Introduction to RSS – WIKIs

Chris Panayiotou

What is RSS?

- RSS (originally RDF Site Summary, often dubbed Really Simple Syndication) is a family of web feed formats used to publish frequently updated works in a standardized format
  - Blog entries, news headlines, audio, and video
- An RSS document includes full or summarized text, plus metadata such as publishing dates and authorship
  - Called a "feed", "web feed", or "channel"
- RSS feeds benefit publishers by letting them syndicate content automatically
  - A standardized XML file format allows the information to be published once and viewed by many different programs
  - They benefit readers who want to subscribe to timely updates from favorite websites or to aggregate feeds from many sites into one place
- RSS feeds can be read using software called an "RSS reader", "feed reader", or "aggregator", which can be web-based, desktop-based, or mobile-device-based
  - The user subscribes to a feed by entering into the reader the feed's URI or by clicking a feed icon in a web browser that initiates the subscription process
  - The RSS reader checks the user's subscribed feeds regularly for new work, downloads any updates that it finds, and provides a user interface to monitor and read the feeds
  - RSS allows users to avoid manually inspecting all of the websites they are interested in, and instead subscribe to websites such that all new content is pushed onto their browsers when it becomes available

RSS History

- The RSS formats were preceded by several attempts at web syndication that did not achieve widespread popularity
    - The basic idea of restructuring information about websites goes back to as early as 1995, when Ramanathan V. Guha and others in Apple Computer's Advanced Technology Group developed the Meta Content Framework.
- **RDF Site Summary**, the first version of RSS, was created by Dan Libby and Ramanathan V. Guha at Netscape
    - It was released in March 1999 for use on the My.Netscape.Com portal (RSS 0.9)
    - In July 1999, Dan Libby of Netscape produced RSS 0.91 which simplified the format by removing RDF elements and incorporating elements from Dave Winer's scriptingNews syndication format
        - Libby also renamed RSS to **Rich Site Summary**
        - This would be Netscape's last participation in RSS development for eight years.
- As RSS was being embraced by web publishers who wanted their feeds to be used on My.Netscape.Com and other early RSS portals, Netscape dropped RSS support from My.Netscape.Com in April 2001
    - Two entities emerged to fill the void, with neither Netscape's help nor approval: The RSS-DEV Working Group and Winer, whose UserLand Software had published some of the first publishing tools outside of Netscape that could read and write RSS
    - Winer published a modified version of the RSS 0.91 specification on the UserLand website, covering how it was being used in his company's products, and claimed copyright to the document
        - A few months later, UserLand filed a U.S. trademark registration for RSS, but the request was rejected in December 2001

RSS History

- The RSS-DEV Working Group (included Guha and representatives of O'Reilly Media and Moreover) produced RSS 1.0 in December 2000
  - This new version, which reclaimed the name RDF Site Summary from RSS 0.9, reintroduced support for RDF and added XML namespaces support, adopting elements from standard metadata vocabularies such as Dublin Core
- In December 2000, Winer released RSS 0.92 a minor set of changes aside from the introduction of the enclosure element, which permitted audio files to be carried in RSS feeds and helped spark podcasting
  - He also released drafts of RSS 0.93 and RSS 0.94 that were subsequently withdrawn
- In September 2002, Winer released a major new version of the format, RSS 2.0, that redubbed its initials Really Simple Syndication
  - RSS 2.0 removed the *type* attribute added in the RSS 0.94 draft and added support for namespaces.
  - To preserve backward compatibility with RSS 0.92, namespace support applies only to other content included within an RSS 2.0 feed, not the RSS 2.0 elements themselves
- Because neither Winer nor the RSS-DEV Working Group had Netscape's involvement, they could not make an official claim on the RSS name or format
  - Ongoing controversy in the syndication development community as to which entity was the proper publisher of RSS
  - Lead to the creation of an alternative syndication format, Atom, that began in June 2003
    - The Atom syndication format has been adopted as IETF Proposed Standard RFC 4287

