What is RSS and how can it serve libraries?

Zeki Çelikbaş

Istanbul Technical University
Faculty of Aeronautics and Astronautics, Istanbul, Turkey
celikbas@itu.edu.tr

Abstract: Examines RDF Site Summary, or Rich Site Summary (RSS) feeds, a means of distributing news or other Web content from an online publisher to users. Outlines RSS background and XML-based structure. Explains feed readers and web-based content aggregators, the two chief tools used for reading RSS feeds. Lists some advantages of RSS use and some reputable RSS feed sources, also known as RSS feed directories. Offers suggestions on potential library uses of RSS including: general communications and marketing; library user education (via book lists); current awareness services (via table of contents analysis for journals); and, augmenting of reference services.

Introduction

Think for a moment about why and for what you use the Internet. The answers are probably many, but in a nutshell they can be reduced to: for almost everything and anything, and because you need up-to-date information quickly.

The Internet is indeed a powerful tool for our everyday lives. But the Internet has a different meaning for us as information professionals, and that is as a part of our daily work. Internet-connected computers are indispensable for our official, everyday chores, but, at the same time, support for Internet services is becoming a weightier duty among our present information service responsibilities. Further improvements in network and Internet technologies will only increase the big changes in our lives. As an information source, the Internet contains volumes of information, from personal homepages to portals, search engines to directories, and we are responsible for successful management of this vast and chaotic information source.

People may add web site links to their bookmarks according to their own personal interests and have to visit their sites again and again to check for updates. Each time, they have to reload the pages and check various different links at each site for any potential updates, all the while remembering the previous content and page styles so that they might notice all the changes possibly there.

This may be workable for many individual users, but whether one approaches this from a computer science or library science perspective, for information professionals, especially those managing large amounts of continuously changing, increasing and/or updating information, this would quickly become a nightmare. With hundreds of sites, consider the hours it would take to update each link manually, surfing to each site, recoding pages, and so forth. But RSS can help!
With its scheduled, periodic ‘feed’ broadcasts offering instant updates and allowing users to target the latest updated information, RSS helps both simplify and speed the access we want to provide to the information our users want.

What is RSS?

Simply put, RSS is an XML-based format to share web site content. An excellent definition of it is provided by Whatis.com (searchWebServices.com) [1]:

RSS (RDF Site Summary – formerly called Rich Site Summary) is a method of describing news or other Web content that is available for “feeding” (distribution or syndication) from an online publisher to Web users. RSS is an application of the Extensible Markup Language (XML) that adheres to the World Wide Web Consortium's Resource Description Framework (RDF). Originally developed by Netscape for its browser's Netcenter channels, the RSS specification is now available for anyone to use.

A Web site that wants to "publish" some of its content, such as news headlines or stories, creates a description of the content and specifically where the content is on its site in the form of an RSS document. The publishing site then registers its RSS document with one of several existing directories of RSS publishers. A user with a Web browser or a special program that can read RSS-distributed content can read periodically-provided distributions. Some current directories of RSS files include Meerkat, GropSoup, NewsIsFree, UserLand, and XML Tree; these sites are sometimes known as content aggregators. RSS browsers include Headline Viewer and Novobot.

News is only one form of content that can be distributed with an RSS feed. Other possibilities include discussion forum excerpts, software announcements, and any form of content retrievable with a URL.

So, if a web site wants to share and publish some of its content to other sites at the same time, the publisher can create an RSS document. This document is published in the web site and any user can read it and use different sites at the same time. This is the syndication element referred to in the acronym: information is being shared with other sites at the same time.

