



Smart Libraries

Formerly Library Systems Newsletter™

50 East Huron Street, Chicago, Illinois 60611-2795, USA

November 2006 Volume XXVI Number 11

ISSN 1541-8820

Friendlier Information Finding in Libraries

Part of **The ILS Scoop** by Marshall Breeding

Acclimated to search systems elsewhere on the Web, library users, more and more, may find their libraries offering more powerful tools for finding information and accessing collections of electronic resources and content.

Serials Solutions has added a new level of sophistication to its Central Search federated-search offering. Last month, the company integrated a results-clustering feature; the new functionality, licensed from Vivísimo, transforms a simple listing of results into an interface that offers labeled groupings, which allows the user to quickly hone in on items of interest.

The Cluster Cadre

One of the downfalls of a federated-search environment involves the large sets of results returned by the search targets. Although it's great to be able to retrieve results from multiple information resources via a single search, delivering results in a single, long list can be unwieldy for the searcher to navigate. Even with duplication deleting and sorting, the results can be difficult to interpret and make the process of pinpointing interesting and relevant results tedious. Thus, organizing the results in clusters of related items can enable users to more easily identify and view the results best associated with their research topics.

The clustering approach developed by Vivísimo has similarities to faceted navigation, which is beginning to gain interest in libraries. Clustering is built on automatically created groupings based on textual simulations in the results. Faceted navigation works on the basis of categories assigned within the metadata records, and facets can usually be implemented hierarchically so users can click through multiple facets to incrementally narrow the results to a manageable set.

Whether accomplished through clustering, hierarchical facets, or other approaches, delivering a better, friendlier search environment for library users is a hot topic these days in the library field.

continued on next page

IN THIS ISSUE

The **ILS Scoop** by Marshall Breeding

Friendlier Information Finding in Libraries
PAGE 1

Medialab Solutions Offers
AquaBrowser Online
PAGE 4

NISO Issues Best Practices for Web Services
in Libraries Report
PAGE 5

Audio-Read Expands into U.K. Market
PAGE 6

Google Corner(ed) by Tom Peters

Google Infiltrates Wisconsin, Spain, India,
and News Archives
PAGE 7



Receive *Smart Libraries* via e-mail

Subscribers that would like an e-mailed version of the newsletter each month should forward one e-mail address and all of the mailing label information printed on page 8 of the newsletter to jfoley@ala.org. Type "e-mail my Smart Libraries" into the subject line. In addition to your monthly printed newsletter, you will receive an electronic copy via e-mail (to one address per paid subscription) at no extra charge each month.

THE ILS SCOOP

BY MARSHALL BREEDING

Vivísimo's Verve: Search

Vivísimo is one of the leading companies in search technology. Its flagship product is the Velocity Search Platform, which consists of three layered components: the Velocity Search Engine, the Velocity Content Integrator, and the Velocity Clustering Engine. The company's technologies power the expansive FirstGov.gov Web site and are widely used by corporations, government agencies, and research organizations.

Based in Pittsburgh, Vivísimo, the company, began as a group of scientists formerly associated with Carnegie Mellon University. The firm was formally established in 2000 and operates with

funding from the National Science Foundation, Innovation Works (a Pittsburgh-based venture capital firm specializing in technology start-up companies), and other private investors.

The Vivísimo Clustering Engine performs an analysis of search results as they are delivered from the remote targets to organize them into groups and determine appropriate text labels for each group. The clustering algorithms work only with the information provided in the results and do not rely on manually created taxonomies or controlled vocabulary lists. This clustering feature, powered by the Vivísimo Clustering Engine, will be provided to

Central Search subscribers without additional cost.

Users can get a feel for how Vivísimo's clustering technology works by trying the Clusty search engine at <http://clusty.com>. Clusty applies clustering to general Web searches, and it provides users with an opportunity to see this search approach in action.