RSS History

- In July 2003, Winer and UserLand Software assigned the copyright of the RSS 2.0 specification to Harvard's Berkman Center for Internet & Society
- In December 2005, the Microsoft Internet Explorer and Microsoft Outlook teams announced on their blogs that they were adopting the feed icon first used in the Mozilla Firefox browser
  - In February 2006, Opera Software followed suit
  - This effectively made the orange square with white radio waves the industry standard for RSS and Atom feeds
- In January 2006, Rogers Cadenhead relaunched the RSS Advisory Board without Dave Winer's participation, with a stated desire to continue the development of the RSS format and resolve ambiguities
- In June 2007, the board revised their version of the specification to confirm that namespaces may extend core elements with namespace attributes, as Microsoft has done in IE7

RSS Example

- RSS files are essentially XML formatted plain text
  - The RSS file itself is relatively easily read both by automated processes and by humans alike
    - Could be placed on any appropriate communication protocol for file retrieval, such as http or ftp, and reading software would use the information to present a neat display to the end user

RSS Example

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
<channel>
        <title>RSS Title</title>
        <description>This is an example of an RSS feed</description>
        <link>http://www.someexamplerssdomain.com/main.html</link>
        <lastBuildDate>Thu, 16 Mar 2012 00:01:00 +0000 </lastBuildDate>
        <pubDate>Wed, 15 Mar 2012 16:45:00 +0000 </pubDate>
        <ttl>1800</ttl>

        <item>
                <title>Example entry</title>
                <description>Here is some text containing an interesting
                                        description.</description>
                <link>http://www.wikipedia.org/</link>
                <guid>unique string per item</guid>
                <pubDate>Wed, 15 Sep 2012 16:45:00 +0000 </pubDate>
        </item>

