

Web Search Exercise
ILS 501-S72
Sharon Porter

Introduction

This exercise taught me that I have been a sloppy researcher in the past! I thought that by hooking up two Boolean operators at a time was showing expertise. Little did I know how many variables could be involved in the simplest of searches. I compared three search engines and one meta search engine by using the inquiry: I need information on how academic libraries are using messaging (instant or text) in virtual reference services offered to distance students.

Searching for search engines that feature truncation (librar* to avoid having to search for “library” and “libraries”) and nesting, I found three search engines that fit that criteria. (“Best Search Tools Chart”; Sullivan, 2001) and (Sullivan, 2004) They are not engines I reach for often, however, so I had a huge learning curve to surmount.

Not long after I started searching, I discovered that using “IM” as the instant messaging acronym returned many hits because the various engines would grab any word that had “IM” in them. I even wondered if the word “messaging” was so specific to text and instant messaging that perhaps I could do without using the qualifiers-text and instant. I ruled in favor of asking for those terms. I did not want to cloud the waters with returns about bicycle messaging or other messaging services like doctors use.

I also chose to use “distance learning” as a phrase and did not include OR arguments for “distance education,” “distance students,” or “distance learner” to define the number of search terms to a controllable number.

I also did not want to filter out blogs as I wanted some grassroots experience to show up in the return lists. I did not, however, want hits that included the lists of sessions at conferences. I grapple with that later in the discussion of the specific searches.

To aid in keeping the variables of each search straight, I will present the search string variables in a chart by search engine. I will capitalize terms found on a search engine page that denote links or choices to be made and will write Boolean logic as I did in the search engine boxes. For example, Advanced Search denotes a link on a page and dalmations AND “fire houses” as a search string.

The Searches

Altavista <http://www.altavista.com>

Altavista features full nesting and Boolean capabilities on its front search strategy and offers an Advanced Search Strategy behind that as a link on the page. I chose to search

Worldwide and in All Languages, figuring I could get a translation if necessary from Babel Fish.

Boolean/ Advanced?	Search string	# of hits	Rationale/Process	Analysis of hits (first 30)
Boolean	("instant messaging" OR "text messaging") AND ("academic library" OR "academic libraries") AND reference AND distance	586	Without a truncation feature, I used the OR statement about libraries. I thought that adding "student" to "distance" might be too limiting as with using "service" with "reference"	Undesirable commercial sites for books, conference proceedings, bibliographies, and library newsletters
Boolean	Added "virtual" to "reference"	214	To narrow the aspect of reference services to "click" services	Picked up a hit from the Virtual Reference Librarian site
Boolean	Added "learning" to "distance"	100	To try to eliminate "running the distance" type results	Picked up a registry of real- time reference services (a term I had not seen before) with libraries, the academic search term kept out commercial distance learning companies
Advanced	All of These Words: "text messaging" "instant messaging" reference distance Any of These Words: "academic library" "academic libraries"	2,020	Lost the ability to nest the terms using Advanced Search and so my net cast out wider. On the Search Tips page, no provision for a phrase using quotations to hold terms together	Lots of conference listings

Advanced Search- Boolean Expression box	("instant messaging" OR "text messaging") AND ("academic library" OR "academic libraries") AND reference AND distance	568	Lost the nesting which allowed more hits	Index to papers from college libraries, blogs, archival postings, proceedings, less articles
Advanced Search- Build a Search Query	All of These Words: reference distance academic This Exact Phrase: instant messaging, text messaging Any of These Words: library libraries	3 118,000	Gambled on using two phrases in the box separated by comma because there was nothing on the help page to help me chose or add another phrase box. Got four hits when I removed the comma to separate the two phrases. Removed "text" and "instant" and moved "messaging" to All of These Words	Off the wall hit: chat/date room 2 hits seen previously-the first one excellent Started to get hits using the word "library" as a set of instructions or manuals, word "messaging" too generic

