WiLS

[Tech Notes are short articles discussing library-related technology.]

[Tech Note:] A Discursive Discussion of Resource Discovery

I am, much to my disappointment, a genuine novice at searching for information. I am not a librarian, so I do not have any formal training to bring to bear. Consequently, I do not have the experience of a reference librarian. Like so many others, my research begins and ends with Google and Wikipedia. Okay, that's somewhat of an exaggeration, although my research usually does *begin* with Google and I do find Wikipedia to be an invaluable resource. Of course, sometimes I move to Google Scholar or Google Books, but I'm still down in the trenches rather than high on the plain.

Resource discovery is very important. If library users aren't able to find what they need it annoys or confuses them or both. What we all seek as our search utopia is a way to ensure that we obtain high degrees of both precision (obtaining *only* relevant information) and recall (obtaining *all* relevant information). This is, of course, impossible for all but the simplest of needs. It is relatively easy to find a book that will tell me how to build a chair, and I only need one relevant book. But if I want to search for Plato's conception of a chair as instance of the form or ideal "chair", things get more complicated. In the former case, precision is easy to achieve and recall is immaterial. In the latter case, finding the appropriate resource is much more complicated. Obviously, precision and recall are important in all libraries, whether academic, public, special or school.

More specifically, what I am concerned with here is *machine-mediated* or *automated* resource discovery, not human-mediated. What I want to be able to do is to go to a computer site, perhaps associated with a particular library, key in some search terms, and obtain all of the relevant results and only those that are relevant. That is not now an attainable outcome, for many reasons. Part of what is missing is what's accomplished in the reference interview; currently, search software cannot ferret out what someone *really* wants from what they *say* they want. Additionally, there is not a unified database containing the world's knowledge or pointers to it, no matter what the ambitions at Google or OCLC.

One reason for the "holes" in searching is that many databases, including those containing or referencing articles from publications, are opaque to index crawlers. In other words, Google or services like Google can't "see" these materials. This is the so-called "Deep Web" or "Dark Web" problem. It is possible for libraries that have subscriptions for some of these databases to unify access to them through federated search. However, federated search tools often do not work very well to bring you the full distance from search to the specific article you want.

Additionally, traditional catalogs have done poorly in incorporating local and worldwide digital collections and archives as well as other local items (e.g., finding aids) into the resource discovery picture. .

In an imperfect information world, what's a library to do? One of the issues is that there are not only questions of retrieving relevant information, but ensuring that the patron is not overwhelmed by too much information. After all, capturing or citing *all* of the information to answer a question or plumb the depths of a subject would present the patron with the proverbial drink from a fire hose. It should be noted that this doesn't seem to bother people when they use Google so long as the first screen of entries holds something pertinent to their needs.

How can a library provide good information to users from both the offline and online worlds given these limitations (and others) to electronic retrieval? The UW-Madison Resource Discovery Exploratory Task Force did a quite good, very thorough job of characterizing the current generation of resource discovery tools (see Executive Summary [1] or the entire report [2]). Unsurprisingly in the current discovery "climate", they recommended "decoupl[ing] the interface from the ILS so that it is sleek, lean and enabled for rapid change." The catalog face of the ILS, regardless of vendor, has been criticized and reviled for years, with libraries having little to no control over its look and feel.

The Task Force also did a very nice job of laying out the desiderata for a leading-edge "discovery environment," including the integration of information held at the campus, e.g., digital collections, library web sites. Customization and personalization are also called for. Delivery of search functionality at off-campus destinations (Amazon, iGoogle, Facebook, WorldCat) is another required feature. Read the entire summary for other desirable features. I don't think all of these are achievable immediately, but they are good targets to shoot for. Equally valuable are the Recommendations of the report, which outline in concrete terms what can be done now at UW-Madison to enhance resource discovery. A number of these recommendations have been implemented.

One difficult problem in resource discovery is the integration of the library catalog with the various subscription databases. As I've said, I want a single, unified interface to *all* of the resources of the library. I want it all, and I want it now. And that's what other patrons want, too. But it is difficult to achieve, given the inability of many of the databases to "play nice" by accepting queries in standardized formats and return fine-grained information in a standard form. Each database is different, even when it does attempt to conform to standards such as query language standards and openURL.