</channel>
</rss>
```

More on RSS

- **Modules**
  - The primary objective of all RSS modules is to extend the basic XML schema established for more robust syndication of content
  - This inherently allows for more diverse, yet standardized, transactions without modifying the core RSS specification
  - To accomplish this extension, a tightly controlled vocabulary (in the RSS world, "module"; in the XML world, "schema") is declared through an XML namespace to give names to concepts and relationships between those concepts
  - Some RSS 2.0 modules with established namespaces are:
    - Media RSS 2.0 Module
    - OpenSearch RSS 2.0 Module
- **Interoperability**
  - Although the number of items in an RSS channel are theoretically not limited, some news aggregators do not support RSS files larger than 150KB (if all elements are provided on a new line, this size corresponds to approx. 2,800 lines)
  - For example, applications that rely on the Common Feed List of Windows might handle such files as if they were corrupt, and not open them
  - Interoperability can be maximized by keeping the file size under this limit
- **BitTorrent and RSS**
  - Some BitTorrent clients support RSS
    - RSS feeds which provide links to .torrent files allow users to subscribe and automatically download content as soon as it is published

Comparison with Atom

▸ Both RSS and Atom are widely supported in all major consumer feed readers
  ◦ RSS has gained widespread distribution through feed reader implementations earlier than Atom, which led to higher popularity
  ◦ However, Atom has several advantages over RSS, such as less restrictive licensing, IANA registered MIME type, an XML namespace, support for relative URIs, and Relax NG support
  ◦ Technically, Atom should be considered the more advanced syndication format between the two

Comparison with Atom

RSS 2.0	Atom 1.0
author	author
category	category
channel	feed
copyright	rights
description	subtitle
description	summary and/or content
generator	generator
guid	id
image	logo
item	entry
lastBuildDate (in channel)	updated
link	link
managingEditor	author or contributor
pubDate	published (subelement of entry)
title	title
ttl	–

RSS Required Channel Elements

- The required channel elements of RSS 2.0 are title, link, and description
- **The title element**
  - The title element represents the name of the channel
  - It often coincides with the title of the website it is associated with
    <title>John Smith Headlines</title>
- **The link element**
  - The link element is a URI representing the domain where the news feed is located
    <link>http://example.com/</link>
- **The description element**
  - The description is a sentence or sentence fragment that describes the channel
    <description>The latest news about rock star John Smith.</description>

RSS Optional Channel Elements

- In RSS 2.0 news feeds the channel element has 16 optional sub-elements
  - category, cloud, copyright, docs, generator, image, language, lastBuildDate, managingEditor, pubDate, rating, skipDays, skipHours, textInput, ttl, and webMaster
- A common feature of all RSS 2.0 elements providing a URL is that they should begin with a URI scheme defined by IANA
  - http://, https://, news://, mailto://, or ftp://
  - Note that the http:// and ftp:// schemes cannot be used in earlier versions

RSS Namespaces

- The default namespace for RSS is http://purl.org/rss/1.0/, which is the permanent URL form of the RDF Site Summary (RSS) 1.0 namespace, http://web.resource.org/rss/1.0/. The namespace can be provided in the form
  <rss version="2.0" xmlns:rss="http://purl.org/rss/1.0/">
- Additional data on channel updates can be provided by the Web syndication namespace of RSS (http://purl.org/rss/1.0/modules/syndication/)
  - It extends the RSS channels with three elements:
    - The period over which the news channel is updated can be described by the sy:updatePeriod element
      - Allowed values are hourly, daily, weekly, monthly, and yearly. If omitted, daily is assumed
    - The frequency of updates can be expressed in relation to the update period with the sy:updateFrequency element
      - Its value is a positive integer
    - To calculate the publishing schedule, a base date can be defined by the sy:updateBase element
      - It should be a #PCDATA date in one of the W3C date and time formats

RSS Namespaces

- By default, news feed entries are plain text contents
  - However, news aggregators often support (X)HTML markup that are not allowed in XML
  - Entity-encoded and CDATA-escaped contents can be provided with the content:encoded element defined by the http://purl.org/rss/1.0/modules/content/ namespace
    - The content:encoded element is especially useful if the hyperlink delimited by the link element is not enough and additional hyperlinks are needed (in the news item content)
    - Although text formatting and other markup codes can also be written this way, they are ignored by many RSS readers

RSS Namespaces

- There is an Atom element, atom:link, that can be used to provide the self-link of the news feed channel
  - To apply this element, the Atom namespace *http://www.w3.org/2005/Atom* should be declared
- Advanced news feeds typically contain at least the following namespace declarations:

```
<rss version="2.0"
xmlns:content="http://purl.org/rss/1.0/modules/content/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:sy="http://purl.org/rss/1.0/modules/syndication/"
xmlns:atom="http://www.w3.org/2005/Atom" >
```

  - Thus, elements can be used from these namespaces in the channel:

```
<dc:creator>John Smith</dc:creator>
<sy:updatePeriod>daily</sy:updatePeriod>
<sy:updateFrequency>1</sy:updateFrequency>
<sy:updateBase>2011-01-01T12:00+00:00</sy:updateBase>
<atom:link href="http://www. example.com/example.xml" rel="self"
                                        type="application/rss+xml" />
```

  - or in item elements such as

