To Google or Not to Google: Metasearch Design in the Quest for the Ideal User Experience

Abstract

The goal of metasearch systems is to facilitate access for researchers—both novice and expert users—to the ever-growing number of scholarly electronic resources. Researchers today are typically Web-savvy and have high expectations regarding ease of access to information for their research needs. Facing the variety and complexity of the interfaces provided by information providers, researchers often prefer to use Google or other Web search engines. Libraries need to address the expectations of today's users, encouraging them to take full advantage of the scholarly information resources in which libraries invest so heavily.

From a user's point of view, metasearching has numerous benefits over other methods of conducting research. Metasearch workflows can be tailored to different types of users and can provide two-fold access to information—via the metasearch interface and via the native interface of the specific information provider.

The presentation of results is a major issue in searching and retrieval. According to Roy Tennant, of the California Digital Library, "the size of a result set isn't as important as how it is presented"—what he calls the "Google lesson." Metasearch systems employ a variety of methods to display results to end users.

Case studies on the effects of metasearch systems that were introduced into academic environments have yielded valuable information about the change in patterns of electronic resource usage on campus.

The paper concludes with some ideas on future directions that metasearch systems might take.

Introduction

Anne is an undergraduate student at a European university. She is taking a course on cognitive therapy and needs to submit her final paper, on the use of dreams in psychotherapy. Unlike many of her friends, who lack the patience to explore the school's library collections and instead use Google[1] to find information, Anne is motivated to produce an excellent paper, so she sets out on a search for quality resources.

The university's library Web site directs Anne to the library OPAC and to a list of databases, displayed alphabetically. Anne starts her search in the library OPAC,

[1] *How students search: information seeking and electronic resource use*, a study carried out by the JISC-funded formative evaluation of the Distributed National Electronic Resource project (EDNER, 2002), found that 45% of students used Google as their starting point to find information; 9% used Yahoo; 6% used Lycos; and 8% used other search engines. Only 10% of the students selected the university OPAC as their first resource for locating information.

where she locates a book called *Dreams as a Tool in Psychodynamic Psychotherapy,* by Vincenzo Conigliaro. She recalls that she is also entitled to search in a union catalog and request books that do not reside in her library. She links to the union catalog and searches again, using a different interface. This time the search takes longer, as she is not experienced with the interface, but she finally finds two other books that look particularly relevant: *Cognitive Therapy and Dreams*, by Rachael Rosner, et al., and *Working with Dreams in Psychotherapy*, by Clara Hill. She writes down the titles so that she can take them to the librarian for help in obtaining the books.

Anne would also like to find some recent articles on her topic. She locates the library's database page and wonders which of the several hundred databases are relevant. Checking them one by one is out of the question, so she goes to the letter *P* and conducts searches in a few of the databases listed there: EBSCO's *PsycARTICLES*® and Psychology and Behavioral Sciences Collection™, OCLC's PsycFIRST®, and SilverPlatter's PsycINFO®. This process takes hours. After selecting several papers, she concludes that not all of them are available online, since no links are displayed. By the end of this long, frustrating day, Anne has managed to find only one book and four papers, with two more books and several papers noted on her list.

In the meantime, Anne's friend Robert used the first few hits from Google and managed to write his paper. It may not be comprehensive or brilliant, but it's good enough for him…

Both Anne and Robert would have obtained better results—and Anne would have had a better experience—if their institution had a library portal like the one at the university where Anne's friend Michael is studying. The library portal provides Michael with various functions related to the research process, including directing him to the appropriate resources. But the main benefit for Michael, no doubt, is the metasearch function: using one interface, he can simultaneously search a variety of relevant resources, including the library OPAC, union catalogs, reference databases, and even Google, and the system displays the results in a uniform manner. Furthermore, the results are merged into one list, although Michael can choose to view the results by resource. The system also offers related services, such as the full text of an article and, if the article is not available online, a document-delivery service. Users have various options for managing the results—for example, saving them or sending them by e-mail—and can also access the resource directly in the native interface of the information provider.

Michael's research skills do not differ from those of Anne or Robert. However, Michael is lucky to have better tools for conducting his research and, indeed, takes advantage of them. As recent case studies have demonstrated, the introduction of such tools significantly increases the use of the electronic collections in which libraries invest so heavily.