Fast Forwarding Federated Search

Vivísimo's Clustering technology enhances Central Search and provides a rapid path to advancement for Serials Solutions, a relative newcomer in the

The screenshot shows a Clusty search interface. At the top, the search query is "ALA TechSource". Below the search bar, there are navigation links for "web", "news", "images", "wikipedia", "blogs", "jobs", and "more". The main content area displays "Top 103 results of at least 15,000 retrieved for the query ALA TechSource". A sidebar on the left lists "clusters" such as "Library 2.0 (12)", "Tame, Libraries and Technology (10)", "Reports, Library Technology (8)", "Opacs (7)", "Flickr (6)", "Technorati (6)", "Rights (6)", "Shifted, Librarian (6)", "Free Range Librarian (6)", and "Blog Archive (6)". The main results list includes:

- ALA TechSource | Home**: Web 2.0 and Libraries: Best Practices for Social Software...
- Flickr: Photos from ALA TechSource**: Flickr is almost certainly the best online photo management and sharing application...
- Tame the Web: Libraries and Technology: ALA TechSource Blog Archives**: Tame the Web: Libraries and Technology This is the old version of Tame the Web by Michael Stephens...
- Tame The Web: Libraries and Technology**: ... ALA's national leadership must help effect that change, ...
- LibraryCrunch: ALA TechSource Discussion with Michael Stephens**: Michael Stephens IMed me the other day and asked if we could have a brief discussion about Library 2.0...
- Case Study: Re:Members revitalizes ALA TechSource | Gulo Solutions**: Gulo - Simple and effective software for organizations of all sizes since 2002...
- Library 2.0 - Wikipedia, the free encyclopedia**: A Library 2.0 Conversation with Michael Casey, ALA TechSource Blog, December 2005...
- ALA TechSource: Bookstore**: About the Bookstore: ALA TechSource provides practical information about the evolving technologies...

A Clusty (<http://clusty.com>) search on the words "ALA TechSource" yields clustered results (see box on left side of screen shot). Image courtesy of Vivísimo.

federated-search domain. The product was debuted at the January 2005 ALA Midwinter Meeting in Boston—while some other companies, such as WebFeat, MuseGlobal, and Ex Libris, have been developing and marketing federated-search products much longer.

Serials Solutions created Central Search through its own development efforts, though some components were licensed from other companies. Besides the recently integrated Vivísimo component, the system uses translation technology licensed from WebFeat.

Other companies involved in providing information resources and technologies to libraries have also taken interest in the Vivísimo technology. In January 2006, Swets announced a new federated-search product, the SwetsWise Searcher, based on metasearch and clustering technology from Vivísimo. In June of this year, Ex Libris announced it had licensed the Vivísimo Clustering Engine and reported plans to integrate this technology into the upcoming Version 4 of the company's MetaLib federated-search product.

Central Search complements other applications offered by Serials Solutions, providing library customers with an integrated suite for the management of and access to collections of electronic resources. Other components include AMS (Access & Management Suite), Article Linker, and ERMS (Electronic Resource Management System).

All these products build on the vendor's foundation product, a comprehensive database for e-journal titles

and holdings. Mastery of holdings data has been the core competency of Serials Solutions from its inception, and in the arena of e-content management, the quality of the holdings data distinguishes a product as much as the capability of the software.

Central Search, like the other members of the Serials Solutions suite of products, follows the vendor-hosted model. Previously called ASP (Application Service Provider), this approach is now referred to as "Software as Service," and it represents one of the major information-technology trends.

Software as Service involves accessing a product on a server managed by the vendor, saving the organization from expense and complexities involved in installing, configuring, and maintaining the software, operating system, and hardware associated with a given technology product. Given the complexity of establishing connection configurations with each of the search targets, the hosted approach is especially attractive in federated-search offerings. The disadvantage lies in more limited ability to customize the application and integrate results into local portal environments, but Serials Solutions offers an XML API (application programming interface) to address this disadvantage.

