RSS as medium for information and communication technology

Alen Šimec, Mia Čarapina, Sanja Duk
Tehnical Polytechnic in Zagreb/Vrbik 8, 10000 Zagreb, Croatia
alen@tvz.hr, mia.carapina@tvz.hr, sanja.duk@tvz.hr

Abstract - In today's modern time we exchange thousands of information with the environment. Over time various forms of information exchange were developed in order to facilitate our daily lives. Today we can use a different approach to exchange information such as correspondence by mail, telegraph, telephone, radio television and internet, etc. In order to preserve data and information they were transferred to a physical media such as paper, and we can conclude that it is a matter of technology that man has evolved over time to facilitate daily life. A term that encompasses the application of such technology is called Information and Communication Technology (ICT). Today, when we use the term information and communication technologies (hereinafter referred to as ICT) primarily we allude to a new technology. With development of ICT came an increasing need for mass information transfer, as for individuals, as for many people. New technologies develop because of the need for a simple way to inform. One such technology that is becoming more popular is the RSS. RSS stands for Really Simple Syndication or Rich Summary Site (at the very beginning of creation, RDF Site Summary). RSS content can be traced through browser or RSS aggregator. RSS aggregator comes in the form of client applications or web application accessed via the web browser. RSS aggregators have the function of collecting titles, descriptions and links to subscribed topics, mostly blogs, podcasts, accounts, news, etc.

I. INTRODUCTION

RSS (meaning "Really Simple Syndication") comes from a family of web feed formats used for publishing frequently updated content on the Internet, such as blog, news, audio and video in a standardized format. RSS document (which is called "feed", "web feed" or "channel") includes full or summarized text, plus "metadata" such as publishing, dates, and content authors.

II. WHAT IS RSS?

RSS stands for Really Simple Syndication or Rich Site Summary (in the very beginning of creation, RDF Site Summary). RSS is a format based on XML and is used for rapid distribution of various facilities. RSS can be considered as service "What is new on the website" and usually comes in the form of a title list for certain website containing a brief description, for which it has become one of the most popular XML formats. The biggest advantage of monitoring the RSS content is that we do not have to provide personal information such as e-mail address what reduces the possibility of virus infection. RSS content can be traced with the most of modern web browsers or special programs that are used for these purposes. Therefore, the monitoring of RSS content is much more elegant than browsing through individual web pages "manually" searching for specific information. RSS was developed by Netscape with purpose of filling in short news in their My Netscape portals. After some time, Netscape has stopped the development of RSS and further development was taken over by Userland Software which took control over the RSS specification. After publication of RSS version 2.0 further development was taken over by Harvard University Law under a Creative Commons license (RSS 2.0 at Harvard Law).

III. RSS ADVANTAGE

RSS allows us to engage in selected information sources, as well as independent reporters, researchers or industry analysts. It allows us immediate visibility with no subscription. Because of no mailing lists payment requirements it offers great accessibility and low cost information delivery. Data format does not have to be different from the news and media format contained on the website.

Compatibility is ensured by allowing e-mail subscribers with text, HTML, AOL or multipart MIME capabilities to receive well-formatted news updates that are perfectly compatible with the e-mail client. E-mail client is not required but can be fully integrated within the local e-mail client. RSS feeds and their fields can be easily read online and gathered on the website. Feed add-ons facilitate the organization of content. Important messages can be easily and automatically stored, sorted and organized by subject. Subscriber has complete control over the content. Subscription services and the removal of some RSS

1 http://cyber.law.harvard.edu/rss/rss.html
news fields is controlled by the user, unlike other forms of content subscription where users are not able to fully determine the exact content they want. Subscribers never need to provide e-mail address to their chosen information supplier. It is difficult to spam RSS subscribers because the source of news is always known. RSS is safe and cannot transmit viruses. Fields are not blocked by spam and e-mail filters.