What Is Metasearch?

When a user submits a query to a metasearch system, the system broadcasts the query simultaneously to heterogeneous information resources residing in numerous locations. These resources can differ in many ways—for example, in the type of information that they contain and the format in which they present information. Each information resource, using its own search engine, performs a search and

transmits the results to the metasearch system, which, in turn, displays the results to the user in a single interface.

Because each information resource has its own search engine, the queries that it receives must conform to that engine's requirements. Even if the resource supports a standard interface, such as the Z39.50 protocol, the metasearch system must make some adjustments so that the resource's engine will interpret the query correctly.

The detailed technical issues involved in searching and retrieval are discussed elsewhere (Sadeh, 2004). This paper focuses on searching and retrieval through the metasearch functionality—also called cross-database searching, federated searching, and a variety of other names—from the point of view of the end user.

The Google Spell

It seems that *Google* is the name most frequently found in every paper that discusses searching options, particularly library searches. Although earlier search engines effected a great change in the information research environment, Google is the one that shattered all previous searching patterns. Most of us use Google when we plan our vacation, when we wish to buy a new camera, and when we need to find a government Web site. However, when it comes to the scholarly environment, Google is not an appropriate research tool, primarily because it does not differentiate between quality resources and other resources. Google's PageRank algorithm always presents the most popular Web sites at the top of the result list. This algorithm is fine in a nonscholarly environment, but those who deal with academic research treat it with caution. Also limited in the scope of its searches, Google typically disregards the deep Web, particularly licensed repositories[2].

We can learn several lessons from the success of Google, as Roy Tennant, of the California Digital Library, points out in this list of principles:

"• Only librarians like to *search*, everyone else likes to *find*
- All things being equal, one place to search is better than more
- 'Good enough' is the sum of gain minus pain; users aren't lazy, they're *human*
- The size of a result set isn't as important as how it is presented ('the Google lesson')" (Tennant, 2004).

So why is Google so appealing to end users? First, it is simple. Second, it usually provides excellent results; and third, you typically find what you are looking for on the first page. It has other benefits, too, such as amazing speed and fuzzy searching, a feature that takes into consideration, for example, the differences between English and American spelling, and other types of spelling variations. In short, Google addresses most users' needs perfectly.

[2] Publishers have now begun to allow Google to search their repositories. An example is the CrossRef Search pilot project, which already covers nine publishers and permits free full-text, interpublisher searching of scholarly research content. This service launches a typical Google search but filters the result set to the content from participating publishers (CrossRef, 2003).

Our library users have grown up in this world of the Internet, of quick and focused response, of immediate satisfaction. Library services should accommodate such users rather than fight them. And above all, librarians need to maintain their role as information experts and make sure that the tools that they provide will direct the users—perhaps in a new way—to the right places. Researchers need to have quality information resources readily available to them.

Designing a Metasearch Interface

Metasearch systems are relatively new products, whose user interface poses various challenges, particularly if designers keep Google in mind as both a model and a competitor. The development of the interface in the new version of MetaLib, the Ex Libris library portal, is a good example of the design process. Numerous factors contributed to the design of the interface:

- The experience gained from working with staff at hundreds of sites where previous versions of MetaLib are in use
- Results of case studies conducted by institutions to obtain feedback from their users
- Discussions with a focus group consisting of representatives of ten institutions from around the world; the representatives were selected by the SFX and MetaLib user group (SMUG)
- The work of user-interface experts

One of the main goals of the new MetaLib user interface design was to enable the system to serve various communities of users. It is obvious that undergraduates have different needs from those of advanced researchers, but all need to feel comfortable with the system. Each user should be able to identify the workflow and search-and-display functions that best suit his or her needs without being hindered by irrelevant information and functions. The system should expose users information and functionality when necessary—and if users desire such information or functionality, they should be able to obtain it with ease.

The Search Flow

The workflow for a user of a metasearch system typically consists of several steps:

- Entering a query and indicating the scope of the search
- Viewing a summary of the search results (the total number of hits and perhaps also the number of hits per resource)
- Viewing a summary of the record for each result
- Viewing the full record for a specific result
- Selecting additional options, such as a link to the full text of an article or to the user's OPAC for ordering a book

Although the basic functions of a metasearch system are similar for all users, advanced users need more flexibility and can take advantage of extended options for obtaining results. Also, the system can display the results in different ways.

