Library Mashups for the Virtual Campus: Using Web 2.0 Tools to Create a New Current Awareness Service

Linda Absher, Adriene Lim, and Kerry Wu

Abstract
This article explores the ways academic libraries can exploit Web 2.0 technologies and techniques to create Current Awareness Service (CAS) mashups for campus virtual environments. Described are Portland State University Library’s efforts to create a CAS site called “Topic Watch;” the criteria used to select humanities and business content for the project; and an overview of the technology employed to combine data from disparate sources into the service. Also summarized are the challenges and issues the library faced during development and future steps to be taken to incorporate the new resource into the campus virtual environment.

Introduction
The need for academic library support of remote learning continues to expand, as many campuses strengthen distance learning programs and users of all types continue to access library collections and services from off-campus. Users rely on libraries to help sift through sometimes overwhelming amounts of information, but may at times need alternatives to traditional library instruction and reference services. Libraries have responded to this reality by publishing electronic research guides and tutorials while integrating help content and virtual reference services into their Websites. Complementary components of these efforts are Web-based Current Awareness Services (CAS), designed to help time-constrained users keep abreast of the latest developments, issues, and activities related to selected topics in their academic fields.

Current Awareness Services in Libraries: An Overview
CAS has had a long history in the library world, although mostly within the special and corporate library realm. But what exactly is current awareness? Kemp defines it as:

A system or publication for reviewing newly available documents, selecting items relevant to the needs of an individual or group, and recording them so that notifications may be sent to those…to whose needs they are related (Kemp 1999).

According to Fourie, academic libraries did not historically implement CAS because such services were...
Developed by the University of Washington’s School of Law Library. Trial Ad Notes, authored by a librarian at the University of Manchester and the British Library; E-Commerce Alert from the University of Alberta; and Zetoc: A CAS providing relevant literature titles by way of the British Library’s Electronic Table of Contents, Zetoc is a joint project of the University of Manchester and the British Library. Subscribers are not only apprised of new titles via a search of its database or via e-mail alerts, but are also provided with an OpenURL article link and/or a link to document delivery options available with the British Library. Recently a RSS (Really Simple Syndication) journal service feed was added, syndicating new arrivals into Zetoc. Along with citations, subscribers receive the bibliographic citation information as well as document delivery options (Apps 2006).

**E-Commerce Alert:** Developed by the University of Alberta libraries, E-Commerce Alert was devised to help students and faculty cope with an overabundance of e-commerce information resources, particularly from electronic newsletters. With librarians as intermediaries, subscribers receive e-mail alerts with irrelevancies and repetitive information removed. Though Williamson does not describe in detail the process by which librarians filter the information, she admits the service is “labor-intensive” and takes “a disproportionate amount of time.” However, she asserts that once the initial focus was improved, the intermediation time was significantly reduced (Williamson 2004).

**Trial Ad Notes:** One library-related blog specifically labeling itself as a CAS is Trial Ad Notes. Begun in 2005, it provides faculty and students with updates regarding legal and trial advocacy information. Though the author is unclear as to its impact upon her target audience, she mentions it is “too early” to determine if the service warrants the time involved, implying intermediation may be too unwieldy to continue the service (Wisner 2006).

Though not specifically labeled a CAS, many libraries use Web logs (or “blogs”) and RSS feeds to update or alert readers about new journal holdings, articles, book titles or general library news. In the article “This Just In”, Gerry McKiernan lists some services that could be construed as CAS, such as new acquisition RSS feeds from the University of Louisville Libraries and the Hong Kong University of Science and Technology (McKiernan 2005). At the time of this writing, there is a dearth of information in the library literature about any library CAS involving the merger or mashup of Web 2.0 technologies or tools.