Some products and projects have attempted to address some or all of these issues. VuFind [3] is becoming an important resource for several libraries (I count 19 in various stages of testing and implementation at the site), allowing records from a number of sources ("catalog records, locally cached journals, digital library items, institutional repository, institutional bibliography, other library collections and resources" [3]) to be served up via a single interface. To see VuFind in action, take a look at MnPALS

Plus[4], which is based on VuFind and extended by PALS staff. (PALS is a Minnesota library consortium.) VuFind is open source. VuFind scales up nicely: there are a number of institutions with 1+ million records. Given its development at a large university (Villanova), it may not be suitable for public libraries, and in fact all of its users so far are academic institutions. It recently won a Mellon Award for Technology Collaboration. I think it's the most important open source product to look at for resource discovery, so long as serious development of it continues.

Blacklight is another open-source project [5]. Blacklight is billed as an open source OPAC, although the banner on its opening page promises "findability for your whole collection", which seems to me to be a broader claim. The project has been active for about a year, and you can try out the one example system [6]; I don't know how much traction Blacklight has attained outside its home at the University of Virginia library. It can handle MARC records as well as several flavors of XML (eXtensible Markup Language), which adds versatility to search beyond the core library collection. Documentation is abysmal, which is not unusual for many open source projects.

For both VuFind and Blacklight it is necessary to have programming staff available who can install and customize the out-of-the-box code. This, of course, makes them non-starters for libraries with limited budgets.

Endeca is perhaps the best-known commercial package because of its adoption by North Carolina State University (NCSU) several years ago and the extensive publicity surrounding that adoption. It is an enterprise information tool that was not originally nor now targeted at the library marketplace, so I'm sure its implementation at NCSU was a learning experience for all concerned. It's interesting to note that libraries are not mentioned as an Endeca market either on the opening page of its website or its Solutions page [7]. It continues to find library users, although it is not the only commercial system available.

Primo [8] is one of those other systems, a product of Ex Libris. It offers access through a single search to both local and remote resources. It has over 150 users worldwide, which may make it the most-adopted of all of the discovery products, commercial or open source. It has the advantage that it is capable of being coupled to SFX, a link server in wide use. Additionally, as a commercial product there is less need for technical staff involvement in implementation.

I would be remiss if I did not give at least a passing nod to WorldCat Local [9] from OCLC. It goes beyond displaying local content to "see" all of the WorldCat world for which the library has access. Additionally, it can interface with Ex Libris for even wider access to documents. It is worth considering as a catalog replacement if you are willing to allow your patrons to search beyond the local library. To see it at work, take a look at the University of Washington Libraries [10].

One of the pluses of all of these products is that they start at the library, unlike a Google search. Thus, libraries preserve their "brand identity" when users are able to search for

what they need starting with the library's tools. People have very positive opinions of libraries, but they tend to gravitate toward search engines, particularly Google, partly because they find current library catalogs unintuitive and generally difficult to use. The resource discovery packages provide easier interfaces for users and more features than current catalogs offer. Getting people to start with the library can be a problem, but providing useful and easy to use tools may draw users away from the reflexive resort to search engines.

Takeaways From This Article

- Resource Discovery is a basic function of any library; only some libraries will be able to afford an automated Resource Discovery System.
- Resource Discovery software can supplement or replace the library's traditional online catalog.
- There are a number of Resource Discovery products available, both commercial and open source. Several of these are not mature systems.
- All of the automated Resource Discovery tools require setup time and the open source products may require additional programming.
- It is important for all types of libraries to consider Resource Discovery systems, if they have the budget to support such a system.

This note was considerably informed by the Resource Discovery Exploratory Task Force Final Report [2]. *It is well worth a close reading by anyone who wants to think deeply about automated Resource Discovery.*

[1] Resource Discovery at UW Libraries: Executive Summary (http://uwlibdiscovery.blogspot.com/2008_06_01_archive.html).
[2] Resource Discovery Exploratory Task Force Final Report (http://staff.library.wisc.edu/rdetf/RDETF-final-report.pdf).
[3] VuFind web site (http://www.vufind.org/index.php).
[4] MnPALS (http://plus.mnpals.net/).
[5] Blacklight (http://blacklight.rubyforge.org/).
[6] Blacklight example (http://blacklight.betech.virginia.edu/)
[7] Opening page (http://www.endeca.com) and Solutions page (http://www.endeca.com/solutions).
[8] Primo Overview (http://www.exlibrisgroup.com/category/PrimoOverview).
[9] WorldCat Local (http://www.oclc.org/worldcatlocal/default.htm).
[10] University Libraries, University of Washington (http://www.lib.washington.edu/).

(All links tested 3/16/09)

Many thanks, as always, to Jane Richard for her invaluable editing skills. Any errors or stylistic blunders which remain are, of course, mine.

— Tom Zillner

Other Tech Notes