Among the federated-search products marketed to libraries, some follow the Software as Service model and others rely on local installations. WebFeat, like Serials Solutions, offers its federated-search product in the Software as Service model. MetaLib (from Ex Libris) and

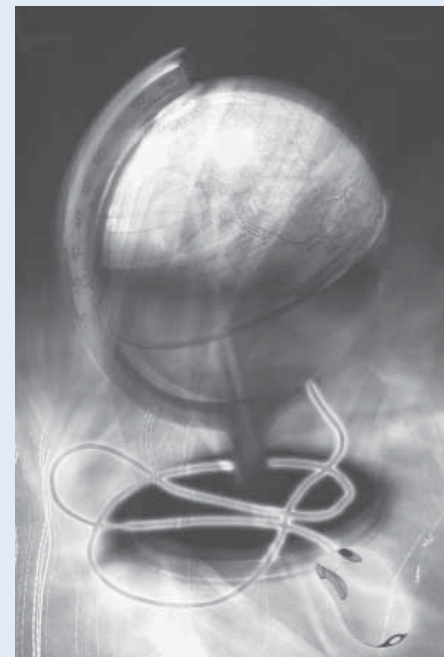
MuseGlobal are offered to libraries as locally installed applications.

Additionally, in an interesting turn, Frank Bilotto, formerly head of publishing at Vivísimo has joined MuseGlobal as its new VP for Publishing and Digital Media. ■

More Info. @:

"Clustering versus Faceted Categories for Information Exploration," by Marti A. Hearst, *Communications of the ACM* 49:4 (April 2006): 59–61.

Central Search Results Clustering, www.serialssolutions.com/promotion/clustering.asp
Vivísimo, <http://vivísimo.com>



Medialab Solutions Offers AquaBrowser Online

In tune with the Software as Service trend (mentioned in the cover story of this issue), Medialab Solutions reports it will offer AquaBrowser Library as a hosted solution and will market the new service under *AquaBrowser Online*. This version of the product will offer the same features as the locally installed AquaBrowser Library, but it will reside on servers managed by the vendor.

AquaBrowser Online targets libraries with collections under 150,000 titles, and it provides an alternative interface for a library-automation system, offering a visual cloud of associations and other navigation features. The operation of AquaBrowser involves regular extraction of records from the library's integrated library system (ILS) into the search platform. For smaller libraries, this transfer of bibliographic records to remote servers can be easily accomplished. AquaBrowser Library will continue as the preferred approach for libraries with collections exceeding 150,000 titles.

Medialab Solutions partners with The Library Corporation (TLC) to distribute AquaBrowser in North America. TLC has licensed the product to more than 100 libraries. Although the majority of TLC's AquaBrowser customers run one of its own ILS products, the company has also achieved some success in marketing it to libraries running an ILS provided by its competitors. In the Netherlands, Medialab Solutions reports more than eighty percent of public libraries use AquaBrowser Library. ■

More Info. @:

AquaBrowser Online, www.aquabrowseronline.com



NISO Issues Best Practices for Web Services in Libraries Report

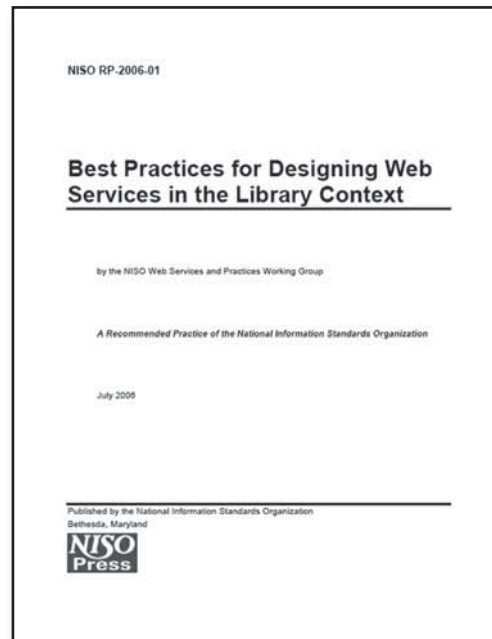
In July NISO, the National Information Standards Organization, released a brief, informative report (RP-2006-01) on *Best Practices for Designing Web Services in the Library Context*. It is worth perusing.