```
<content:encoded>
    <![CDATA[ An escaped RSS item can contain markup elements such as
    <a href="http://www.example.com/">hyperlinks</a> that work in all major
    news feed readers]]>
</content:encoded>
```

Styling RSS Feeds

- The browsers that support news feeds usually provide a basic styling or no styling at all (rendering a tree structure instead)
- Developers who are not satisfied with that or want to ensure an advanced look can format RSS channels using CSS or XSLT
  - A CSS reference can be given as:
    ```
    <?xml version="1.0" encoding="utf-8" ?>
    <?xml-stylesheet type="text/css" href="css/feed.css" ?>
    <rss version="2.0">
    ```
  - Much information is not necessarily relevant and can be omitted, for example:
    ```
    channel link, channel language, channel copyright, channel
    managingEditor, channel webMaster, channel docs, channel
    lastBuildDate { display: none; }
    ```
  - XSL Transformation provides more control (e.g. hyperlinks can be activated and node order changed). The XSL file can be linked as follows:
    ```
    <?xml version="1.0" encoding="utf-8" ?>
    <?xml-stylesheet type="text/xsl" href="css/feed.xsl" ?>
    <rss version="2.0">
    ```
- Note that additional functionality such as searching or category listings provided by the built-in RSS reader of certain browsers are not available when custom style sheet are applied to a news feed

What is a WIKI?

- A **wiki** is a website whose users can add, modify, or delete its content via a web browser using a simplified markup language or a rich-text editor
- Wikis are typically powered by wiki software and are often created collaboratively by multiple users
- Examples include community websites, corporate intranets, knowledge management systems, and notetaking.
- Wikis may serve many different purposes
  - Some permit control over different functions (levels of access)
  - For example, editing rights may permit changing, adding or removing material
  - Others may permit access without enforcing access control
  - Other rules may also be imposed for organizing content
- Ward Cunningham, the developer of the first wiki software, WikiWikiWeb, originally described it as *"the simplest online database that could possibly work"*.
- "Wiki" is a Hawaiian word meaning "fast" or "quick"

WIKI History

- WikiWikiWeb was the first wiki
- Ward Cunningham started developing WikiWikiWeb in Portland, Oregon, in 1994, and installed it on the c2.com on March 25, 1995
- It was named by Cunningham, who remembered a Honolulu International Airport counter employee telling him to take the "Wiki Wiki Shuttle" bus that runs between the airport's terminals
  - According to Cunningham, *"I chose wiki-wiki as an alliterative substitute for 'quick' and thereby avoided naming this stuff quick-web."*
- Cunningham was in part inspired by Apple's HyperCard
  - Apple had designed a system allowing users to create virtual "card stacks" supporting links among the various cards
  - Cunningham developed Vannevar Bush's ideas by allowing users to *"comment on and change one another's text."*
- In the early 2000s, wikis were increasingly adopted in enterprise as collaborative software
  - Common uses included project communication, intranets, and documentation, initially for technical users
  - Today some companies use wikis as their only collaborative software and as a replacement for static intranets, and some schools and universities use wikis to enhance group learning
  - There may be greater use of wikis behind firewalls than on the public Internet
- On March 15, 2007, *wiki* entered the online *Oxford English Dictionary*

WIKI Characteristics

- Ward Cunningham and co-author Bo Leuf, in their book *The Wiki Way: Quick Collaboration on the Web*, described the essence of the Wiki concept as follows:
  - A wiki invites all users to edit any page or to create new pages within the wiki Web site, using only a plain-vanilla Web browser without any extra add-ons
  - Wiki promotes meaningful topic associations between different pages by making page link creation almost intuitively easy and showing whether an intended target page exists or not
  - A wiki is not a carefully crafted site for casual visitors. Instead, it seeks to involve the visitor in an ongoing process of creation and collaboration that constantly changes the Web site landscape
- A wiki enables communities to write documents collaboratively, using a simple markup language and a web browser
  - A single page in a wiki website is referred to as a "wiki page", while the entire collection of pages, which are usually well interconnected by hyperlinks, is "the wiki"
- A wiki is essentially a database for creating, browsing, and searching through information
  - A wiki allows for non-linear, evolving, complex and networked text, argument and interaction
- A defining characteristic of wiki technology is the ease with which pages can be created and updated
