INTRODUCTION

Rich site summary (RSS) is a type of XML document used to share Web contents. Originally designed by Netscape (http://www.netscape.com) to create customize Web channels, RSS has been adopted by news syndication services, Weblogs, Webcasting and online information services. RSS is thus also known as “Really Simple Syndication”. While around for many years, it is now quickly gaining momentum owing to RSS’s active “content-push” technology. RSS is also attractive because of the growing problems of spam making e-mail content delivery extremely challenging. As the data is in XML, RSS information can be handled by a large number of devices. The strength of RSS is its simplicity and universality. It is exceptionally easy to syndicate and deliver site content using RSS; and it is also very easy for the users to read RSS data feeds.

BACKGROUND

RSS is architecturally a distributed data network implemented using the XML standard (W3C, 2006). Contrary to traditional client-server model where data contents are contained in large centralized application servers, RSS has no central content repositories. RSS contents are totally distributed, not even existed as data bundles in a few servers, but as individual items in computers scattered all over the Internet. Users do not need to install complicated client programs to access
RSS CONTENTS; A SIMPLE RSS READER WILL BE ABLE TO GATHER THE INFORMATION AND DISPLAY THEM AS WEB PAGES.

An RSS network is built on three major components: provider, aggregator and reader (Hammersley, 2005). The network has numerous content providers. They provide contents such as news articles with RSS files describing and containing these articles. The network is served by a smaller number of RSS aggregators. They read the RSS files from various content providers, Web sources, collecting, indexing and providing customized “RSS feeds” of topic-specific contents to the readers. Reader application connects itself to an RSS aggregation. Based upon user input, contents are fed to the reader. Once the RSS feed is received, the user can select an item and view the information directly from the content provider (see Figure 1).

RSS VARIANCES & BASIC SYNTAX

The original RSS, version 0.90 and 0.91, was designed by Netscape as a format for building portals of headlines to mainstream news sites (Libby, 1999). When Netscape lost interest in portal business and subsequently halted the RSS development process, a non-commercial group RSS-DEV continued the effort (Dornfest, 2000). They redesigned RSS 0.9 to version 1.0 that conformed to the W3C RDF specification. RDF is an abbreviation for Resource Description Framework, which is a framework for describing and interchanging metadata for resources from Web sites, Web pages, XML documents and so forth.

In the meantime, UserLand Software (http://www.userland.com/) also picked up RSS as the basis for its Weblogging products and continued the development process from RSS 0.91 through versions 0.92, 0.93, 0.94, and finally to version 2.0 (Winer, 2005). RSS 2.0 is easy to read and learn, but with some notable design issues existed in the specification (Berlind, 2004). A third group, AtomEnabled Alliance, began a new syndication specification effort taking a evolutionary rather than revolutionary approach. The specification, ATOM 1.0, has been approved as an Internet engineering task force (IETF) standard (Nottingham & Sayre, 2005).

RSS files do not have a standard file extension, although they frequently end in either “.xml” or “.rss”. A single RSS file is typically called an RSS channel. The channel’s attributes include the name of the channel, a home URL and an image for the channel. Like a cable channel, it contains multiple news items from the same source channel. Usually, each item contains at least a title and URL, but other information such as unique identifier, publication date, and summary may also be presented (see Figure 2).

RSS 1.0 feeds look very similar to RSS 2.0 feeds but with more verbose as it needs to be compatible with other versions of RSS while containing the markups for RDF. The entire feed is wrapped in <rdf:RDF> ...</rdf:RDF> elements; and there is an <items> element in the channel metadata that contains a list of items in the channel for the RDF processors to keep track of the relationship between items. A more extensive exploration of RDF format is available in the W3C RDF specification (W3C, 2004); and such detail is outside the scope of this article.