**Web-based Current Awareness Services in Academic Libraries**

Most electronic CAS offered by academic libraries to date provide literature/title updates from table of content resources or alert users to new library acquisitions, with blogs being an increasingly preferred delivery mode. Three services stand out as noteworthy: Zetoc from the University of Manchester and the British Library; E-Commerce Alert from the University of Alberta; and Trial Ad Notes, authored by a librarian at the University of Washington’s School of Law Library.

Selection of one or more systems that provide notification of the existence of new entities added to the system’s database or of which the system took note (e.g., documents, Websites, events such as conferences, discussion groups, editions of newsletters). Current awareness services automatically notify users or allow users to check periodically for updates (Fourie 1999). In terms of CAS during the pre-Web 2.0 era, academic librarians provided services still requiring constant intermediation, such as monitoring a variety of online services, filtering results, and delivering updates via an electronic list or e-mail updates (Fourie 1999; Williamson 2004). With the advent of Web 2.0 technologies and their abilities to push and pull content in new ways, CAS mashups give libraries the potential to maintain CAS offerings for end users with less need for significant intervention.

**Background and Definitions**

Web 2.0 is broadly defined as “second-generation Internet-based services,” which includes social networking sites, wikis, tagging, podcasts/videocasts and other interactive services, technologies that are currently popular topics in the library world (Wikipedia 2007). These new applications and their associated RSS feeds are being leveraged by many users to access content in innovative ways. Libraries are taking advantage of these developments as content continues to evolve from its...
location-bound past to its dynamic, just in time, re-usable present and future. Many content providers currently employ RSS to create resources to provide users with updates without the need of an external RSS feed reader, embedding content directly into a Web page. (Please note that the aforementioned definition of Web 2.0 and subsequent definitions in this section, unless listed otherwise, come from the well-known Web 2.0 site, Wikipedia.)

**Really Simple Syndication (RSS)**

From the latest weather forecast to the most recent blog post, RSS is the backbone by which Web content providers deliver the latest updates. RSS is an XML format designed to allow online content generators to syndicate and distribute news and/or other types of information. The typical RSS file describing a feed or “channel” contains three mandatory elements: a title, the URL and a description (Pilgrim 2002). Other elements or tags are optional (Tennant 2003). The format’s popularity is primarily due to the fact that it allows Web publishers to stream their content to subscribing readers, rather than forcing the user to periodically check websites for updates. When RSS use first became widespread, most users had to use an intermediary program called a reader or aggregator which detects updates from subscribed sites. To subscribe, one simply adds the URL or RSS feed information to their aggregator, which then checks and displays updates from the site. To view updates, readers visited their aggregator on the Web (such as bloglines.com, my.yahoo.com, etc.) or launched it as an application (i.e., iTunes, Thunderbird, etc.). Though this is still the case, users may now use RSS-enabled browsers to subscribe to feeds, while some content providers embed RSS-derived content directly into a web-based resource.

**Blogs**

A blog or Web log is a “user-generated website”, usually commenting upon subjects of the user’s choosing, with entries posted in a journal format in reverse chronological style. Though mostly textual, many blogs focus on other formats, such as video (videocasts), audio (podcasts) and photos. Most, if not all, blogs use RSS to syndicate their content.

**Podcasts and Vodcasts**

Podcasts are files that are gradually replacing the old-fashioned, static Web recordings. Vodcasts, or Videocasts, are relatively new, and offer video content. As broadband access becomes more readily available, vodcasts will surely gain momentum in the near future.

**Wikis**

Wikis are websites “that allows the visitors themselves to easily add, remove, and otherwise edit and change available content...” Due to its collaborative environment, offering a means for users to modify content, wikis fall into the array of tools associated with Web 2.0. Perhaps the most well known wiki is Wikipedia, a site originating in 2001 which is now the largest reference Web site on the Internet (Wikipedia).