The report, developed by the NISO Web Services and Practices Working Group, which includes members from both libraries and vendors, defines a Web service as “. . .an alternative to fully developed application programming interfaces (API) for circumstances in which the additional overhead is not warranted.” The goal of these best practice guidelines is to support interoperable digital library services.

The report outlines the tradeoffs between a simple Web service interface and more complex protocols. It also describes several different models for describing or documenting a Web service interface. The information model articulates the types and formats of information to be exchanged. Information models describe both the structure (syntax) and meaning (semantics) of the information being shared between information systems. The behavior model focuses on system actions and responses as well as the order of and dependencies between actions taken by an information service. The action model describes the actions possible within the information service as well as the broader implicit effects of the possible actions. The process model emphasized the required time-order of events and actions within an information system.

The report also contains recommendations and best practices for caching Web-based information as well on how to re-use output formats. The strengths and weaknesses of various document output formats, such as DTD (Document Type Definition), XML (Extensible Markup Language) Schema, RDF (Resource Description Framework) Schema, Relax NG (Regular Language for XML Next Generation), and DSD (Document Structure Description), also are enumerated.

Appendix A contains a useful categorization (with examples) of Web services. Discovery services, for example, are used to discover metadata, full objects, or services. Locate services are used to “resolve an object to its location, either physically or in a process.” Requesting services are used primarily to request information and information objects. Delivery services facilitate the “delivery of objects, information, formatting, transactional information, etc.” What the working group calls



The PDF of *Best Practices for Designing Web Services in the Library Context*, published by NISO Press in July 2006, is available at www.niso.org/standards/resources/rp-2006-01.pdf.

“common services” are services that focus on administrative, financial, systems integration, and other back-office activities.—*Tom Peters*

More Info. @:

Best Practices for Designing Web Services in the Library Context, www.niso.org/standards/resources/rp-2006-01.pdf

“Web Services and the Service-Oriented Architecture,” *Library Technology Reports* 42:3 (May/June 2006), by Marshall Breeding, www.techsource.ala.org/ltr/web-services-and-the-service-oriented-architecture.html

Audio-Read Expands into U.K. Market

The burgeoning field of digital audiobook services for institutional customers (such as libraries, corporations, and governmental units, including prison systems and branches of the military) has many interesting facets. Although digital audiobooks can be listened to on one's "mothership" computer—and burning to CD remains a flickering, if time-consuming, option with some digital audiobook services—listening to audiobooks on portable audio playback devices, commonly called MP3 players, is the dominant mode for interacting with digital audiobooks.

The problem with audio playback devices is they are designed primarily for teenyboppers who want to listen to music pretty much all of their waking hours. The devices were not designed for listening to and navigating through audiobooks, and they were not designed to be eminently accessible to other segments of the population, such as older adults and people who are blind, have low-vision, or with mobility issues involving their wrists, hands, and fingers. For many people beyond their teenage years, using today's "mainstream" MP3 players can be anything from a minor to a major annoyance.

Niche device manufacturers, such as Humanware, offer specially designed portable audio playback devices that are easier to use and more accessible for blind and low-vision users. The problem is price. Compared to the low and declining prices for mainstream playback devices, the prices for specialty playback devices can seem outrageous.

An all-too-common phenomenon in the information-technology and telecommunication fields is that some of the more interesting and innovative initiatives emanate from countries other than the United States. When it comes to digital-audiobook services for institutional customers, one small company I try to keep tabs on is Audio-Read, based in Ultimo, a neighborhood in Sydney, Australia.

Several years ago Audio-Read developed the Audio Navigator portable playback device. Although not as slim and sexy as the teenybopper MP3 players that sell like hotcakes in the U.S., the Audio Navigator is one device that is joyfully useful to just about everyone. It even has its own built-in speaker, in case prolonged use of ear buds makes your ears ring.

In late September, Audio-Read announced plans to expand into the United Kingdom market. The company is targeting its service to public libraries, agencies that service blind and low-vision people, and U.K.-based publishers.