RSS is practical because it is structured as multi content protocol that allows content to be reused for different purposes: victualling other channels with news, integration of dynamic libraries and facilities for learning.

RSS can be searched and fully addressed. The user is offered a possibility to modify RSS by changing current position or to correct the list of printing errors. Monitoring of RSS is supported by free and canvas content distribution.

IV. MONITORING RSS FEEDS

RSS content can be monitored using an RSS browser or aggregator. RSS aggregator comes in the form of client application. Application can be run on personal computer or as web application accessed via web browser. RSS aggregators have the function of collecting titles, descriptions and links to topics to which we are subscribed, mostly blogs, podcasts, accounts, news, etc. In order to subscribe to an RSS feed or newsfeed you will need two things, an RSS reader (also known as a news aggregator) and URL (web address) of the RSS feed that you wish to subscribe. A news aggregator or RSS reader is a software application that collects and displays news headlines and summaries from sources that you have designated. Today's RSS browsers have the option of displaying the formatted RSS content from the source with complete withdrawal of news appearance along with pictures, font size and type, etc. RSS content can be monitored without usage of any specific software. RSS can be monitored using any of popular web browsers such as Mozilla Firefox or Opera, or by e-mail clients like Mozilla Thunderbird, but their options are quite limited and usually offer only text and titles of RSS Feeds.

V. HOW NEWS AGGREGATORS OR RSS READERS WORK

Websites summarize content in an RSS feed (XML document). Visitors download an RSS reader (news aggregator). There are generally two different types of RSS readers. The first kind of feed reader is a self contained program and the second kind of feed reader uses a web browser. Many of the programs are free, but generally those that have a small fee are more robust. Visitors select the content and summaries what they wish to view in a news aggregator or RSS reader (news aggregator and RSS reader are essentially interchangeable). Content is added to the news reader by entering the URL or web address of the XML file. Clicking on the RSS link will provide the URL of the feed. Some RSS readers will auto-detect an XML file on a site indicating that a feed is available (assistance for those wanting to subscribe to RSS feeds). Each time the feed is updated the content being viewed in the RSS reader indicates that there is new content. This insures that the customer has current information related to the topics they choose.

Professionals estimate that RSS will soon rival e-mail's popularity as a content delivery method.

There are online services, a proxy between news reader application and third-party web sites that don't support RSS natively. It allows you to create your own news feeds in RSS format for any web site, i.e. monitor any web site using your favorite news reader. This engine converts free-form HTML or XML documents to valid RSS feeds by extracting pieces of text or HTML by means of applying search patterns, and then joining these pieces together using output templates to form user-friendly content of feed's items. The principle of extracting specific data from source documents is also known as “HTML scraping”. The procedure of setting up a feed is the following: first find a web page with the content that you want to follow and create a new feed on the site, which points to that web page. In feed parameters, define search patterns and output templates for this feed, and get the link to your feed. Subscribe to this feed using your favorite feed reader (aggregator). At the end, there is option to protect your feed with password so no one could alter it. After the feed is set up, the service works as follows: feed reader sends request to server to download the feed, service downloads the original source URL, processes it, converts to valid RSS feed on the fly, and returns it to feed reader application and feed reader displays the contents of this feed to you. Engine acts like a proxy and checks the source site at the exact moment you request the feed, all you need is to setup your news reader so it requests the feed at 9:00 am and 6:00 pm. However, some news readers may not support fixed feed update schedules. The owner of the site may request not to allow monitoring his site (or some specific pages) via this kind of service. This may happen if the respective site has its native news feed already and doesn't want anyone using 'unofficial' news feeds.

VI. CREATING RSS FEED

RSS has become the standard technology for automatically publishing information to large audiences. A great way to start is by creating your own RSS feed and adding to it as often as you can. One way to create your own custom RSS feeds using PHP and MySQL. First step is to create
database tables and then to retrieve data from them which will be formatted into an RSS feed.