In the original string search -("instant messaging" OR "text messaging") AND ("academic library" OR "academic libraries") AND reference AND distance- I got several repeats of commercial sites advertising a table of contents for a book about messaging in reference services. The content was not there on those sites. The first ten hits were virtually identical to the search I did by adding "virtual" to form the phrase "virtual reference." The fifth hit on both searches was a top-notch article from the ALA with many references to academic libraries who are using messaging. However, interspersed in the "virtual"-term search brought some blogs to the result list that proved interesting as they contained grassroots information about utilization in "real life." The blogs provided the nuts and bolts of using library staff and resources to support messaging activities. In both of the results list, I got lists of sessions at conferences and proceedings of professional organizations. While this may be useful to identify key players in the development of messaging services for future follow-up, it is not the immediate information I was seeking. I will have to contact those people after the conferences for handouts, Powerpoint presentations, web pages and the like. Again, on

both results lists, I got bibliographies that will need follow-up to locate the actual articles in a journal database or from archival material available on the author’s website or academic file directory.

When I added the word “virtual” to reference, the results were more fruitful by skewing the information toward more current material. There was an excellent resource that showed up in hit #10 titled *Virtually Yours, Real-Time Library Reference Services*, with an excellent bibliography and some helpful case studies. There appeared a hyperlinked list of *Virtual Reference Projects* which was much more directly linked to resources than a standard bibliography or annotated list. The Digital Reference Conference included the entire text of a messaging paper. *Ask Us Illinois* provided a hyperlinked bibliography and useful information on marketing messaging services. There was a sobering study from Ontario, Canada titled *The Library Visit Study* that outlined the research interview as being problematic no matter what the venue. UCLA provided an extensive bibliography on digital reference but the sources were dated no later than 2002- an eternity in digital trends and development. All in all, this search had much more precision than the search using the entire string search. The recall was still a bit high.

Again, there were hits that were commercial product “testimonies” of successful ventures with academic libraries. There was a hit that had some good information in a Powerpoint slide show for about two slides and then the commercial began for a specific messaging service, price tag attached. A puzzling return in the “virtual reference” search was a purely commercial site for *Yahoo! Messenger* that did not even have all the search strings I had asked for.

Adding “learning” to “distance” at the end of the search string, yielded less precision and less recall. I did not find the results are desirable as the middle search I performed by adding “virtual” to “reference.” I got some strange hits that did not include all the search terms. I got a bibliography from the Virtual Reference Desk that had more current resources on it. The foreign language sites were much more numerous. There was a site for librarian’s jobs that included all the search terms but contained no information about the content I was searching for. There was definitely a skew evident in using the phrase “distance learning.” The content reflected away from academic librarianship to phrases in the search that we further apart in the results, yielding less content. In summary of this search, the recall was down but the precision was not desirable.

Yahoo! Search <http://search.yahoo.com/>

Once known for its subject directory arrangement, Yahoo! Search has metamorphosed as a subject directory, web portal and search engine. In 2004, it started serving up its own database supplemented heavily from Inktomi. (Notess, 2005) It features full Boolean logic (except truncation) but includes some listings that have paid to be there.

Boolean/Advanced?	Search string	# of hits	Rationale/Process	Analysis of hits
Boolean	(“instant messaging”	593	Again, without a	While some

	OR “text messaging”) AND (“academic library” OR “academic libraries”) AND reference AND distance		truncation feature, I used the OR statement about libraries	dictionary entries picked up and some employment listings, the first thirty hits looked more like the content I wanted
Boolean	Added “virtual” to “reference”	164	More narrowing of “reference” service	Hits with phrases like “distance themselves,” more hits from papers and professional journals, hip websites (My Web Librarian, The Shifted Librarian)
Boolean	Added “learning” to “distance”	68	To eliminate the use of “distance” in other contexts without using NOT	Hits showed more emphasis on distance learning, rather than academic response to reference services
Advanced Search	All of These Words: reference distance academic This Exact Phrase: instant messaging, text messaging Any of These Words: library libraries	139,000, 000 140,000, 000	Lost nesting, lost the ability to hold phrases together Removed “text” and “instant” and moved “messaging” to All of These Words	Dumping the apple cart-lost all control of logic. First few hits look useable but then getting usages of words not intended, example: “reference model.” Tons of vendor information and academic library listings of services

An interesting feature of the Yahoo! Search engine is the ability to search for material that has a Creative Commons license to denote material that offers possible change to the user as opposed to material that is static, for commercial purposes.