In an e-mail message I received from Anthony Blackwood, the director of Audio-Read, he indicated the company plans to expand into the European Union and the United States soon, so keep your ear to the ground and listen for Audio-Read.—*Tom Peters*

More Info. @:

Audio-Read, www.audio-read.com.au/home.htm
Humanware, www.humanware.ca

Compared to the low and declining prices for mainstream playback devices, the prices for specialty playback devices can seem outrageous.



Google Corner(ed) BY TOM PETERS

Google Infiltrates Wisconsin, Spain, India, and News Archives

The list of libraries officially partnering with Google for the Google Books Library Project continues to grow. In mid-October, the University of Wisconsin joined the project, and approximately a month prior, Google announced that Complutense University of Madrid had joined too.

Complutense has the second largest collection in Spain and the largest academic collection. In addition to obvious strengths in Spanish-language books, the Complutense collections are also strong in other languages.

Complutense joins the Bodleian Library at Oxford University as the two non-U.S. libraries—to date—participating in this massive digitization project. Only books in public domain will be scanned.

On the publishing side of the Google Book Search project, in the last couple of months, Google also announced it had reached agreements with several large Indian publishers to include their titles in Book Search. India is the third largest publisher of English-language books in the world, after the United States and the United Kingdom. Because the overall Indian publishing industry has been growing at the rate of approximately fifteen percent per annum for the past several years, these alliances could prove to be very strategic for Google.

In addition, earlier this fall Google announced and began beta testing—in a public manner, of course—yet another searchable index. This one is a searchable compilation of news archives. Being able to perform a relevancy-ranked search of a large corpus of archived news items is clearly valuable, but Google, of course, pushes the envelope a bit. According to the “About” page on Google News Archive Search, “In addition to searching for the most relevant

articles for their query, users can get an historical overview of the results by browsing an automatically created timeline. Search results include both content that is accessible to all users and content that requires a fee.”

Examples provided by Google of the free content include the archives at BBC News, *Time*, and *The Guardian*. Examples of fee-based content include the *Washington Post* Archives, the Newspaper Archive, and the *New York Times* Archives.

Like many things in Google, a computer algorithm automatically generates the timelines. Sampling the timed articles about popular information-technology developments, such as PDAs and iTunes, can be interesting and informative.

More Info. @:

“University Joins Google Library Project,” http://news.zdnet.com/2100-9588_22-6125226.html

“University of Wisconsin-Madison and Google Make Rich Historical Collection Available Online,” www.google.com/press/ann/uniwisconsin_books.html

“University Complutense of Madrid and Google to Make Hundreds of Thousands of Books Available Online,” www.google.com/intl/en/press/ann/books_madrid.html

Complutense’s Announcement (in Spanish), www.ucm.es/BUCM/biblioteca/11979.php

Google Book Search, <http://books.google.com>

Google News Archive Search, <http://news.google.com/archivesearch>



Smart Libraries Newsletter
American Library Association
50 East Huron Street
Chicago, IL 60611-2795 USA

NON PROFIT
US POSTAGE
PAID
PERMIT 3020
BIRMINGHAM, AL

November 2006

Serials Solutions Integrates Results Clustering in Central Search Offering

Smart Libraries Newsletter

Smart Libraries Newsletter delivers hard data and innovative insights about the world of library technology, every month.

Contributing Editors

Marshall Breeding
615-343-6094
marshall@breeding.com

Tom Peters
816-228-6406
tpeters@tapinformation.com

Editor

Teresa Koltzenburg
800-545-2433, ext. 3252
312-280-3252
tkoltzenburg@ala.org

Administrative Assistant

Judy Foley
800-545-2433, ext. 4272
312-280-4272
jfoley@ala.org

TO SUBSCRIBE

To reserve your subscription, contact the Customer Service Center at **800-545-2433, press 5 for assistance**, or visit **www.techsource.ala.org**.

The 2006 subscription price is just \$85 US.

Production and design by Angela Gwizdala,
American Library Association Production Services.

Smart Libraries Newsletter is published monthly by ALA TechSource, a unit of the publishing dept. of the American Library Association.

Copyright American Library Association 2006. All rights reserved.