The first question that the user-interface designer asks is how the users will indicate their search scope. It is obvious that novice users need more help in determining the scope of their search, as they may not be familiar with the resources offered by the library. In many cases, they are not even interested in knowing; just as they trust Google to search in the "right" places, they trust the

metasearch system to examine the appropriate resources. For such users, the system needs to determine the scope based on the user affiliation or to offer the user some alternatives described in general terms (for example, "Cognitive Psychology Course CP014" or "resources for medical students"). Of course, the system also needs to accommodate users who want to look "inside" and find out which resources make up the selected scope.

Advanced users, on the other hand, need to have full control of the research scope. The system should enable them to locate resources through various attributes, such as name, vendor, subject, and language, and to use ad hoc lists of resources or create long-term lists, which they keep as a part of their profile and modify as necessary. Each user should be able to define multiple lists and use them as needed.

Some metasearch systems do not limit the number of resources that can be searched simultaneously, and some even provide an option to search everywhere. This lack of restrictions seems problematic. First, there is a limit to the amount of data that a user can process. A large number of hits does not necessarily help the user. Second, squandering the use of information resources for searching has an unfavorable effect on both the access permissions available to institutions and the load on the information providers' systems. Concurrent-usage licenses, if employed, are exhausted rapidly and unnecessarily, and the resource may not be available to users who require access. Information providers, on the other hand, are concerned by an increase in the search requests and the load on their systems in unnecessarily opening multiple sessions. Hence, limiting the number of resources in which a search takes place seems preferable; and the system can achieve this goal by helping users focus on the relevant resources.

Users differ in the ways in which they prefer to see and manage search results. Novice users might want to see one set of merged and deduplicated results, whereas experts might want to browse the results by resource, as they are probably more aware of the differences between the types of information provided by the various resources. For example, OPAC records, article citations, finding aids, and image collections offer different kinds of information.

Hence, a metasearch system needs to accommodate multiple workflows and allow users to choose the one that best suits their needs. Furthermore, users need to have options for moving from the default (simple) interface to a more complex one without having to start all over again. Users should also be able to customize the interface so that it reflects both their area of research and their preferences. The success of an interface is measured by its adaptability to both fixed and changing needs.

The Search Interface

Like most search interfaces, including Google, metasearch systems typically offer a simple search option and an advanced search option. Novice users are not likely to take advantage of the advanced search functions provided by systems in scholarly environments. Experts, on the other hand, need the advanced search options to specify their query in more exact terms. Hence, regardless of the search scope and the way in which it is defined, a metasearch interface should include both simple and advanced searching, with the simple search function as the default interface. Note that the search interface does not replace searches in the native interface,

and a metasearch system's advanced search is only "advanced" in the context of metasearch. That is, it offers those query features that are supported by the majority of information resources.

The Display of Results

The display of results is the most challenging aspect of metasearch interface design. Keeping the Google paradigm in mind, we can assume that users would like to see one merged list of results, with the best results at the top—and they would like to see this list immediately! Those who find the source of the results important would benefit from an option to view the results organized by resource.

One of the ways in which metasearch systems differ from Google is that the former carries out "just in time" searching. A metasearch system does not maintain any local indexes of its information landscape; only when the information is required does the system access the various resources to obtain the results.

The Google approach is based on just-in-case technology. Immense effort is invested in preparing the information prior to users' requests so that when the information is needed, it is obtained immediately. Google maintains indexes for the entire World Wide Web; these indexes include not only pointers to sites but also other types of information, including PageRank data (which reflects the number of links to a page from other pages and takes into consideration the prominence of those other pages). When a user searches with Google, only the indexes are scanned (and not even all the indexes at first—just those with high PageRank scores); the information that Google displays on the screen does not come from the sites themselves but from this vast repository of indexes. Google provides the actual access to a certain Web location only when the user selects it from the list. Needless to say, significant computing power and disk space along with sophisticated technologies for harvesting, evaluating, and maintaining the information are necessary for such powerful tools.

A metasearch system, then, confronts at least two considerable challenges: the results are obtained on the fly, and they are subject to the vagaries of the network. The metasearch system must respond to these challenges to provide results to end users in the most comprehensive manner possible.