  - Generally, there is no review before modifications are accepted
    - Many wikis are open to alteration by the general public without requiring them to register user accounts, can be made in real-time and appear almost instantly online. This can facilitate abuse of the system
  - Private wiki servers require user authentication to edit pages, and sometimes even to read them

Editing WIKI Pages

- There are many different ways in which wikis have users edit the content
- Ordinarily, the structure and formatting of wiki pages are specified with a simplified markup language, sometimes known as wikitext (for example, starting a line of text with an asterisk often sets up a bulleted list)
  - The style and syntax of wikitexts can vary greatly among wiki implementations, some of which also allow HTML tags
- Designers of wikis often take this approach because HTML, with its many cryptic tags, is not very legible, making it hard to edit
  - Wikis therefore favor plain-text editing, with fewer and simpler conventions than HTML, for indicating style and structure
  - Limited access to CSS promotes consistency in the look and feel, and having JavaScript disabled prevents a user from implementing code that may limit access for other users

WIKI Example

MediaWiki syntax	Equivalent HTML	Rendered output
"Take some more [[tea]]," said Alice.	"Take some more tea," said.	"Take some more tea," said.
"I've had '''nothing''' yet," John replied, "so I can't take more."	"I've had nothing yet," John replied, "so I can't take more."	"I've had **nothing** yet," John replied, "so I can't take more."
"You mean you can't take "less"?" said Alice. "It's very easy to take "more" than nothing."	"You mean you can't take <i>less</i>?" said the Alice. "It's very easy to take <i>more</i> than nothing."	"You mean you can't take *less*?" said the Alice. "It's very easy to take *more* than nothing."

Editing WIKI Pages

- Increasingly, wikis are making WYSIWYG editing available to users, usually by means of JavaScript or an ActiveX control that translates graphically entered formatting instructions into the corresponding HTML tags or wikitext
  - In those implementations, the markup of a newly edited, marked-up version of the page is generated and submitted to the server transparently, shielding the user from this technical detail
  - However, WYSIWYG controls do not always provide all of the features available in wikitext, and some users prefer not to use a WYSIWYG editor
  - Hence, many of these sites offer some means to edit the wikitext directly
- Most wikis keep a record of changes made to wiki pages; often, every version of the page is stored
  - This means that authors can revert to an older version of the page, should it be necessary because a mistake has been made or the page has been vandalized
  - Many implementations, like MediaWiki, allow users to supply an *edit summary* when they edit a page; this is a short piece of text summarizing the changes
  - It is not inserted into the article, but is stored along with that revision of the page, allowing users to explain what has been done and why
    - Similar to a log message when making changes to a revision-control system

Navigation

- Within the text of most pages there are usually a large number of hypertext links to other pages
  - This form of non-linear navigation is more "native" to wiki than structured/formalized navigation schemes
  - That said, users can also create any number of index or table-of-contents pages, with hierarchical categorization or whatever form of organization they like
    - These may be challenging to maintain by hand, as multiple authors create and delete pages in an ad hoc manner
  - Wikis generally provide one or more ways to categorize or tag pages to support the maintenance of such index pages
- Most wikis have a backlink feature, which displays all pages that link to a given page
- It is typical in a wiki to create links to pages that do not yet exist, as a way to invite others to share what they know about a subject new to the wiki

Linking and Creating Pages

- Links are created using a specific syntax, the so-called "link pattern"
- Originally, most wikis used CamelCase to name pages and create links
  - These are produced by capitalizing words in a phrase and removing the spaces between them (the word "CamelCase" is itself an example)
  - While CamelCase makes linking very easy, it also leads to links which are written in a form that deviates from the standard spelling
  - To link to a page with a single-word title, one must abnormally capitalize one of the letters in the word (e.g. "WiKi" instead of "Wiki")
    - CamelCase-based wikis are instantly recognizable because they have many links with names such as "TableOfContents" and "BeginnerQuestions."
  - It is possible for a wiki to render the visible anchor for such links "pretty" by reinserting spaces, and possibly also reverting to lower case
