4.6 NEWSFEEDS and ATOM

NEWS FEED

News feeds are an example of automated syndication. News feed technologies allow information to be automatically provided and updated on Web sites, emailed to users, etc. As the name implies news feeds are normally used to provide news; however the technology can be used to syndicate a wide range of information**.**

The **BBC ticker** is an example of a news feed application. A major limitation with this approach is that the ticker can only be used with information provided by the BBC. The RSS (Really Simple Syndication)standard was developed as an open standard for news syndication, allowing applications to display news supplied by any RSS provider.

RSS (Really Simple Syndication)

RSS is an XML dialect used to publish frequently updated content, such as blog posts or news headlines.

It is a way to easily distribute a list of headlines, update notices, and sometimes content to a wide number of people. It is used by computer programs that organize those headlines and notices for easy reading.

The content feed is identified by a unique URI, and this URI is used by an RSS Reader application to retrieve and display the content feed from a web site. An RSS feed can contain one or more images or items. An item can be a synopsis of an article with a link to the full article or the entire article itself. An RSS has a well-defined structure. Some Web APIs use RSS as the data format returned when a request is made.

Working of RSS

RSS works by having the website author maintain a list of notifications on their website in a standard way. This list of notifications is called an **RSS Feed**. People who are interested in finding out the latest headlines or changes can check this list. Special computer programs called **RSS aggregators** have been developed that automatically access the RSS feeds of websites of user's interest and organize the results.

Producing an RSS feed is very simple and hundreds of thousands of websites now provide this feature, including major news organizations like the New York Times, the BBC, and Reuters, as well as many weblogs.

RSS provides very basic information to do its notification. It is made up of a list of items presented in order from newest to oldest. Each item usually consists of a simple title describing the item along with a more complete description and a link to a web page with the actual information being described. Sometimes this description is the full information the user want to read (such as the content of a weblog post) and sometimes it is just a summary.

Creating RSS feed

The special XML-format file that makes up an RSS feed is usually created in one of a variety of ways. Most large news websites and most weblogs are maintained using special **content management** programs. Authors add their stories and postings to the website by interacting with those programs and then use the program's publish facility to create the HTML files that make up the website. Those programs often also can update the RSS feed XML file at the same time, adding an item referring to the new story or post, and removing less recent items.

Blog creation tools like Blogger, LiveJournal, Movable Type, and Radio automatically create feeds. Websites that are produced in a more custom manner, such as with Macromedia Dreamweaver or a simple text editor, usually do not automatically create RSS feeds. Authors of such websites either maintain the XML files by hand, just as they do the website itself, or use a tool such as Software Garden, Inc.'s ListGarden program to maintain it.

There are also services that periodically read requested websites themselves and try to automatically determine changes (this is most reliable for websites with a somewhat regular news-like format), or that let the users create RSS feed XML files that are hosted by that service provider.

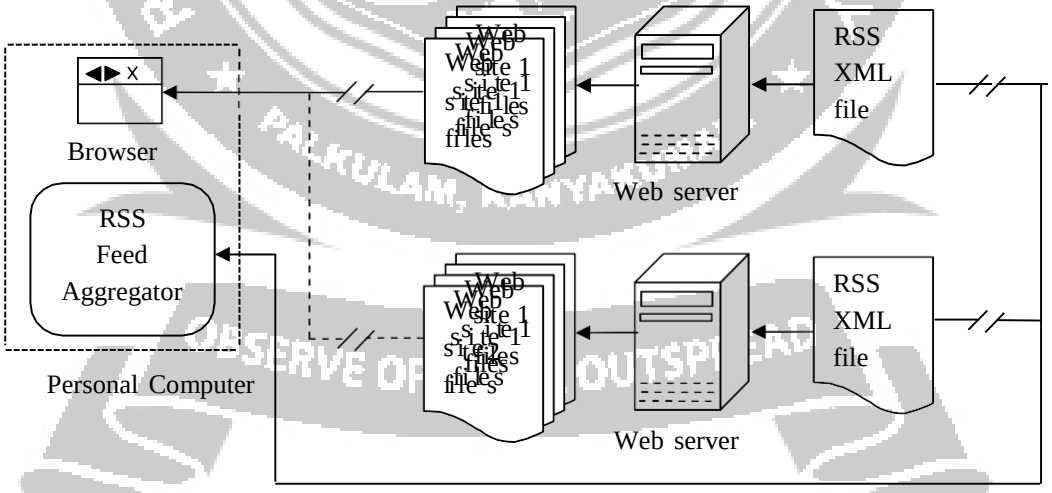

Fig 4.3: Communication between websites, RSS feed and PC

In the above diagram, a web browser being used to read first Web Site 1 over the Internet and then Web Site 2. It also shows the RSS feed XML files for both websites being monitored simultaneously by an RSS Feed Aggregator.