Typically, a metasearch system performs a search in two stages: first, it delivers the query and obtains the number of hits along with a reference to the result list. Then the system starts retrieving the records. Because a result set can be very large, only the first results are retrieved initially. After examining the number of hits and the first results, the user can decide whether to stop, ask for more hits, or launch a new query. For example, if the number of result hits is very large, the user might want to narrow down the search by providing more terms in the query.

In most cases, a metasearch system has no control over the order in which results are returned from information resources. Some information providers offer sorting options, but most return the results based on their own algorithms. For example, most abstracting and indexing databases return results in reverse date order; some library systems return results based on date and author name; and Google returns results ranked by relevance, as determined by the Google relevance-ranking algorithm.

Other factors can also influence result sets. For example, some resources are slower to provide results, so their results might be retrieved last, regardless of their relevance. In addition, results from heterogeneous resources can differ considerably from each other in structure, content, and quality.

Current metasearch systems deal with these issues in various ways. Some systems retrieve only very few hits per resource and display all these hits in one merged set based on the order in which the results arrived; other systems use various algorithms to accumulate the results into one list. To overcome the network-delivery obstacle, one system might show the results as they arrive, refreshing the screen until the process ends, whereas another will allow the user to browse the results retrieved thus far while waiting for the search to be completed.

A number of systems organize the results by resource, instead of or in addition to a merged list. The display order can be based on sorting capabilities supported by the resource, if such capabilities exist. For example, a user views a result list that is displayed by resource and wants to re-sort the list by author name. The metasearch system can send the query once again, this time defining the author as the basis for the sort order. The whole result set is re-sorted—not just the records that were already retrieved.

However, the real challenge faced by metasearch systems is the display order of the merged list; listing the results by relevance would probably meet the needs of the users best. Just as in Google.

Relevance Ranking

A metasearch system would ideally use a relevance-ranking algorithm that addresses the problem of displaying results in the most convenient form for users. Such an algorithm would be applied to every retrieved record, regardless of the resource from which it originates, and would assign each record an objective relevance ranking. The algorithm would take into account the user's query, the result record, and the resource-ranking factor (the quality of the resource as determined by the librarian). Of course, the algorithm would use the field type and other factors to calculate the match between the query fields and the result fields.

A system that applies such an algorithm would enable the user to see the best hits first. However, in a metasearch environment, the result set would be limited to the records that have been retrieved, and the best records might be farther down the list of the information resource results. So does this mean that the solution is not appropriate?

To answer this question, we need to examine the initial result set and the display alternatives. The result set depends on the way in which the system is configured to work: typically, the initial result set consists of the first records from each resource. Hence, the result set might contain several hundred records, depending on the number of resources and the number of records initially retrieved. A set of a few hundred results is probably going to satisfy the needs of most users, especially if the best hits appear first. Also, in most systems, users can request more hits, thus expanding the set as a whole, and particularly the number of good results.

Now let's consider the alternatives. When relevance ranking has not been applied, the user browses the merged result set in the order in which the metasearch

system displayed the results. This order usually does not relate to the quality of the results or the quality of the various resources; more often, it relates to the performance of the networks involved. Some systems offer users options for sorting the merged set, but if a user is not aware of a specific title or author, the sorting does not provide much help.

If we compare the display of a merged set to that of results listed by resource, we can see that an end user will not necessarily realize that the latter display is limited to the first results (because the system can continue retrieving results while the user is browsing the list). However, a per-resource list does not provide better options for viewing the most relevant results first—unless the information provider returns results by relevance. If this is the case, the order of the results serves the merged set as well.

In the current environment of most institutions, where users need to search individual resources, the solution offered here provides a great advantage: not only are the results ordered by relevance ranking, but the system also offers a way to compare the results that come from various resources.

Two-fold Access to Information

While most library users have limited research needs and can be expected to find the searching options of metasearch systems adequate, experts might consider them too limited. Such experts are well acquainted with the information resources that are relevant to their field and take advantage of searching options tailored *to* these resources.

To comply with the search options of various resources, the search interface of a metasearch system must be based on a common denominator: the options that are widely supported by resources. Even so, the process of adapting search options to the requirements of individual resources has to overcome normalization issues such as the lack of support for a search field or a different way of supporting a certain function. For example, a resource might not support any searching by phrase or might require a change of syntax for its engine to interpret a phrase search correctly.

