Editorial Board Thoughts:
Services and User Context in the Era
of Webscale Discovery

Mark Dehmlow

Early implementations of discovery systems were revolutionary in their ability to bring together metadata from masses of scholarly resources into super indexes but they were also somewhat unsophisticated in that their initial focus was primarily on the aggregation of data. The data is critical, but the debates over vendors' unwillingness to share data have overshadowed other important features and functionality that are important to making discovery systems valuable, especially in the areas of getting users the actual resources that they find, delivering the most relevant results, and being more aware of the full user context not merely their search query. The user context is a multidimensional matrix encompassing: (1) level of experience, (2) comprehensiveness of their research need, (3) type of scholarly materials the user primarily works with, (4) user discipline, and also (5) what physical or virtual location the user is performing their research from.

Missing Services and Opportunities

A major issue that continues to confound me is the lack of fully integrated request and delivery services that many discovery systems lack. Of course, all of them implement full text linking to every online article that they can create a link to, but as the sphere of scholarly data stretches beyond just articles, library print collections and delivery services have continued to be neglected primarily because implementing those services in an intuitively integrated way, beyond the "link to your old OPAC" methodology, remains a complex task. My main concern with this deficit is that there is a significant amount of scholarly material only available in print and to focus primarily on electronic access limits the ability of our users to perform comprehensive research and reduces access to significant resources and services that libraries provide. During the transition from print to online, we need to consider user behaviors and preferences while eBook technology becomes more user-friendly, especially for academic materials which are often difficult to access and use because of clunky digital rights management and distribution models.

The most recent Ithaka study on faculty research behaviors noted that 80% of faculty respondents still find it much or somewhat easier to read an entire book in print versus in electronic format.[1] This will change over time as the technology evolves, but it will be quite some time before all written works are available online and researchers often needs access to the long tail of unique resources that support very narrow and niche knowledge areas.

Mark Dehmlow (mdehmlow@nd.edu), a member of LITA and the ITAL editorial board, is the Director, Information Technology Program, Hesburgh Libraries, University of Notre Dame, South Bend, Indiana.

Some of the discovery systems have integrated the DLF recommended interoperability standards seamlessly[2] into their interfaces, but some don't provide much more than a link back to the libraries' traditional OPAC interface, often in a completely different user interface that requires the user to make a cognitive shift. And even those vendors who do integrate basic ILS functionality don't intuitively integrate other important library services such as localized delivery or interlibrary loan services. For many of our users, particularly faculty and graduate students, their goal is to exhaustively evaluate every useful resource in the pursuit of their research, not only the easiest and quickest to access.

In addition to these deficits also lies untapped opportunity. The aggregation of so much bibliographic data in these systems could allow for providers to build more intuitive connections between primary resources and bibliographic records that relate to those resources. One of our librarians recently commented that it would be really valuable if we could connect all of the book reviews in the webscale index to the actual bibliographic record for the book - placing the related information for the book at the primary point of access for the user. Another librarian commented on how valuable it would be to integrate user generated comments about books from worldcat.org in a similar way. I believe one of the next major opportunities for discovery systems is to begin building holistic connections between the objects in their indexes, perhaps using linked data techniques.

Relevancy Refactored

Determining actual relevancy of an item to a user query is a much more nuanced business than standard term frequency — inverse document frequency (td-idf). Even when used in concert with field weighting for metadata like author and title, this type of algorithm has mixed effectiveness because while math can easily indicate how often a series of words occur in a document, it requires considering the whole user context to best derive what a user is looking for. Perhaps one of the more interesting developments in relevancy enhancements is the incorporation of social and usage analytics. This more intricate adjustment to relevancy calculation is what makes Google so powerful — they can take how often a website is referenced in other websites and how often users go to a website to impact a website's relevancy to a user query. Ex Libris is currently using similar strategies with something they call Scholar Rank. They enhance their document relevancy through using their bX recommender service (a "users who downloaded this article also down loaded these articles" – style service that is based on link resolver log analysis) and data like impact factor.[3]

The physical and virtual location of a user provides another unique context that doesn't get much attention. Recently, we made our default search box the library catalog for searches performed in our stacks, the idea being that a user is most likely looking for things in our collection if they are in the tower. They can easily jump to the webscale search if they like by clicking an adjacent tab. Another specialized context is for our users are our subject pages, which have lists of library resources that relate to specific disciplinary areas. Where possible, each of these pages features a

subject quicksearch box which currently utilizes metasearch to provide discipline scoped results. We would like to eventually replace these popular metasearches with the benefits of a webscale search, but scoped by. Serials Solutions' Summon is able to provide through its "discipline-scoped search" feature.[4]

CONCLUSION

Libraries are still struggling with how to market and position these massive indexes in the context of their websites and the other services they provide. In part, discovery systems induce discomfort for library professionals because the ways in which relevancy is calculated are becoming more opaque. In the face of all this ambiguity, I still believe that these systems are incredibly useful and valuable for support of different research needs at different levels – getting novice users quick access to a handful of scholarly resources and augmenting the deep research process of our expert users. Many of these systems on the market are evolving into more complex search tools, but there is a lot of room for these systems to grow as well. If you were to take the best of each of the systems and put them together, you might actually be on your way to having a complete solution. For libraries as we continue to license products and implement these solutions, it is important to develop a more complex understanding of what our users need and then ensure our discovery systems will translate their simple keyword query into results they are actually looking for and then help them easily request the materials that turn up in those results.

REFERENCES

[1.] Ross Housewright, Roger C. Schonfeld, and Kate Wulfson, *Ithaka S+R US Faculty Survey 2012*. (Ithaka S+R, 2013), Pg. 32. http://sr.ithaka.org/sites/all/modules/contrib/pubdlcnt/pubdlcnt.php?file=http://sr.ithaka.org/sites/default/files/reports/Ithaka_SR_US_Faculty_Survey_2012_FINAL.pdf&nid=502 (accessed June 3, 2013).

[2.] John Mark Ockerbloom et al., *DLF ILS Disovery Interface Task Force (ILS-DI) Technical Recommendation: An API for achieving effective interoperation between traditional integrated library systems and external discovery applications.* https://project.library.upenn.edu/confluence/download/attachments/5963787/DLF_ILS_Discovery-April08_draft.pdf (accessed June 3, 2013).

[3.] "Primo Scholar Rank plain and simple," Youtube.com http://www.youtube.com/watch?v=YDly9qPpPYQ (accessed June 3, 2013).

[4.] Summon: Features & Functionality," Serialssolutions.com http://www.serialssolutions.com/en/services/summon/features-functionality/search (accessed June 3, 2013).