CONTENT SYNDICATION: A NEW SOLUTION TO THE OLD PROBLEM

N.S. Harinarayana  Mallinath Kumbar  M V Sunil

Abstract

The Paper discusses about content syndication and its application in Libraries not only the storage but also the retrieval problems. Needless to say that computer technology overshadowed all the previous technologies.

Keywords: RSS, Content Syndication, Weblog, Alert Service

1. Introduction

Information explosion is a problem inherited from the twentieth century. The information professionals of the yester century were engaged in finding out ways and means of managing mostly the printed materials. Keeping themselves abreast with the latest information has been a nightmare, major among them being aspects like identification, storage and dissemination of information. Librarians after a long struggle came up with some kind of solution through their current awareness (CAS) and selective dissemination of information (SDI) services, which were available to a segment of users. Last quarter of the century provided them a sigh of relief as librarians could experiment with suitable technologies to deal with other kinds of problems which were also the offshoots of information explosion. During seventies microfilming technology was perceived to have solved one aspect of the problem i.e., the physical space problems. CD-ROM as an alternate technology proved too much for the microfilming technology to sustain itself in the market. The emergence of computer technology provided an opportunity to explore the solutions for not only the storage but also the retrieval problems. Needless to say that computer technology overshadowed all the previous technologies.

Computer technology in conjunction with communication technology enforced a paradigm shift in the way people generate, communicate, retrieve and use information. The access to information became easier than ever before. Technology substantially democratized the generation of information. Anybody from any part of the world at any point of time can generate and publish information instantaneously. Giant search engines such as Google, AltaVista etc extended the retrieval capabilities to everyone without any cost from the users of information. All these things sound good as long as we prefer to ignore the other face of the coin.

Technology, to some extent, has been responsible for aggravating the problem of information explosion. It has allowed information to grow without any control mechanism. Internet is a classic example for all the chaos. Googling for a piece of information in the Internet throws at us mountains of information from which we have to find what we want. This makes the system very ineffective and scary from the point of view of users. More over information searching and retrieval are always a user-initiated process. In other words, information user has to deliberately initiate the process of searching for information. Failing to do so from the part of user would leave him ignorant of the availability of information. In comparison to previous century, in a way we are “back to square one” situation as far as keeping oneself abreast of new information. For example, the user has to make a deliberate effort to find out the new developments in databases, search engines, new sites, changes in sites, news, etc.
A possible solution to this ever growing problem lies within technology itself, thanks to the two fastest growing technologies on the web, RSS and Atom. These are basically syndication technologies which allow websites to share information with other applications in a standard way by providing what is called "Content syndication feed". Syndication services help the user in being informed of latest developments in their fields of interest.

2. Syndication

Syndication means "sharing". A few web definitions are:

- The process by which a web site is able to share information, such as articles, with other web sites.
  http://www.sideroad.com/Article_Writing/article-marketing-terms.html

- Web syndication is a form of syndication in which a section of a website is made available for other sites to use. This could be simply by licensing the content so other people can use it, but more commonly these days web syndication refers to making Web feeds available from a site so other people can display an updating list of content from it (for example one's latest forum postings, etc.). This originated with news and blog sites but is increasingly used to syndicate any information.
  http://en.wikipedia.org/wiki/Newsfeed_syndication

- Syndication means the distribution a news article through a syndicate - in this case an RSS feed - for publication in a number of newspapers or periodicals simultaneously. Used in the context of RSS because RSS syndication is all about distributing content for reuse or redistribution on other websites.
  http://www.rstoolchest.com/rss-glossary.html

- The supplying of information or content for simultaneous publication in several periodicals or web sites.
  http://commnet.gc.ca/news_nouvelles/glossary_e.html

- Either offering your own content for use on other sites in exchange for a plug or link or using content from other sites with permission. This is a good strategy and effective quid pro quo for both parties. One receives free publicity and the other receives unique and quality articles for keyword optimization without having to write each proprietarily. Offering your articles on other web sites is a great way to build incoming links.
  http://www.proshay.com/glossary_c.html

Content syndication is a blanket term used to refer to accessing and publishing web content (text, images, etc.) in one or both of these formats: RSS and Atom.