An RSS feed consists of the main details for the feed, such as the title, description, URL, image and so on. Next are the items, probably the most important parts of the feed. For example, different stories from a newspaper, posts from a blog and so on. Therefore, database will contain two tables, the first is called web_rss_details, which contains the details for the feed and the second is called webref_rss_items, which contains all of the items. First table, web_rss_details, contains columns such as the id, title, description, link, language, image_title, image_url, image_link, image_width and image_height. Each of these columns are self-explanatory once you are familiar with the structure of an RSS feed, the non-RSS related column is the id, which is used to represent the id for the row of data in the database. The second table, web_rss_items, contains columns which include the id, title, description and link. URL typically links to the location of the original text, such as a specific blog page where an entry has been posted. When the MySQL tables are created, next step is to add data that you want to syndicate and proceed to construct a valid RSS feed with PHP. Using methods such as dbConnect, GetFeed, getDetails and getItems, PHP files connects to database and shows data as RSS.

VII. RSS SPECIFICATIONS

According to RSS v1 specification RDF Site Summary (RSS) is a lightweight multipurpose extensible metadata description and syndication format. RSS is an XML application, conforms to the W3C’s RDF specification and is extensible via XML.

Proposed RSS Spec. Changes - Changes proposed to the current specification.

RSS Specifications v2 - RSS originated in 1999, and has strived to be a simple, easy to understand format, with relatively modest goals. After it became a popular format, developers wanted to extend it using modules defined in namespaces, as specified by the W3C. RSS 2.0 adds that capability, following a simple rule. An RSS feed may contain elements not described, only if those elements are defined in the namespace.

RSS Specifications v1 - RDF Site Summary (RSS) is a lightweight multipurpose extensible metadata description and syndication format. RSS is an XML application, conforms to the W3C’s RDF Specification and is extensible via XML-namespace and/or RDF based modularization.

RSS Specifications v.93 - RSS specification v.93 enclosures per item (previous version only allowed one). Also all dates need to conform to the Date and Time Specification of RFC 822.

RSS Specification v.9 - places restrictions on the first non-whitespace characters of the data in the link and URL tags. RSS 0.9 supports the full ASCII character set, as well as all legal decimal and HTML entities. RSS 0.9 does not support other types of character data, such as UTF-8.

VIII. RSS VERSIONS

<table>
<thead>
<tr>
<th>Owner and Versions</th>
<th>Description</th>
<th>Status</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netscape 0.90</td>
<td>Replaced by 1.0</td>
<td>Do not use</td>
<td></td>
</tr>
<tr>
<td>UserLand 0.91</td>
<td>„Drop dead simple“</td>
<td>Replaced by 1.0 but still very popular</td>
<td>Used only for basic Easy migration path to 2.0 if you need more flexibility</td>
</tr>
<tr>
<td>UserLand 0.92</td>
<td>Allows richer and more metadata than 0.91</td>
<td>Replaced by 2.0</td>
<td>Use 2.0</td>
</tr>
<tr>
<td>UserLand 2.0</td>
<td>Extensible via modules, easy migration path from 0.9x branch</td>
<td>Stable core, active module development</td>
<td>Use for general purpose metadata</td>
</tr>
<tr>
<td>UserLand 2.0.1</td>
<td>Added &lt;rating&gt; element</td>
<td>Minimal changes</td>
<td>Use</td>
</tr>
</tbody>
</table>

IX. CONCLUSION

Although RSS is present in the market for several years it is relatively new term to many Internet users. The latest research reveals that even less than 10% of Internet users use RSS. The reason for this is a low awareness what RSS is and for what it is used. Customers in less developed countries still have no need for such type of information and time saving. RSS is technology in the swing that is still expanding and upgrading. It is
just a question of time before its meaning becomes equated with the usage of the web or e-mail.

X. LITERATURE

[6] Bruce C. Brown, How to Use the Internet to Advertise, Promote and Market Your Business or Website with Little or No Money, 2006.