Short Paper #2 – "Development of Digital Libraries" Search Process Project

Michael J. Giarlo

17:610:540:01 - Professor Marie Radford, Ph.D.

October 27, 2004

For assignment #4, I located four online indexes which seemed fitting candidates for retrieving information on my topic, "development of digital libraries." After scanning the various indexes and databases available through the Rutgers University Libraries (RUL), I decided to stick with my original four choices: *Educational Resources Information Center (ERIC), IEEE Xplore, Library and Information Science Abstracts (LISA), and Library Literature and Information Science.* In order to verify my choices, I issued a few searches on Dialog's DiaLindex file. The search I eventually settled on was "(digital()librar?)(3n)develop?", seeking instances of "digital library" or "digital libraries" within three words of "develop," or "development," in any field. The results of this search confirmed that *ERIC*, *LISA*, and *Library Literature* contained many articles on my topic. *Xplore* is not a file in Dialog, and thus I did not expect to receive confirmation therein. To confirm this choice, I scanned RUL's list of indexes and databases, reading the descriptions of those whose titles seemed relevant. Based on this scan, I chose to stick with *Xplore*.

The general strategy I formulated and used frequently in the Principles of Searching course offered by SCILS is the one I employed within the four indexes chosen for this assignment. For two topic searches, such as "development of digital libraries," I have found searches to be rather successful when using proximity operators to link the concepts together. Using only Boolean AND typically yields too many results, or too few which deal with the development *of* digital libraries. That is, it is not entirely rare to see some form of the word "develop" used in titles and abstracts, and perhaps also in the subject terms depending on the levels of specificity and exhaustivity. By using proximity operators—for instance, within three words—a searcher can increase the probability that

the subject of development is the digital library itself. Additionally, I used phrase operators wherever possible to group the words “digital” and “libraries” together, in addition to truncation operators, or “wildcards.” Fortunately, truncation operators are a feature common to nearly all production-class search engines, as they do come in quite handy. Without them, I would have had to employ more Booleans than I would have liked, for instance “digital library OR libraries” and “develop OR developed OR development” rather than “digital librar*” and “develop*.” This was the general strategy I tried to employ for all four online indexes. In the subsequent sections of this paper, wherein I relay my experiences and evaluations for each online index, I will note how my strategy needed to change for each, if at all necessary.

The main interface of *ERIC*, which it shares with other *OVID* databases, is more or less Spartan; at the top of the screen the date range is listed, so that determining coverage and currency is a cinch, and there is a blue “Help” button. The choice to use the color blue is a good one, as it contrasts with the light beige background of the page, and nothing else on the main interface uses that color. A user's eyes are drawn to it, therefore, and the user's “lifeline” is not buried in a sea of otherwise useless links. Though other elements are included on the main search interface, such as a table showing a history of one's searches and a series of graphical icons which link to other functions, the prominent item is the ubiquitous search box, which is presented in a more or less intuitive layout. To employ my search strategy, I began by scanning the page for “quick help” information on how to use operators. As I could find none, I began my search by clicking the Help button, at which time I discovered that this rather simple screen is considered OVID's “advanced mode,” which led to a bit of disappointment. Search engines which do not

give one the option of using the operators I described above benefit greatly from providing an advanced search, whereby one can search multiple fields in an interface which is becoming rather common on the web. Therefore, the original Help page was not entirely helpful to me, so I scanned for other topics. One such topic which looked quite promising was labeled "Command Line Syntax and Advanced Searching Techniques." In retrospect, it would have been most helpful if a very small table of commonly used operators were included on the main page, rather than requiring users to click two or three times to retrieve this information. Through this help page, I discovered that *ERIC* did indeed support phrase grouping and proximity operators (in addition to a few other potentially useful options). The first search I issued was "digital librar$ adj3 develop$," which retrieved 59 documents. Clicking the "Limit to journal articles" and "Limit to English language" checkboxes, I was able to pare this down to 42 documents, which I used to select some of my fifteen "best" articles.

While *ERIC* has a rather simplistic interface, *IEEE Xplore* provides a number of different searches via the site's navigation bar. The main page for *Xplore* does not contain a search box, so I began by locating the submenu labeled "Search" in the navigation bar, under which four options are listed: "By Author," "Basic," "Advanced," and "Cross-Ref." As I was searching by topic only, the "By Author" and "Cross-Ref" searches were not useful to me. I started out by entering the basic search interface, at which time I noticed a prominent red link at the top of the page announcing a new "full-text search prototype." Diving right in, I went straight to the full-text search, whose interface is identical to that of the basic search: three input fields, three field-limited dropdowns, two Boolean dropdowns, and a number of options for limiting and sorting

searches, such as by publication type and date range. Unsure which operators were offered, I clicked the Help link which is rather small and inconspicuous. The help pages were difficult to navigate, and after ten to fifteen clicks, I discovered that two wildcards are available in addition to a few operators that did not seem entirely helpful, such as <word> and <stem>. Without any proximity operators available, I needed to rely upon the Booleans, and thus my first search at the full-text interface was "digital librar* AND develop*," making sure to limit based on the publication types of IEE Journals and IEEE Journals. The search returned 2,371 documents out of a total of 327,901, a number which I was not surprised to see given the lack of proximity operators and the full-text nature of the search. To limit the results, I issued the same search via the basic interface and retrieved 33 documents, a much more manageable number! The advanced search is more or less a command search, and having already researched the available operators which are available in each of these interfaces, I felt as though the advanced search offered nothing new.