The evolution of the technology has entailed some confusion, specifically in what the acronym RSS stands for (and even more confusion, for novices, as to what it actually means). Communications media pundit Amy Gahran [2] is therefore advocating use of the term webfeed as more comprehensive and comprehensible. This term would be equivalent to RSS but would also include other feed types (such as Atom feed), in a manner that we could define as “news on your desktop”. For the purposes of the present paper, however, we are leaving aside the possible future adoption of webfeed and will refer to RSS feeds throughout.
The Background of RSS: A Short History

RSS comes from Netscape’s innovative MyNetscape portal service and Netcenter channels. Through the portal services, Netscape allowed anyone who had a MyNetscape account to publish news headlines on his or her pages. Thus was born the first version of RSS (0.9), introduced by Netscape in 1999. Not long after, Netscape decided to get out of the portal business. At this point, another firm, UserLand, that developed weblogging products, picked up RSS 0.91 and continued to develop it. Weblogging or blogging is a web-based authoring system and a weblog or blog as it is most commonly known, is a web page that serves as a publicly available personal journal for an individual. UserLand published RSS versions 0.92, 0.93 and 0.94 with this system. Meanwhile, a non-commercial group, the RSS-DEV Working Group (<http://purl.org/rss/1.0/spec>), started to work on RSS and developed version 1.0, basing it on W3 Corporation’s RDF standards and redefining the acronym RSS as “RDF Site Summary”. UserLand, however, was not satisfied with RSS 1.0 and continued development of their version of RSS, eventually releasing RSS 2.0. Development continues unabated [3]. In Figure 1, one can see the proportional increase in the total number of RSS feeds from September 2001 to May 2004 as reported by Syndic8, a large and important RSS directory on the Internet.

![Figure 1: Increase proportion of RSS between years 2001-2004](source: Syndic8 statistics page <http://www.syndic8.com/stats_feedcount>bar.php> accessed ca. May, 2004)

RSS Feed Structure

Content Syndication with RSS by Ben Hammersley is an excellent starting point for further exploration of (all aspects of) RSS. The book covers the main standards and
specifications, details of RSS modules which can be used to extend the format, and some fairly in-depth explanations of how RSS can be created using tools such as the XML::RSS module.

Before creating an RSS feed document, one must decide which version of RSS will be used. Currently, all RSS versions developed to date are still in use on the Internet. It is important to note that RSS 2.0 is not a continuous development of/from RSS 1.0; the different versions have been developed by different groups. Nevertheless, all RSS versions are XML-based and generally have the same structure. Independent of version then, the basic RSS structure entails the following: [4]

1. XML-based, it must be formatted like all XML documents.
2. The first element in the document is the <rss> element. This includes a mandatory version attribute.
3. The next element is the <channel> element. It may include up to 15 items.
4. The <title> element gives the title, either of the entire site or, if it is within an <item> element, the title of a specific item.
5. The <link> element indicates the URL of the web pages that corresponds to the RSS feed or if it is within an <item> the URL to that specific item.
6. The <description> element describes the RSS feed or the item.
7. The <item> element is the meat of the feed. These are all the headlines (<title>), URLs (<link>) and descriptions (predictably, <description>) that will be in your feed.

Using the basic elements then, one could have a basic RSS feed, perhaps something like this example: [5]

```xml
<?xml version="1.0"?>
<rss version="0.91">
  <channel>
    <title>Meerkat: An Open Wire Service</title>
    <link>http://meerkat.oreillynet.com/</link>
    <description>Meerkat is a Web-based syndicated content reader</description>
    <language>en-us</language>
    <image>
      <title>Meerkat Powered!</title>
      <url>http://meerkat.oreillynet.com/icons/meerkat-powered.jpg</url>
      <link>http://meerkat.oreillynet.com/</link>
    </image>
    <item>
      <title>The First Item</title>
      <link>http://www.oreilly.com/example/001.html</link>
      <description>This is the first item.</description>
    </item>
    <item>
      <title>The Second Item</title>
```

This RSS feed example would appear online as shown in the Figure 2 screenshot:

![Figure 2: Screenshot of an RSS 0.91 feed transformed into HTML](image)

**Working with RSS: Reading and Writing**

As RSS synchronous updates between web pages can be read with browsers, RSS documents can also be read by specialized desktop software programs called **feed readers** or **RSS aggregators**. RSS readers are one of the biggest new categories of software in recent memory. Here is a list of the categories (with specific examples) of the new tools available that enable reading of RSS feeds:

- Desktop feed readers (examples: Feedreader; Amphedadask; FeedDemon)
- Web-based readers (examples: Bloglines; Oddpost)
- Plug-ins for web browsers and e-mail clients (examples: NewsGator, which adds feed support to Microsoft Outlook; Mozilla Firefox browser’s RSS reader panel)
- Browsers with built-in feed readers (example: Opera 7.50)

There are even web services that deliver subscribed RSS feeds to an e-mail address. All of these kinds of tools can be seen in different lists. For example, RSSfeeds.com offers an alphabetical list of readers. *PCWorld* gives this same kind of list, but with the addition of ratings.