As in Altavista, I did not limit by domain. I was tempted to eliminate the.com sites to get rid of the book reviews and the distance learning packet materials, but elected to keep it to catch some content available in monographs. Since the topic is current and not likely to have many published monographs on this daily-changing topic, the amount of hits was negligible for books.

Yahoo! Search uses a unique shortcut to access its directory services by adding a “!” after a topic to take you to that sub-directory. It is a clever way to access the directory features on the fly but it took me some looking to find the tip for that shortcut. The Tips page is a fairly standard explanation of the advanced search operations.

With the results of my first full string search that I used with Altavista, I got a similar top ten list. A new site that listed high in the results list was from the University of South Alabama Medical School Library. The search terms were all there but not very contiguous and the content as not helpful for my purposes. In the top five was a metatag list document from some educational institution in Arizona...puzzling. A very interesting blog came up in the top ten called *Technogeekery for Librarians* that had a lively discussion of using messaging in virtual reference. It was cheeky and real-world. Another blog called schwagbag that focused on distance students and libraries popped up at #11 that contained a survey of academic libraries using IM. In general, there were many more blogs in the hit list than I had seen so far. This type of current data I had not seen before. A fairly recent (2003-ish) bibliography came early on the list from ALA’s Reference Services Section. I hadn’t seen this *Reference Research Review 2003* before and it was useful even though not hyperlinked. I looked for a newer version but I did not see it on the internet. There was high recall in this search and some unique hits that were useful. The number of hits, though, was unmanageable.

As in Altavista, I added the word “virtual” to “reference” to narrow down the result list and dig deeper into the more reflective current term of “virtual.” I got many of the same results from the full search string search but got many more blogs and current material than I had gotten before with the whole string. I started to get *Library Journal* full-text articles. The use of “virtual” brought me many, (and in my humble opinion) more hip writing about messaging in academic libraries. *The Lethal Librarian* is another blog with the author reflecting on research on the topic and relating it to a working library situation. A webcast archive picked up in this search that I had not seen before. It listed the main players of this event so I could follow-up on the material. All in all, there was better precision with more current content with a lower recall.

When I added “learning” to “distance” to lower the recall, I got the same foreign language sites I had previously gotten in Altavista. I wonder why that is? The top hits were changed little but toward the middle thirty and forty hits, I noticed that I was getting more college catalog entries and hits from trade and commercial schools. The scholarly content took a nosedive with many more stream of consciousness-type blogs with very little to offer in terms of solid information. Like in Altavista, the recall may have dropped to an acceptable level but I lost the precision past acceptable levels.

MSN Search <http://www.msn.com/>

MSN Search is the portal of the MSN network and sports a large database with full Boolean logic, minus truncation. In Greg Notess' review of this database, he cites as a disadvantage the lack of advanced searching capabilities. The home page is glitzy, offering a subject directory on the left navigational area with plenty of popular culture, shopping, sports, weather, and news for a variety of interests. One feature that is unique is the ability to narrow down these interests to reflect local items. However, when searching for our town in the shopping, sports, and weather topics, the searches yielded a few maps and local entries and then proceeded to serve up a list of national items.

Boolean/Advanced?	Search string	# of hits	Rationale/Process	Analysis of hits
Boolean	("instant messaging" OR "text messaging") AND ("academic library" OR "academic libraries") AND reference AND distance	1,231	Without a truncation feature, I used the OR statement about libraries. I thought that adding "student" to "distance" might be too limiting as with using "service" with "reference"	Hits tend to be more academic than I expected
Boolean	Added "virtual" to "reference"	577	Narrowing down the "reference" term	Wide range of the meaning of "distance" in this context, hits seemed more current
Boolean	Added "learning" to "distance"	270	Ridding the search of uses of the word "distance" in phrases- not as in "distance education" etc.	More conferences, listings of sessions
Boolean	Added "effectiveness" to end of string	34	For the purpose of the exercise, I wanted to narrow the number of hits.	"Effectiveness" not necessarily (actually not often) tied to the main idea of the search
Advanced Search	Not available as an option		Help section was apropos to MSN itself, not for search tips	