    - However, this reprocessing of the link to improve the readability of the anchor is limited by the loss of capitalization information caused by CamelCase reversal
    - There is no easy way to determine which capital letters should remain capitalized
    - As a result, many wikis now have "free linking" using brackets, and some disable CamelCase by default

Implementations

- Wiki software is a type of collaborative software that runs a wiki system, allowing web pages to be created and edited using a common web browser
  - It is usually implemented as an application server that runs on one or more web servers
  - The content is stored in a file system, and changes to the content are stored in a relational database management system
  - A commonly implemented software package is MediaWiki
- Alternatively, personal wikis run as a standalone application on a single computer
  - WikidPad is an example
  - Or even single local HTML file with JavaScript inside – like TiddlyWiki
- Wikis can also be created on a "wiki farm", where the server side software is implemented by the wiki farm owner
  - PBwiki, Socialtext, Wetpaint, and Wikia are popular examples of such services
  - Some wiki farms can also make private, password-protected wikis
  - Note that free wiki farms generally contain advertising on every page

Notable WIKI Software

- Java-based
  - JAMWiki is a J2EE application released under the LGPL. JAMWiki is a Java clone of the MediaWiki software and uses the same wiki syntax
  - JSPWiki is a J2EE application released under the Apache License
  - XWiki is a Java wiki engine with a complete wiki feature set (version control, attachments, etc.) and a database engine and programming language which allows database driven applications to be created using the wiki interface
- JavaScript-based
  - TiddlyWiki is a HTML/JavaScript-based server-less wiki in which the entire site/wiki is contained in a single file.
  - Lively Wiki is based on Lively Kernel and combines features of wikis and development environments. Users can create and edit application behavior as well as other content.

Notable WIKI Software

- .NET-based
  - FlexWiki is written in C#, uses the .NET Framework, and stores data in files or Microsoft SQL Server.
  - MindTouch is an open-source application that began as a fork of MediaWiki; it has a C# back-end and a PHP front-end.
  - Screwturn Wiki is based on Microsoft ASP.NET 3.5 and can store data in a Microsoft SQL Server database or using the file-system.
- Perl-based
  - ikiwiki, a "wiki compiler" – can use Subversion or git as the backend storage mechanism.
  - TWiki is a structured wiki, typically used to run a project development space, a document management system, a knowledge base, or any other groupware tool. Also available as a VMware appliance.
    - Foswiki is a fork of TWiki.
  - UseModWiki (created by Clifford Adams in 2000) is a clone of AtisWiki.
    - Oddmuse is a fork of UseModWiki.

Notable WIKI Software

- **PHP-based**
  - DokuWiki is aimed at the documentation needs of a small company. It uses plain text files and has a simple but powerful syntax which ensures the datafiles remain readable outside the wiki.
  - MediaWiki was custom-designed for the high-volume Wikipedia encyclopedia website; it is also used for all other projects run by the Wikimedia Foundation, which operates Wikipedia. It is also publicly available for use in other wikis, and has widespread popularity among smaller, non-Wikimedia wikis.
    - SMW+ is a package of MediaWiki with semantics-oriented extensions
  - PhpWiki is a WikiWikiWeb clone in PHP.
  - PmWiki is a PHP-based wiki. Features include: GPL-licensed, easy installation/customization, designed for collaborative authoring and maintenance of web sites, and support for internationalization. Does not require a database.
  - PukiWiki is intended for Japanese wikis
  - TigerWiki is a minimalist and simple wiki framework, that has been discontinued
  - WakkaWiki is a now-defunct PHP/MySQL-based lightweight wiki engine, that notably spawned a large number of forks, including WackoWiki and WikkaWiki.

Notable WIKI Software

- **Python-based**
  - ◦ MoinMoin is a wiki engine written in Python. Offers good access control based on user groups
  - ◦ Zwiki is a Zope-based GPL wiki engine. It can integrate with the content management framework Plone, and supports several kinds of markup as well as WYSIWYG HTML editing
- **Other languages**
  - ◦ Gitit is a Happstack-based wiki server employing git or Darcs to manage wiki history, and the Pandoc document conversion system to manage markup – among other things permitting the inclusion of LaTeX mathematical markup
  - ◦ ProjectForum is a self-contained wiki server for Windows, Mac OS X and Linux, written in Tcl, C and C++
  - ◦ Swiki is written in Squeak. It runs on common platforms, including Mac, Windows, Linux, as well as others
  - ◦ Wiki Server is proprietary software distributed with Mac OS X Server
  - ◦ Canvas Wiki is a ColdFusion-based wiki server http://canvas.riaforge.org/