If a web site has an RSS feed, you will see a little blue or orange icon, bearing the label XML or RSS on it, as in these examples (Figure 3):
When you click this icon, your web browser will download the RSS document page and display it to screen.

However, what you will probably see at this point is a rather ugly page filled with code: that is how a feed looks when viewed through most traditional web browsers. This code must be processed to come out in a meaningful and structured, i.e. a readable, format. This process is called parsing. To do this, all you have to do is copy or enter the address (URL) of the page in question into computer memory and tell your feed reader that you wish to add it as a new feed. It will ask you for the address of the feed; simply paste in the URL.

If you are connected to the Internet with a fixed position, you will find desktop feed readers are very easy to work with. In general, these readers have three panels (boxes, frames) and are layed-out like most e-mail clients, which are doubtless familiar to most computer users (e.g. see Figure 4 below). One of the panels shows a list of the feeds it recognizes (also sometimes known as channels). Next is the box for headlines (similar to the subject/date received frame of an e-mail client). The last, our ‘browser box’, if you will, is used for content display (like the main frame of a divided page seen via a web browser, or again, like the letter frame in an e-mail client, it displays the actual content of the selected feed item). Clicking on any of the headlines will bring up the article – or more commonly still, a description of the article and a hyperlink to it – in the browser box. When you click that link you can continue to read the article, from its original site of web publication, within the ‘browser box’ panel.

For the purposes of this paper, both a desktop feed reader, FeedReader, and a web-based aggregator, Bloglines, have been tested.

**FeedReader** (<http://www.feedreader.com/>) is a GPL (GNU Public License) licensed free software, which means that it is free to use anywhere, under the conditions specified in the license. Below is a screenshot of the FeedReader interface (Figure 4). Like the generic feed reader described above, its display is divided into 3 panels or boxes. The feeds available are listed on the left side. Headlines are in the upper right panel and the content of the feed can be seen in the bottom right panel. To add a new feed, one clicks the New button (upper left corner) or presses F2. This opens a window that asks the URL of the RSS feed. Simply paste the URL (previously copied the clipboard or computer memory) or enter it directly into this box and then click Next. FeedReader will then automatically verify the feed and download the latest news. FeedReader will update a feed at any interval the user sets (e.g. every 30 minutes, every hour, etc.). When new content from a feed becomes available, FeedReader will open a box on the screen and alert the user with a beep. FeedReader will also keep a certain number of feeds in its memory for future use.
If you are working on computers from different locations, you will likely find it easier to follow your RSS feeds if you employ a web-based aggregator. Use of a web-based aggregator, like that available via Bloglines (<http://www.bloglines.com/>), gives one the familiarity and functionality of a typical desktop feed reader with the added ‘mobility’ bonus of access to one’s RSS feeds via any computer with an Internet connection.

You must register to start using Bloglines, but the process is as easy as registering for a web-based e-mail service. After you get into the system, to add a new feed, click the Add link in the upper left corner of the screen. This will open a page with an input
text box in the right frame; again simply enter or paste the desired URL into the box, click Subscribe, and Bloglines will start your subscription immediately.

Advantages of RSS

For users, there are several advantages in subscribing to RSS feeds, a few of which are:

1. **PRIVACY**: One does not need to send one’s e-mail address anywhere, to anybody, in order to subscribe to any RSS feed.

2. **SPAM PROTECTION**: Spam-like publishing is impossible with RSS feeds. If a feed becomes annoying, one can cancel the subscription with only a few clicks of the mouse.