When information providers design the interface for their systems, they typically present specific search options that are relevant to the types of data that they offer and their target audience. Thus, experts might prefer to continue using the native interface of a resource in addition to, or even instead of, the options provided by the metasearch system. A case study conducted at Loughborough University found that academic staff and postgraduate researchers "also [use MetaLib to] carry out a search across more than one database for the required topic and then carry out more detailed searching within that database's native interface" (Stubbings and Hamblin, 2004, p. 28).

A metasearch system, therefore, needs to provide access to the native interface of all resources at each stage of interaction with the user—be it when the user locates information resources or views those that the librarian has recommended; when the user examines brief results; or when the user focuses on a specific record. Furthermore, the link to the native interface needs to be sensitive to the context in which the link has been clicked. For example, when the user wants to access the native interface while searching for a resource, the link should direct the user to

the home page or even the search interface of that resource. However, when the user clicks the link in the context of a specific record, the link should go to the same record presented in the native interface.

More Than Discovery

Although the immediate search result is typically a bibliographic record or other form of metadata describing the actual material, the bibliographic record from the end user's perspective serves only as a means of obtaining the material itself, which could be an article, book, video recording, thesis, image, or any other type of material that is available in print or electronic form. Hence, the discovery phase should be followed by additional services that enable users to obtain the objects of interest, either electronically or through other means. A metasearch system needs to integrate multiple software products in a seamless interface.

The link from a bibliographic record to the actual material can be direct, an explicit URL embedded in the metadata, as in the MARC 856 field of a bibliographic record in library catalogs. However, in many instances, the system must perform calculations to create such a link. For example, the bibliographic record might reside in an information repository such as an abstracting and indexing database but the actual material resides elsewhere, such as in an e-journal repository or the library's printed collection. Nevertheless, the user expects to reach the actual material. A library can make this possible by configuring a context-sensitive linking server, which links the user to the material as part of a set of extended services and onward-navigation options. And, indeed, as the source of OpenURL linking, metasearch systems are typically available with a linking server or are able to employ one.

Does It Work?

Metasearch systems have been in use since 2001. A few case studies have been published, such as one that Loughborough University, in the United Kingdom, conducted after it implemented the MetaLib library portal (Stubbings and Hamblin, 2004). In that study, a promising trend in the use of libraries' information resources emerged: the use of databases rose by 609% over the previous year. "Some web-based databases showed a dramatic increase in usage, for example, Zetoc rose by 1,385%, and SportDiscus by 1,207%. Even some of those databases that were not cross-searchable saw an increase in usage, for example, ICEA went up by 225% and ANES by 36%. This illustrates that MetaLib is a useful promotion tool" (p. 29). These figures clearly show that the introduction of a friendly and easy-to-use environment has encouraged users to take advantage of the library's e-collections.

NISO Metasearch Committee

Having taken hold in many libraries and information centers around the world, metasearch systems are now becoming the focus of standards-setting institutions. Following discussions at the ALA Midwinter 2003 meeting, where three vendors—EBSCO, Gale, and ProQuest—drew attention to the significant increase in the number of searches in their systems resulting from the implementation of

metasearch systems, the National Information Standards Organization (NISO) agreed to take the lead in setting up a framework for metasearch activities. The resultant NISO metasearch initiative was established "to enable:

- metasearch service providers to offer more effective and responsive services
- content providers to deliver enhanced content and protect their intellectual property
- libraries to deliver services that distinguish their services from Google and other free web services" (NISO, 2004).

Jenny Walker (Ex Libris) and Andrew Pace (North Carolina State University) are cochairing the committee, which consists of three task groups: Access Management, Collection Description, and Search/Retrieve. European interests are well represented, with Juha Hakala, of the National Library of Finland, heading the Collection Description task group. Other group leaders are Mike Teets, of OCLC (Access Management); and Sara Randall, of Endeavor Information Systems, with Matt Goldner, of Fretwell-Downing (Search/Retrieve).

The initial work by the Search/Retrieve task group has focused on record retrieval. The group intends to recommend data elements to describe a result set and each record within a result set. In addition, the task group will conduct an inventory of information vendors to determine the range of current practices.