**Social Bookmarks**

Social bookmarking is a method whereby users share resources and links with others via a Web site. Users store, share, and even categorize their resources with a homed grown classification system (often called “folksonomy” or “tagging”). Though subjects are often criteria for social bookmarking sites, format (i.e., videos, images, music, etc.) is often the basis for building such sites. Arguably the most famous example would be del.icio.us. According to the site, “the primary use of del.icio.us is to store your bookmarks online, which allows you to access the same bookmarks from any computer and add bookmarks from anywhere” (del.icio.us).

**Mashups**

The term “mashup” is used in the music world referring to the combination and/or remixing of existing musical elements into a new form, according to many experts, but now has been adopted by information technologists to refer to new Web applications or sites created through the combination and re-use of data and/or functionality from two or more disparate and usually third-party Web sources (Fichter 2006; Wikipedia 2007). With the abundance of human energy, creativity, and knowledge unleashed in public sites on the Web, the proliferation of mashup services capitalizing on high-value content and functions was perhaps inevitable. Sources for mashups include: content extracted with Application Programming Interfaces (APIs) from existing Web services; data streams from blogs, wikis, and other social networking and news sites delivered via RSS feeds; and Web pages “screenscraped” to harvest desired information. A substantial list of mashups is available on ProgrammableWeb, where growing numbers of APIs available for mashups are also registered.
Related Work

There are seminal articles that describe the concept, vision, and implications of Web 2.0 and its relationship to Library 2.0. The Talis white paper “Do Libraries Matter? The Rise of Library 2.0” defines Library 2.0, discusses the challenges libraries face in a new technological environment, and concludes libraries must “evolve and begin to deliver its services in the ways that its modern users expect” (Chad and Miller 2005). Talis is optimistic the new information environment is not a threat to libraries, but offers “great opportunities for progressive libraries to reach out…and to engage with an increasingly literate body of information consumers” (Miller 2006). Paul Miller, the Talis “Technology Evangelist” points out “there is much of value with which libraries should be seeking to engage, that participation doesn’t just come from the users, but also from all parties in the information industry.” (Miller 2005). In his twelve-page Library Technology Report entitled “Web 2.0 and Libraries: Best Practices for Social Software,” Michael Stephens examines the definitions, implementation, and best practices for Web 2.0 tools such as blogs, RSS, instant messaging, wikis, and Flickr (Stephens 2006).

Blogs also serve as an information service, providing library updates; forums for book discussions; listings for popular or award-winning books; a feedback tool for soliciting comments and conducting polls; and as a current awareness aid. (Kajewski 2006). Some libraries have gone further by converting the library Web site, or at least part of it, into a blog. The Ann Arbor District Library not only has a homepage that looks like a blog, it offers in its “Catalog” section, three separate blogs on books, audio resources, and videos. University of Minnesota supports blogging by faculty, staff and students; its UThink site contains approximately 4,000 blogs with over 8,000 authors (Coombs 2006).

An early article about RSS is Karen J. Bannan’s aptly titled “RSS” in the January 2002 issue of EContent. (Bannan 2002) Since then there has been a plethora of articles on various Web 2.0 applications and how libraries exploit these technologies to provide innovative services. Some libraries provide simple new titles alerts (for example, Catalogue RSS from Curtin University of Technology Library), while others, such as the Hennepin Public Library, offer a comprehensive list of RSS feeds ranging from library events announcements to subject guides updates (Kajewski 2006). In their recently published article, Corrado and Moulaison explain how they succeeded in integrating RSS feeds of new subject-specific books into the campus course management system (Corrado and Moulaison 2006). For additional information about RSS feeds and libraries, visit the RSS(sm): Rich Sites Services Web site (McKiernan 2004).

Wikis can benefit both librarians and their users. Examples of wikis made by and for librarians include LISWiki Library Instruction Wiki and Library Success, A Best Practices Wiki. Wikis intended for library users include Ohio University’s Biz Wiki, BookLoversWiki from Princeton Public Library (Kajewski 2006), and Subject Guides from St Joseph County Public Library (Coombs 2006).