3. **EASY CANCELLATION**: One does not need to send an ‘unsub’ e-mail message as with most e-mail listservs or go through a complicated process on any web site. To cancel a subscription, simply delete (or ‘drop’) the feed from the list of feeds followed.

4. **CURRENT CONTENT MANAGEMENT**: RSS feeds identify the newest updates and additions, and one can set feed readers and content aggregators to give an alert when new content has arrived (in essence, providing an instant ‘current awareness’ service for users).

Finding RSS Feeds

The web sites you already visit may contain an RSS feed. To **find the RSS feed, look for one of the small icons we have mentioned above (Figure 3: Different RSS icons)**; most RSS feeds are linked to an icon. If you still have not found an RSS feed of interest, there are several directories available to help those who want to discover what RSS feeds might offer them. [6] The following list provides a quick introduction to some reliable and reputable RSS feed directories:

**RSSfeeds.com (<http://www.rssfeeds.com>)** has put together a very nice structure for listing RSS feeds. Based on the Open Directory Project structure, it allows one to drill down through familiar-looking categories, like ones from known search engines such as Google. The search listings themselves may not seem much (name, description, date indexed, and URL) but when one clicks on the “More Details” link for the feed, one finds a more impressive set of information. “More Details” for an individual feed includes: a screenshot of the site; links to validate the RSS’s four different validators (all RSS feeds must be well-formed XML; validators check RSS feed source against the rules defined in the version information; it also validates elements of commonly used namespaces); where it is in the hierarchical listing of the RSS feed directory; and, a link to add the feed to your feed reader
Syndic8 (<http://www.syndic8.com/>) represents the largest directory of RSS channels, with over 75,000 feeds listed. Syndic8 provides search and browse facilities that allow one to find channels in particular areas readily. Syndic8 also offers a very detailed filter option on searches. Users can also browse feeds via subject headings according to various different systems of subject headings, the most known/familiar of which for most users being those of the Open Directory Project.

NewsIsFree (<http://www.newsisfree.com/>) is one of the oldest and most established directories of RSS feeds. Users of this site can browse by subject heading, search the extensive database by name or description, or search the latest headlines from its indexed sites. One of the favourite sections of NewsIsFree is the “Latest Channels” portion, which is updated daily and provides quick access to the newest feeds added to the database. NewsIsFree can also be used as a web-based news aggregator for subscribing to feeds that are listed in the site.

— and last, but certainly not least, especially for the library-oriented reader —

LISFeeds.com (<http://www.lisfeeds.com>), which is, in fact, not merely a directory, but also a web-based RSS headline aggregator. It gathers headlines from sites and services that specialized in library-oriented news and presents these in a convenient, one-stop shop for library-oriented headline seekers.

**RSS and Weblogs: Blogging the Net**

While we are dealing with RSS, we would be remiss not to mention weblogs, commonly known as blogs, since RSS feeds have become especially widespread with the advent of and increase in weblogging.

A weblog or blog is a web site that contains brief entries, arranged in reverse chronological order and updated regularly. Blogs are usually created by just one author, but are sometimes compiled collaboratively by a community of authors. Updating a weblog is as easy as using a word-processor, and blogs can be updated from several times per day to just a few times per week or month. Some blogs encourage interactivity between the writer and audience by allowing readers to post comments and questions about entries.

RSS feeds are the most important component of weblog systems: if someone adds a new entry to her or his blog, its RSS feed is updated immediately to inform all subscribers.

Their flexibility and ease of use have made weblogs the Web’s most important element. Whether it is a question of —

· e-learning projects, helping instructors and students to discuss and share information on a subject using web media (e.g. teachers could post daily news about the course, answers to FAQs, and links to extra materials);
· intra-organizational communication, allowing members to share knowledge, executive officers to disseminate information, and managers to supervise projects;

· formation of a departmental or project resource, assisting different kinds of groups in efforts to build teams and allow team members to share resources, experience, and knowledge about the project or work of the department or group; or,

· marketing tools, affording businesses and other organisations eager to publish information about their activities to the outside world, reliable and facilitating means to do so

— weblogs are to be found.