3. What is RSS?

Originated by UserLand in 1997 and subsequently used by Netscape to fill channels for Netcenter, RSS has evolved into a popular means of sharing content between sites. RSS means different things to different people. Even they are expanded differently. "Really Simple Syndication", "Rich Site Summary" and "RDF Site Summary" denote one and the same RSS concept. According to Dave Winer one of its creators, "There is no consensus on what RSS stands for, so it's not an acronym, it's a name." "Really
Simple Syndication is probably the most accepted/used form. An orange icon with XML (XML) or RSS (RSS) in any website indicates that the site has a RSS feed. In simple terms, Weblogs and an ever-growing number of other sites including the homepages generate a behind-the-scenes code in XML. This code, usually referred to as a "feed" or "content feed" (as in "news feed"), makes it possible for readers to "subscribe" to the content that is created on a particular website so they no longer have to visit the site itself to get it. As is true with traditional syndication, the content comes to you instead of you going to it. RSS works in some what similar way to the Current Awareness Service used by librarians over the years. But the difference is in the flexibility of operation available in RSS.

4. Reasons for using RSS

Quite often we visit a set of websites in our area of our interest and research. Moreover, we visit them frequently. The reason for our frequent visit is to find the new information that might be available since our last visit. The frequency of visit depends upon our interest in those websites and the nature of the websites. We tend to visit, say, our favorite newspaper website more often than a professional journal website. In a newspaper website we are certain that new information is always available every day. On the other hand, we are not sure of finding new information every time we visit our favorite websites. We would rather like to visit them only if they contain new information since our last visit. This is where RSS comes into the picture. The RSS technology provides us an alert service about the new information that was added by sites. In other words, we can easily stay informed about the latest content from the sites that we are interested in. RSS helps us to save time by not needing to visit each site individually. More importantly, it ensures privacy by not needing to join each site's email newsletter. Çelikbaş has summarizes the advantages of RSS as follows:

- Privacy: One does not need to send one's e-mail address anywhere, to anybody, in order to subscribe to any RSS feed.
- 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.
- Easy Cancellation: One does not need to send an 'unsub' e-mail message as with most e-mail listserv 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.
- 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).

5. The RSS Network Architecture

An RSS network consists of three major components:

1. A (large) number of content providers, each providing news articles, and each providing their own RSS files describing these news articles.
2. A (smaller) number of RSS aggregators that read these RSS files from multiple sources, collect them into an index, and provide customized "feeds" of topic-specific news headlines from this index.
3. A (large) number of news viewing applications that, based on user input, connect to an RSS aggregator, access a news feed, and display it to the reader. On viewing the news feed, the reader can then select a news item (by clicking on the headline) and read the article directly from the content provider.

The RSS network architecture looks like this:

![RSS Network Architecture Diagram]

5.1 RSS Syntax

RSS defines an XML grammar (a set of HTML-like tags) for sharing news. Each RSS text file contains both static information about your site, plus dynamic information about your new stories, all surrounded by matching start and end tags.

Each story is defined by an `<item>` tag, which contains a headline `<TITLE>`, `<URL>`, and `<DESCRIPTION>`. Here’s an example:

```xml
   <item>
     <title>RSS Resources</title>
     <link>http://www.webreference.com/authoring/languages/xml/rss/</link>
     <description>Defined in XML, the Rich Site Summary (RSS) format has quietly become a dominant format for distributing headlines on the Web. Our list of links gives you the tools, tips and tutorials you need to get started using RSS. 0323</description>
   </item>
```

Each RSS channel can contain up to 15 items and is easily parsed using Perl or other open source software.