Public libraries, such as the Orange County Library System are also taking the lead in integrating digital video and podcasts into their services (Kajewski 2006). Another multimedia movement to provide content and encourage library involvement is YouTube. Denver Public Library recently held a YouTube Contest for young adults. Social networking and bookmarking sites (e.g. del.icio.us, ma.gnolia.com, FURL, rawsugar.com) are also a popular theme in the library literature. However, many articles have been written about how social networking and bookmarking sites can benefit library services, but few are available about what people have actually done to promote such services. Rachel Bridgewater mentions, in her presentation at the 2006 Online Northwest conference, that the reference department at her library has a del.icio.us account that serves as an internal knowledgebase. Another project is PennTags from the University of Pennsylvania, a “social bookmarking tool for locating, organizing, and sharing” favorite online resources from members of the Penn Community. UPenn-affiliated members can collect and maintain URLs, links to journal articles, and records in their online catalogs. (Salo 2006)
Some libraries have developed mashups with Google Maps, such as Libraries 411, a mashup of public library directory and Google Maps (Fichter 2006). Darlene Fichter lists several ways libraries can take advantage of “a map mashup, such as providing a map of walking tours for their area with links to archival photographs and books in their collection,” or using “color-coded pushpins to indicate which branches are currently open at any given moment.” Fichter also suggests that libraries create non-map mashups, such as an RSS mashup mixing and displaying feeds from multiple sources, or a mashup pulling together “top links, photos, and stories for their community by using a variety of data streams on topics pertinent to users,” similar to services offered by DailyMashup and Diggdot. Libraries still seem however, to be at the beginning stages. For example, Ann Arbor Public Library offers a mashup combining new book RSS feeds and book cover photographs (Fichter 2006). Additional information about library-related mashups is searchable via keyword on the ProgrammableWeb, as well as other sites such as Technorati.

Some vendors and consortia actively encourage the creation of library mashups. Go-go Google Gadget, the winner of the Mashing up the Library 2006 contest sponsored by Talis, provides little wedges of personalized library-related information to be displayed on a personalized Google homepage. Book Burro, a Firefox extension, won the 2nd prize of the OCLC 2nd Annual Software Contest, enabling users to check prices of a book from online stores such as Amazon.com and Half.com, as well as viewing the book’s availability from the local library.

Building the New Current Awareness Service

The availability of powerful new Web 2.0 tools and techniques for automating content inspired the Portland State Library to explore whether its traditional new-titles alerts could be combined with subject-specific content from Web sites and custom search queries into a convenient, interactive CAS interface.

The library hoped to build components leveraging the latest advances in Web 2.0 technology, especially ones that could be exploited to minimize the need for ongoing human intervention, in order to create a content-rich Current Awareness Service (CAS) that would be sustainable for long-term use.

With a plan to use selected RSS feeds as major building blocks for the content, the idea for a CAS mashup called “Topic Watch” was formed. Project participants chose Business Research and Film Studies as proof-of-concept topics. The library hoped to move beyond what are now well-documented efforts to create and consume RSS data in a library context, to incorporate filtered search results from the library’s federated searching system and from Google Custom Searches to offer a new type of CAS tool.

Technical Overview of the Service

The creation of a presentation framework with embedded RSS feeds seemed to be a natural fit for the mash-up’s first stage, since it allowed librarians to leverage their existing del.icio.us collections of topical RSS feeds. External RSS feeds for CAS had been used previously on the Portland State Library’s Web site, but only in a rudimentary fashion, using the RSS-aggregator functionality within the Library’s Millennium Integrated Library System from Innovative Interfaces. The aim of the first stage of the project was to build a more extensive framework with the same look and feel of existing interfaces, one that would provide users with structured CAS results tailored specifically for topics of interest in their academic disciplines. In addition, the project participants planned to group retrieved RSS items by type; e.g., blog posts, images, podcasts, etc., similar to the integrated design used in Google’s SearchMash pilot project. Presentation of the content is achieved with a combination of dynamic HTML, Cascading Style Sheets (CSS), and Asynchronous JavaScript and XML (AJAX). The majority of CAS content in this phase of the project was gleaned and integrated into the interface using PHP scripts that enabled the Library to scrape and cache static content, to retrieve and parse RSS data from external content providers, and to tap into third-party APIs.