You do not need to have very special computer skills and knowledge (html, xml, programming) to create and maintain a lively weblog. Online weblog services such as Blogger (<http://www.blogger.com>) and LiveJournal (<http://www.livejournal.com>) provide all that is needed to create a very professional look with a modicum of effort.

New technologies and the conditions they create force libraries and library workers to seek new ways to inform users and potential users, ways beyond the classic library newsletters, brochures and the like. Some libraries are already using weblogs for these purposes (for one lengthy list, see <http://www.libdex.com/weblogs.html>) A few ways libraries can use weblogs immediately suggest themselves: [7][8].

- To update librarians’ current knowledge base;
- To inform users of new resources (as one aspect of library user education);
- To inform users about library services (as another aspect of the same).

RSS in Libraries

The potential of RSS feeds for the library environment is at least as great as its multiplicity of uses in other contexts presently suggest. Take the case of library publications.

Libraries publish announcements for their target audiences much as other web publishers might – notices, activity bulletins and the like – but libraries also want to present information on local and online (library) resources. While service announcements can be made by e-mail (subscriptions or listservs) or on a library web page, we can use RSS feeds for such service announcements and, at the same time, offer live hyperlinks to online resources. It is not necessary to give up the old and familiar ways used until now; nor is it even necessary to make significant changes to the core of library web sites. All that is required is that a little space be allotted for the little RSS file and the addition of a little RSS icon (see above, Figure 3) to the web page. Users utilizing RSS feeds will be very happy to be able to read their library-related news at/with their aggregators.

And this simple example only touches on the great potential of RSS feeds that are beginning now to be tapped.

Librarians at the National Cancer Institute (NCI) of the National Institutes of Health (NIH) in Bethesda, MD is creating a very special database including RSS
feeds gathered from the Internet: they have named it LION (Library Online). They have employed RSS both to integrate Internet content into the NCI library system and to distribute content from the Institute’s library system to their users via an RSS news feed available on their Intranet. This system connects RSS feeds from a variety of sources, among them: BBC News: Health; Moreover: Breast Cancer News; Moreover: Cancer News; New York Times: Health; and, Reuters Health eLine. LION’s tools are parsing the RSS feeds into its memory along with other item records. In LION, these RSS feed records can then be searched, browsed, displayed and saved in a user’s personal area. Another exciting aspect of this project is the chance to use the RSS technology bi-directionally: in addition to bringing content into LION, it also publishes, or syndicates, content from LION in RSS format, carrying it out to their registered NCI users.

The appeal of this versatile service is clear. According to the NCI staff, after the initial investment of time and technology, the information flows without requiring any additional, hands-on staff effort.

Moreover, there is similar software available that can provide these same kinds of services to any automated libraries. Examples include: ROSA (an acronym for the) RSS Open Source Aggregator and Filter (see <http://www.jisc.ac.uk/index.cfm?name=project_rosa>); and, Urchin (http://urchin.sourceforge.net/), an exciting and new Open Source tool that can be used by library webmasters to aggregate RSS content of their choice on/or their own web sites. Growing numbers of different types of RSS feeds are being produced by publishers and other content providers, and the opportunity is there for libraries not only to reduce information overload for their staff and patrons, but also to reclaim firmly the traditional role of information storage with new, easily manageable and searchable databases (over which they exercise collection development control).

Only a little imagination is needed to see how RSS aggregated feed could improve the average library site and the services the library offers. Any list-oriented content type is suitable for RSS library projects, as the following four options outline. [10]

I. RSS Marketing Ideas: Activities, Event Announcements, and More ...

As has been noted above, RSS feeds can be use to publish any announcement from the library web site on activities, exhibitions, promotions and new library resources, especially databases.

Weblogs provide more interactive solutions to these kinds of services. Librarians could introduce a new database with a short blog note and users could share their experiences, creating an interchange of questions and answers about the database in question which RSS feeds would also be able to inform subscribers of at their scheduled update times.

Libraries establish consortia to achieve better value in making large purchases. Similarly, library consortia could enable member libraries to work together to provide more detailed information – in quantity (volume) and in quality (depth) – about their ‘products’ and services to RSS aggregators. And, of course, library webmasters could aggregate this information to their respective library web sites for the benefit of their more immediate or local user group.
II. RSS Book Lists

Perhaps the most practical potential usage of RSS feeds in library settings is in the easy generation of various kinds of book lists.