- **Sections of an RSS file**

Apart from the root element there are four main sections of the RSS file. These are the channel, image, item, and text input sections. In practical use, the channel and item elements are requirements for any useful RSS file, while the image and text input are optional.

The channel section

The channel element contains metadata that describe the channel itself, telling what the channel is and who created it. The channel is a required element that includes the name of the channel, its description, its language, and a URL. The URL is normally used to point to the channel's source of information.

Channel element

```
<channel><title>MozillaZine</title>
<link>http://www.mozillazine.org</link>
<description>The user source for Mozilla news, advocacy, interviews, builds, and more!
</description>
<language>en-us</language>  </channel>
```

The title can be treated as a headline link with the description following. The **Channel Language definition** allow aggregators to filter news feeds and gives the rendering software the information necessary to display the language properly. The </channnel> tag is used after all the channel elements to close the channel. As RSS conforms to XML specs, the element must be well formed; it requires the closing tag.

The image section

The image element is an optional element that is usually used to include the logo of the channel provider. The default size for the image is 88 pixels wide by 31 pixels high, but it can be enlarged to 144 pixels wide by 400 pixels wide.

Image element

```
<image><title>MozillaZine</title>
<url>http://www.mozillazine.org/image/mynetscape88.gif</url>
<link>http://www.mozillazine.org</link>
<width>88</width>
<height>31</height>  </image>
```

The image's title, URL, link, width, and height tags allow renderers to translate the file into HTML. The title tag is normally used for the image's ALT text.

The items

Items form the dynamic part of an RSS file. While channel, image, and text input elements create the channel's identity and typically stay the same over long periods of time, channel items are rendered as news headlines, and the channel's value depends on their changing fairly frequently.

Item element

```
<item><title>Java2 in Navigator 5?</title>
<link>http://www.mozillazine.org/talkback.html?article=607</link>
<description>Will Java2 be an integrated part of Navigator 5?
    Read more about it in this discussion...</description> </item>
```

Fifteen items are allowed in a channel. Titles should be less than 100 characters, while descriptions should be under 500 characters. The item title is normally rendered as a headline that links to the full article whose URL is provided by the item link. The item description is commonly used for either a summary of the article's content or for commentary on the article.

News feed channels use the description to highlight the content of news articles, usually on the channel owner's site, and Web log channels use the description to provide commentary on a variety of content, often on third-party sites.

The text input

The text input area is an optional element, with only one allowed per channel. This lets the user respond to the channel.

```
<textinput><title>Send</title>
<description>Comments about MozillaZine?</description>
<name>responseText</name>
<link>http://www.mozillazine.org/cgi-bin/sampleonly.cgi</link>    </textinput>
```

The title is normally rendered as the label of the form's ssubmit button, and the description as the text displayed before or above the input field. The text input name is supplied along with the contents of the text field when the submit button is clicked.

ATOM

Atom is a syndication data format like RSS, as well as a publishing protocol. The Atom data format uses XML syntax with one or more entry elements containing the full and/or summary content.

The **Atom Publishing Protocol** (APP) defines a hierarchy for organizing published content (services, workspaces, collections, and resources) and uses the HTTP Get, Post, Put, and Delete methods for retrieving, creating, deleting, and editing published content. Atom's use of HTTP's built-in methods and XML as a data format is in the spirit of a RESTful web services implementation.

Atom was designed to be a universal publishing standard for blogs and other Web sites where content is updated frequently. Users visiting a Web site with an Atom feed can discover a file described as "atom.xml" in the URL that can be copied and pasted into an aggregator to subscribe to the feed.

Atom was originally developed as an alternative to RSS 2.0, the standard developed by Dave Winer and copyrighted by Harvard University, as a means of improving perceived shortcomings of the RSS format by the blogging community.

Features of ATOM

- Atom was developed to be vendor neutral and freely extensible by any user; it is an open standard.
- Atom lies within an XML-namespace.
- Atom includes auto discovery, allowing feed aggregators to automatically detect the presence of a feed.
- Atom differentiates between relative and non-relative URIs.
- Atom has separate summary and content elements.
- Atom explicitly labels a payload as plaintext or HTML.
- Each entry has a globally unique ID.