The *LISA* interface is at first rather disorienting; scanning the page from top to bottom, one is confronted with a series of oddly colored stripes which seem to contain different types of options. To describe it in two words, "information overload." While the layout leaves something to be desired, I appreciate how the bottom half of the screen has a number of helpful additions, such as a small table of search examples showing how to use the wildcards and Booleans available, and some pointers on which type of search to use for the query in the user's mind. There are five main tabs, so to speak, on this page: a basic search, an advanced search, a thesaurus search, an index search, and the user's search history. For this project, I focused on the basic and advanced search screens. To

get started, I adopted the same strategy as in the other indexes, but to do so I needed first to check which operators were available. Though the help link is easily found, it doesn't jump out like the help button in *ERIC*. *LISA*'s help facility is easily the best one among the four indexes I used. Like the help utility included in Microsoft Windows operating systems, *LISA'*s help provides an index and its own search interface. Therefore it did not take very long to discover that the "*" wildcard operator was supported, as was phrase grouping, but there was no support for proximity operators. With this information, I went straight to the advanced search page which has a nearly identical layout to that of *Xplore*'s basic search. My first search was "(digital librar*) and develop*" which retrieved 1,108 hits, far too many to be useful of course. Tightening the net significantly, I restricted my terms to the title field with the following search string, "((digital librar*) in TI)and ((develop*) in TI)," which reduced the number of hits to 60. Were I actually writing this paper, *LISA*'s "Check Iris for other print and electronic versions" and "Check LinkSource" links would have come in quite handy!

The default interface of WilsonWeb's *Library Literature and Information Science (LLIS)*, the advanced search page, shares many common elements with *LISA* and *Xplore*: three input fields, two Boolean dropdowns, three field limit dropdowns, and a number of other limits for date and type. Unlike the other indexes, *LLIS* provides a number of other, rather novel options for limiting search results, namely "limit to peer-reviewed," "limit to page image," and the ability to limit by document type (which seems similar to that of Dublin Core's type definition) and physical description (similar to Dublin Core's format definition). These weren't particularly useful for my searches, but I considered them worth remarking upon. Other high-level options in addition to the default advanced

search are similar to those of the previously evaluated indexes, such as a basic search, a subject browse, a search history, and a thesaurus search. As the basic search did not seem to offer anything the advanced did not, I began at the advanced search page, at which point I diverted temporarily to investigate the available operators in *LLIS*'s help facility. The help facility for *LLIS* is very good, though it is not searchable as was *LISA*'s. The links are well-organized and laid out, which makes up for the lack of searchability, and thus I quickly discovered a list of all available operators. *LLIS* provides far more operators than any of the other online indexes used for this project, offering a host of truncation, proximity, and even "concept operators." My specific search strategy would be to locate documents matching "(digital librar*)<near/3>develop*", a direct adaptation of my general search strategy. The search retrieved 5,942 documents, so I scanned some of the citations in the first ten hits. Surprisingly, these citations did not match my search string! That is, some documents had "library" without the string "digital" right before it and other documents did not have the phrase "digital library" within three words of "development", which would seem to be errors. Looking back at the help pages, I was unable to determine what I was doing wrong until I realized that I was in fact searching full-text, which of course was not displayed on the citation page! I then issued the same search limiting the scope to the title field only. This produced no hits. I tried the same on the subject, keyword, and abstract fields, but none of these produced a single hit. To confirm my theory, I issued the same search limiting the scope to the actual text, and again to my surprise, there were no hits. I still have not figured out what I have been doing wrong on the *LLIS* search interface. For the time being, I have included documents from my original search which did return some relevant hits.

Though I believe my general search strategy to be a sound one, I was required to make different changes to it for each of the online indexes, some because no proximity operators were available and others because the number of hits was either too high or too low. While I was able to come up with fifteen articles I would use for a term paper on the development of digital libraries, I would probably refine my search strategy a bit were I actually engaged in research by utilizing the thesauri more, and also by checking to see what subject terms were applied to articles I judged to be most relevant, and issuing subsequent searches on those specific subject terms. All four indexes yielded a wide range of good articles, but my experiences with each were quite different: some help utilities were much more useful than others; while the interfaces of some were similar, there were often inconsistencies between what one site considered basic and another considered basic; and I also found it interesting to see what each site chose as its default interface. Some chose a very basic search while others chose a more advanced search, but it seemed to me that each chose the search it deemed to be the most useful. I can only surmise that part of the decision was made because of the range of users that come to the indexes. That is, some users are undergraduates who may not be that familiar with database searching, but others are information professionals who generally would prefer to use the most functional interface available. This, of course, also varies depending on the type of information being sought. In summary, I found each of the indexes to be quite useful, though I am as yet confounded by my experiences with *LLIS*.