As patrons are always curious to know about what’s new, a “New Items” list could be attempted. The Koha ILS, an Open Source system developed by Katipo in New Zealand, has started a project for RSS feeds – with newest items, most reserved and longest unseen books featured feeds – although the progress made to date is unclear (see <http://www.kohalabs.com/projects/rssKoha/>).

For large public or academic libraries, it would be impossible, of course: the hundreds or even thousands of books added every month would far exceed RSS limits on maximum items listable. But for smaller libraries – and perhaps for each branch, or section, or collection, of larger libraries – an RSS feed showing new items available would be feasible. Another approach to make more user-friendly, readable lists would be to divide and conquer: use 10, 20, 100 or however many, pre-determined subject headings and have the system publish the RSS feeds on new items by subject. The world’s most well-known online bookseller is trying something similar. Amazon is already involved in the production of a number of RSS feeds which aggregate book news under popular headlines (<http://www.amazon.com/exec/obidos/subst/xs/syndicate.html/>).

Another list (or set of lists) could be created for the library’s “Most Wanted” (books, videos, CDs, CD-ROMs, DVDs, etc.), punning on the classic crime posters. This list could even be named “Most Reserved Items” — this time punning on the stereotype of the dignified and refined library setting — because new patrons would likely reach these items only through use of reservation services. That would serve to hype what the library has in terms of product (the desired items that everyone wants) and services (in this case, patron’s ability to reserve desiderata). This kind of list could be produced and published using data from the library’s integrated library system (ILS) at daily, weekly, biweekly, monthly or other desired intervals.

III. RSS Table of Contents (TOC) for Journals (Current Awareness Service)

Current awareness services such as Table of Contents (TOC) compilation were traditionally manual, menial and mainly slow. Web services have sped up the process and reduced library staff labour, improving efficiency. RSS feeds have taken this further, as aggregated Table of Contents (TOC) for academic journals are now becoming available. It is possible at present, for example, to receive content from BioMed Central (<http://www.biomedcentral.com/info/about/rss/>), and The Scientist (<http://www.the-scientist.com/>), as RSS headline feeds. These contain even more detailed content, including abstracts of the articles contained in each journal issue. Also, it is now possible to subscribe to the RSS feed for Nature (<http://nurture.nature.com/rss/nature_test.rdf/>).

Early on, publishers became aware of the potential of RSS feeds and started to work on creating the Publishing Requirements for Industry Standard Metadata or PRISM (<http://prismstandard.org/>). A TOC module has been developed from this project to publish TOC information from journals online [11][12].
IV. RSS Feeds To Improve Reference Service

The importance of search engines could not be underestimated, especially for reference librarians and their potential clients. XmlHub (<http://www.xmlhub.com/odp_feed.php>), an Open Directory custom RSS-feed generator, recently unveiled two very interesting tools, destined to have an impact.

The first generates a customized RSS feed of Open Directory (<http://dmoz.org>) search results for any search term entered. The second generates a customized RSS feed for any Open Directory category. These tools make it possibly to alert patrons to new resources being added to Open Directory in the areas of their particular interests.

This same methodology could be used to generate customized RSS feeds for other databases, including libraries’ OPACs, bibliographic and full-text databases, and web directories.

Conclusion

New technologies are creating new opportunities for libraries. CD-ROM databases — which not so long ago were the latest thing — have now transitioned into growing online databases. Even as the Internet creates new opportunities for the library, it may also become a rival to library systems. According to Searcher editor Barbara Quint, Google now gets more searches in 3 days than all libraries combined globally get in 1 year [13]. Libraries must follow new technologies very closely and the advantages of the newest innovations, such as RSS feeds, must be reflected in library services. If this is not done voluntarily, this integration will be forced over time.

References

   Online, v. 27 no. 6, November/December (2003) Also available online via
   http://www.infotoday.com/online/nov03/plosker.shtml