6. How to subscribe for RSS feeds?

RSS feed readers or RSS aggregators are required to collect and read the RSS feeds. Different categories of readers are already available. According to Çelikbaş they are:
7. RSS and Libraries

RSS technology is really taking off. According to a study, 5% of Internet users say they use RSS. That's almost 46,700,000 users! Most news sites now publish RSS feeds. Aside from news, these feeds can also be used for announcing new products, Web site updates, blogs and anything else to keep your visitors informed. This is where the information professionals should take the cue from and use the technology for providing innovative services and help the users to keep themselves abreast of new information. The present article tries to identify the possible areas of usage of RSS feeds for information services in libraries.

Although the easiest and most obvious use for content syndication is in the production of relatively current lists of news links on a given topic, RSS developers are beginning to perceive that a wide range of uses will be possible. Ian Graham and Benet Devereux suggest the following:

- New bulletins or news summaries, currently largely distributed using a simple XML dialect called RSS. An example of this is My Netscape.
- Web site content replication or distribution (often done using tools such as rdist, which is Rdists is a program to maintain identical copies of files over multiple hosts. (MagniComp)
- Database-related content distribution, such as gathering event calendar data for use in a local calendar.

- Gnutella-like file/resource sharing services. This is a serve where multiple copies of the same file (for example, a music video) are located on different servers, with syndication information being used to facilitate retrieval.

- Dmoz.org-like catalogues. The Mozilla Open Directory project is a human-created directory of Web-accessible resources. This directory is available as an open-source archive (in RDF), and is integrated into many other Web cataloging systems (for example, Google or Lycos).

- The HEML (Historical Event Markup and Linking) Project. This is a project aimed at creating a world wide collection of history-research related XML resources, with each academic research group being able to create their own resources, which can then be syndicated and distributed amongst the different institutions.

- To aggregate proprietary scientific data, as described by David Detlefsen.

As Graham and Devereux point out, in each of these cases, "one organization publishes 'origin' data and makes it available in some form, and another organization downloads the data and processes the data to integrate it in some way into their own database or application."

Apart from the above mentioned general applications of RSS technology, libraries can set up an alert service, traditionally called "current awareness service". The library can make its website RSS enabled so that interested users can subscribe for RSS feed. Following are some thoughts about the options available for the libraries

- Marketing their information and services: Libraries can provide RSS feed either from their websites or blogs on activities, exhibitions, promotions and new library resources. RSS with weblogs provide a more interactive environment for these kinds of services.

- Book lists: New arrivals can be circulated electronically through RSS feed. Other lists that could be communicated are "most wanted" or "most reserved books". Little more imagination needed to find out what other lists that are locally required by their user community. One limitation of RSS feed is that it cannot take a very big list. The commercial book sellers have already using the RSS technology to promote their business. Amazon for example is providing RSS feeds which aggregate book news under the popular headlines. (<http://www.amazon.com/exec/obidos/subst/xs/syndicate.html>.)

- Table of Contents for Journals: 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>).

Some of these have been tried at SDM-IMD library, Mysore where one of the authors is working.
8. Conclusion

Technology has a democratic way of giving opportunity to every segment of society. Libraries too have many opportunities to utilize them to the benefit of their clients in particular and society in general. 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.

9. References

1. Çelikbaş, Zeki. What is RSS and how it serves libraries? Available at http://eprints.rclis.org/archive/00002531/01/RSS_and_libraries_EN3.pdf

About Authors

Dr. N S Harinarayana is a Lecturer in Department of Library and Information Science, University of Mysore, Mysore.
Email: harinarayana@lisc.uni-mysore.ac.in, ns.harinarayana@gmail.com

Dr. Mallinath Kumbar is a Reader in Department of Library and Information Science, University of Mysore, Mysore.
Email: mkumbar@lisc.uni-mysore.ac.in

Shri M V Sunil is a Librarian at SDM-IMD, Mysore.
Email: sunilmysore@gmail.com