This project provided an opportunity for the library’s newly hired Programmer/Analyst, Michael Flakus, to experiment with custom search queries using the APIs of major Web service providers (e.g. Google and Yahoo!) and Serials Solutions’ Central Search federated-searching product. The chosen topic of Film Studies, for example, lends itself well to the inclusion of new content, images, and videos using Google, Flickr, and YouTube, respectively, all of which offer associated APIs for developers. The most important exploration, however, was with the Central Search system. The library uses Central Search for its locally named Multi-Search federated searching system, and is designing search filters that will target licensed databases and date ranges most appropriate for the selected topics. Users’ searching results will be presented in a prominent way within
the CAS interface by way of custom programming using the vendor’s newly released XML-based API. The use of Application Programming Interfaces (APIs) to access and retrieve results from third-party sources and pull them into the new application is underway at the time of this writing.

Selection of Resources
During the planning stage, the library envisioned profiling this service by presenting an array of RSS feeds featuring interdisciplinary topics to targeted users (undergraduates, graduate students, and faculty). The CAS content was to have been guided by their RSS selections. However, during development, the library decided instead to use specific feeds with comprehensive coverage and offer less choice over the selection of RSS feeds offered to users. To compensate for the lack of personalization, we decided to instead rely on the librarians’ knowledge/experience as faculty liaisons, instructors and subject specialists. At this juncture we intend to provide this service as a beta product, with the hope of refining and expanding our resources and selectivity options with user feedback.

Selecting Content for Business
Business Source Premier, with its extensive coverage, high percentage of full-text, and the capability to generate RSS feeds for search alerts, was deemed a natural choice for the service. Compendex was also selected because it contains articles on technology management, an area of research interest for some School of Business faculty as well as the subject of some coursework.

The business blogs and Web sites were selected based on possible faculty interest and the prestige/reputation of the sources. For example, HBS Working Knowledge from the Harvard Business School “offers business practitioners a first look at cutting-edge research and thinking from more than 200 HBS faculty.” (Harvard Business School). Knowledge@Wharton is a similar site gathering the latest business insights, information, and research from a variety of sources, including analysis and interviews with industry leaders and faculty (Wharton School). Fast Company, a unique magazine that focuses on “new products, new services, and new ways of doing business,” provides fresh insight into innovative people and ideas (Fast Company).

Selecting Content for the Humanities
The resources selected for Film Studies are a mix of licensed databases and publicly available Web sites, including static and searchable resources. Though publicly available, the Web sites selected are highly regarded and considered very scholarly within the Film Studies community. Though there is strong interest on campus for Film Studies, currently there is no degree offered or even an integrated curriculum for the subject. Film-centric courses are taught by a variety of departments, from Political Science to Communication Studies, with the emphasis usually on a topic (e.g., “Asian Women in Film”) rather than the medium. One exception is the Portland State University English Department, which offers a smattering of Film Studies courses. Because of the lack of a cohesive curriculum, the resources used for this project are purposefully broad, covering a wide array of Film Studies topics.

Challenges Encountered
Versatile applications of Web 2.0 technologies and content hold exciting promise for libraries. However, Huwe argues that best practices for blogs, RSS, and SDI (Selective Dissemination of Information) should “emphasize a personal touch.” The information delivered must be “substantive, on-point” for the user community and the tools must be able to create an opportunity to build interpersonal communication with users over the long term” (Huwe 2006). Although the library is trying to personalize the CAS content as much as possible, it admittedly does wish to reduce (though not eliminate) the need for intermediation. This will lead inevitably to content that cannot be guaranteed to be “substantive, on-point” at all times for the user community. It also means that the CAS will be relying heavily on automated content through RSS feeds and other third-party services that may not always be stable or predictable.