Bibliography

Besser, H. (2002). The next stage: Moving from isolated digital collections to interoperable digital libraries. *First Monday, 7(6)*. [Discovered: *ERIC*]

Bruce, B. C., Bishop, A. P., Heidorn, P. B., Lunsford, K. J., Poulakos, S., & Won, M. (2003). The inquiry page: bringing digital libraries to learners. *Knowledge Quest, 31(3),* 15-17. [Discovered: *ERIC*]

Falk, H. (2003). Developing digital libraries. *Electronic Library, 21(3),* 258-261. [Discovered: *LISA*]

Gold, Anna Keller. (2003). Multilateral digital library partnerships for sharing and preserving instructional content and context. *Libraries and the Academy, 3(2),* 269-291. [Discovered: *ERIC*]

Jeevan, V.K.J. (2004). Digital library development: identifying sources of content for developing countries with special reference to India. *International Information & Library Review, 36(3),* 185-197. [Discovered: *LLIS*]

Lynch, C. (2002). Digital collections, digital libraries and the digitization of cultural heritage information. *First Monday, 7(5)*. [Discovered: *ERIC*]

Marcum, D. B. (2003). Development of digital libraries: an American perspective. *Libraries and the Academy, 3(2)*, 344-346. [Discovered: *LISA*]

McClelland, M. (2003). Metadata standards for educational resources. *Computer, 36(11),* 107-109. [Discovered: *IEEE Xplore*]

Paepcke, A., Cousins, S. B., Garcia-Molina, H., Hassan, S. W., & Ketchpel, S. P. (1996). Using distributed objects for digital library interoperability. *Computer, 29(5)*, 61-68. [Discovered: *IEEE Xplore*]

Shiri, A. (2003). Digital library research: current developments and trends. *Library Review, 52(5),* 198-202. [Discovered: *LISA*]

Storey, T. & Van de Sompel, H. (2003). Developing new protocols to support and connect digital libraries. *OCLC Newsletter, (261)*, 12-14. [Discovered: *LISA*]

Tanner, S. & Deegan, M. (2002). Economic strategies for developing digital content and digital libraries. *Library Hi Tech News, 19(3),* 14-18. [Discovered: *LISA*]

Tuominen, K., Talja, S., & Savolainen, R. (2003). Multiperspective digital libraries: The implications of constructionism for the development of digital libraries. *Journal of the American Society for Information Science and Technology, 54(6),* 561-569. [Discovered: *ERIC*]

Xie, H. (2003). Dimensions of human-work domain interaction: A preliminary analysis for the design of a corporate digital library. *Proceedings of the ASIST Annual Meeting, 40,* p109-118. [Discovered: *ERIC*]

Zhang, D. (2001). Review of digital libraries development around the world. *Journal of the China Society for Scientific and Technical Information. 20(4)*, 483-494. [Discovered: *LISA*]

Zhang, X. (2003). Design and development of an open digital library: the practice of CSDL. *Journal of the China Society for Scientific and Technical Information. 22 (5)*, 520-525. [Discovered: *LISA*]

Annotated Bibliography of Journal Indexes

Educational Resources Information Center. (2004). New York, NY: Ovid Technologies. [Available: http://www.libraries.rutgers.edu/cgi-bin/ruonly/eric3.ovwb]

The *Educational Resources Information Center*, or *ERIC*, is a database of national scope, holding "over one million citations covering documents, journal articles, technical reports, program descriptions and evaluations, and curricular materials in the field of education." Materials in *ERIC,* to which Rutgers has access, are dated between 1966 and July 2004, with regular updates beginning again in Janary of 2005.

IEEE Xplore. (2004). New York, NY: IEEE. [Available: http://www.ieee.org/ieeexplore]

IEEE Xplore, otherwise known as known as the *IEEE Electronic Library*, is a full-text portal to dozens of journals, engineering standards, and conference proceedings. Information accessible through *Xplore* typically falls between 1988 and the current year, though there is "select content back to 1950," and is updated on a weekly basis. A number of configurable interfaces are provided, including five different searches: by author, basic, advanced, a "CrossRef Search Pilot," and a full-text search prototype.

Library and Information Science Abstracts. (2004). London, UK: Bowker-Saur. [Available: http://web5.silverplatter.com/webspirs/start.ws?databases=LISA]

Library and Information Science Abstracts "abstracts over 550 periodicals from more than 60 countries and in over 20 different languages" between 1969 and 2004, chiefly in the fields of library studies and information science. In addition to periodicals, *LISA* also indexes and abstracts conference proceedings, book reviews, and research report series. Content is updated monthly.

Library Literature and Information Science. (2004). Bronx, NY: H. W. Wilson Company. [Available: http://vweb.hwwilsonweb.com/?sp.dbid.p=S(Y6)]

Library Literature and Information Science is an index primarily of articles and book reviews related to library studies and information science, providing citation and descriptor information about materials dating back to 1984 in addition to full-text access for post-1994 materials. *Library Literature and Information Science* is updated four times a week.