There is no doubt that metasearch systems will benefit from a standard description of result sets for the display of results. Metasearch vendors are watching the work of this task group particularly closely and are confident that setting standards in this area will benefit the users of metasearch systems and vendors alike.

Future Directions

Metasearch systems have undoubtedly improved access to library collections and have demonstrated that facilitating access indeed encourages users to take advantage of those collections. We see that such systems manage to offer an interface that complies with the "norms" already taken for granted in the Internet yet also manage to direct users to the appropriate resources.

Competing with Google when it comes to performance is not a realistic goal for the near future. Because Google uses computing and hardware resources that are not available to libraries, resources that enable it to prepare information prior to users' requests, it can respond with great rapidity. However, Google offers another advantage, fuzzy searching, that designers of metasearch systems should consider emulating.

Let's take, for example, a query for the name *Dostoievski*. Because other spellings of the name (for example, Dostoyevsky) are often used, Google offers alternative spellings and does not let the user down. Library systems have tools that show the authoritative names, but these tools are no longer obvious in the metasearch environment: the heterogeneous nature of this environment makes the use of any specific subject-heading list or authoritative name ambiguous. Fuzzy searching, including the soundex algorithm, synonyms, and the handling of various forms of names, might provide an important auxiliary tool.

Another direction that designers of metasearch systems might explore is the personalization of the metasearch environment. It is obvious that users expect to be able to set their preferred display options and language. However, user preferences might also be extended to the relevance-ranking algorithm. That is, the relevance ranking of a specific record should be sensitive to the user's affiliation. For example, if a user in the medical school and a user in the history department are both looking for information about the black death in a multidisciplinary resource, the records that deal with the medical history of the plague should be ranked higher for the first user and those that deal with the social aspects should be ranked higher for the second user. This issue is not so simple, as the same person can assume more than one role when accessing information. Therefore, the user should be the one to define his or her profile and set relevance-ranking preferences.

Another aspect of personalization in which a metasearch system might benefit the user involves references to additional information. Responding to a user's selection of an information resource, the system might offer other resources; and responding to the user's selection of a record, the system might suggest other records. Already implemented in other environments (at amazon.com, for example), this capability seems promising.

Conclusions

As Roy Tennant notes, users are not necessarily lazy but they are human. With so many ways in which they can obtain information, they use the tools that require the least effort and still provide what they consider reasonable results. Not all users can distinguish between information resources of good quality and of lesser quality; and an increase in the use of inferior resources will bring about a gradual decrease in the level of research.

Through recently developed technologies and new concepts of interface design, metasearch systems can offer library users an environment that is friendly and encouraging. With effective tools, users' research patterns change, and we see an increase in the use of electronic library collections. As Stubbings and Hamblin put it, "…the Library believes that MetaLib and SFX have played a major role in encouraging readers and in particular, undergraduate students, to make more use of the resources provided by the Library. Hopefully, the ease of cross-searching databases and linking to full text will mean less reliance on searching for information through search engines such as Google and, ultimately, a better standard of work" (Stubbings and Hamblin, 2004, p. 30).

As designers of metasearch systems, we cannot, and should not, attempt to beat Google. But we can certainly learn the Google lesson and provide a system that is appropriate for the academic environment and that scholars at all levels will find effective and will like using. No doubt that the use of such a system will gradually raise the quality of scholarly research.

References

CrossRef (2003). CrossRef search: 9-publisher pilot for full-text scholarly research. http://www.crossref.org/crossrefsearch.html.

EDNER (Formative Evaluation of the Distributed National Electronic Resource) Project (2002). How students search: information seeking and electronic resource use. Issues Paper 8. http://www.cerlim.ac.uk/edner/ip/ip08.rtf.

NISO National Information Standards Organization (2004). MetaSearch initiative. http://www.niso.org/committees/MS_initiative.html.

Sadeh, T. (2004). The challenge of metasearching. *New Library World*, Volume 105, No. 3.

Stubbings, R., and Hamblin, Y. (2004). How to make your e-resources earn their keep. *Serials,* 17(1), March.

Tennant, R. (2004). Libraries through the looking glass. Eighth Annual ALA Endeavor Digital Breakfast, Jan. 10, 2004. http://www.endinfosys.com/news/alice.pdf.