each of the received documents to perform the following actions:
 - Get all item elements and check the title part of each of them for the given keywords
 - Based on the option selected, if the tile contains ANY or ALL of the specified keywords (in any order), select the following parts of that news items to be sent to the client as XHTML
 - Title
 - Description
 - Link
- The servlet would send an XHTML document to the client containing the above parts for each selected news item.

[See this link for an example of a Java program for XML-RSS parsing.](#)