Free utilities, scripts, and tutorials exist on the Web to support amateur mashup creation, but construction typically requires at least an intermediate level of knowledge if not advanced skills in Web development and programming. This is a drawback, obviously, for many libraries with little or no Web programming support. The library recently hired a programmer analyst who was enlisted to work on the project. However his expertise is in the development of Web applications and not with library-related programming.

Because some licensed database providers only support e-mail alerting services and have yet to offer reliable RSS feeds and public APIs, the participants foresee the necessity of including some static, traditional instructions for configuring e-mail alerts for these resources on the CAS site. Since much of the content from libraries
comes from licensed resources, the limited access to automated data streams is an obstacle.

The library’s Integrated Library System (ILS) offers built-in functionality for external RSS aggregation and display, but the library has not yet implemented an “RSS Builder” product that would allow automatic creation of the library’s own RSS feeds. Without this functionality and a direct querying method for bibliographic data via Web programming, the library was forced to create its New Titles RSS feeds via other methods, which padded the development time.

There is also concern in the library literature over relying too heavily upon RSS as a content retrieval/selection tool, due to the inconsistent and repetitive use of keywords, particularly within blog threads. In the article “Are Raw RSS Feeds Suitable for Broad Issue Scanning?” the authors warn that “data cleansing” (which for most libraries means intermediation) is necessary for efficient broad issue scanning (Thelwall, Prawbowo, and Fairclough 2006). Though issue analysis is important to our constituents, the broader scope of academic research is paramount. However, the drawbacks discussed in the article should be a warning not to rely solely on RSS to provide current content at this time. It seems at the time of this writing, complete removal of human intervention from CAS is an impossibility.

Campus Virtual Environments and Future Steps
Implementation of the final product includes efforts to integrate the “Topic Watch” service into the campus’s portal and learning management systems, as well as the library’s electronic course reserve system.

The library maintains its own tabbed section within the myPSU portal, which is based on SCT’s Luminis platform and scheduled to be released for beta-testing in early 2007. Library content has been built into the new portal in the form of multiple channels; e.g., My Library Accounts, Research Basics, My Research Databases, News and Events, etc. Plans are to incorporate the CAS elements as optional channels to which users in different academic programs may subscribe and arrange on their customized portal pages. Channels may be populated using RSS feeds, including custom-bundled RSS feeds, another factor to explore as the project progresses.

Some library content, such as persistent links to full-text articles and links to virtual reference services, is now delivered to users enrolled in online courses via WebCT and Sakai as simple links leading to the library’s Web site, but the inclusion of customized library services and resources, such as the new CAS content, offer viable opportunities to market the value-added benefits academic libraries can bring to users and to build the case for more granular, programmatic integration in the future.

Conclusion
The concept of using Web 2.0 mashups to provide CAS content is promising, especially for libraries that have found the costs and efforts to maintain these services too prohibitive in the past. By synthesizing and displaying RSS feeds from selected sources and using the APIs from major services to present structured search results, libraries can automate major portions of the CAS content rather than tend to this manually. The mashups will benefit users, the authors believe, because users will no longer have to subscribe to multiple feeds, perform CAS-related searches in separate interfaces, and receive numerous e-mail alerts. Instead, they will be able to use the CAS mashups to retrieve search results about their topics and view librarian-selected content from feeds all in one convenient place. There are certainly pitfalls that may be encountered while trying to build and deploy the CAS mashups—not the least of which is the need to rely on external sources to provide stable feeds. But as long as libraries expect to use mashups to reduce, not eliminate, librarian intermediation, the inherent risks may be worth the gains for this type of complementary service.

References


Library Mashups for the Virtual Campus: Using Web 2.0 Tools to Create a New Current Awareness Service