Wow! Was I surprised at the top twenty of the results list for searching the entire string in this search engine! The front page is so consumer-oriented, reflecting popular culture that I was amazed at the list I got back. Even though the recall was too high, the top twenty of

this list were content heavy and very reliable, academic sources. The *ALA Issues and Trends Talking Points* was a wonderful summary of the state of the library today. (Ooh...that statistic that 80% of college students use the internet for all or most assignments makes me nervous!) Bibliotech Web, a blog by Scott Pfitzinger, a reference librarian in Indiana, was high on the list and served up quite a bit of information. Of course, I will cross-check to verify his credentials-that for any blog yielding information I may use. His blog was so nuts-and-bolts that I could use it for a training manual on messaging and virtual reference! I learned a new concept termed “walking paper” in one of the first bibliographies I read. The term came from Aaron Schmidt, reference librarian, at the Thomas Ford Memorial Library in Western Springs, Illinois. The term refers to the electronic “boxes” that contain information for student research-iPods, handheld computers, cellphones, “the next big thing.” I found very little in the first twenty sites that was not useful.

Adding ‘virtual’ to “reference” in searching MSN, brought a similar list forward but an exceptional first hit, *A Registry of Real-Time Digital Reference Services*. Narrowing the list to academic libraries led me to a hyperlinked list of libraries employing messaging techniques in their reference services. That was very helpful information to have. There was a state library report from California that accessed the use of messaging in public library reference and recommendations for reactions to patron survey data. It reinforced the need to survey our patrons here for their interests in receiving data in non-traditional ways. The websites that began to show here are contemporary with more blogs and more research articles about messaging. *Library Student World* from Drexel University posted a fascinating literature review article, “A Review of the Literature of Information Seeking Behaviors and Needs of Young Adults.” It was funky in the browser (exceptional large font and formatting was nonsensical) but once pasted into Word I could read it. A very useful article about research interviewing, “Question Negotiation and the Technological Environment” on the D-Lib Magazine website proved to be essential information. I had not seen that hit before. This return list brought forth some unique material with superb content even though the recall was still pretty high.

As in the other searches, in MSN I added “learning” to “distance” to improve the high recall ratio. As in the other results when I introduced this term, the results list started to move toward the distance learning aspect away from a focus on academic reference services. I got hits about continuing education, more school news and events with just the terms represented somewhere in the text, and more conference listings. This was not a successful precision maneuver even though recall was lower. It was still a list that was too long to be useful. It was too diluted to be interesting.

Dogpile- The Meta Search Engine <http://www.dogpile.com>

Dogpile searches two (out of six) of the search engines that I used previously in this exercise: MSN Search and Yahoo! Search. There is a search comparison tool page that illustrates the process that Dogpile uses to compile its returns. It portrays on the screen the results for each engine you have chosen and then the relationship to the top searches that Dogpile serves up. It lists the duplicates and unique hits of your search term. Claiming in an overlap study that 1.1% of the top returns of the four leading search

engines are the same, Dogpile gives them to you all at once. Dogpile is based on the premise that results from a single engine search miss significant hits. The Best of Breed tool shows you what it has found in each search engine as unique and you can stack the windows against each other for comparison. In the preferences window, you can determine the number of results per page, bold the search terms add a filter for objectionable sites, and check to open sites in a different window. I turned off the new window option but fought it the whole time I was searching. It would not accept my changes.

Boolean/Advanced?	Search String	# of hits	Rationale/Process	Analysis of hits
Boolean	("instant messaging" OR "text messaging") AND ("academic library" OR "academic libraries") AND reference AND distance	62	Without a truncation feature, I used the OR statement about libraries.	Free ringtones began to show in hit 21.
	Best of Breed breakout-Google	0	No Boolean logic in Google	No hits in Google
	Best of Breed-Yahoo! Search	0	Boolean is available in Yahoo! Search so I can't figure out why no hits	No hits in Yahoo! Search
	Best of Breed-MSN	# not listed on screen	Checking for uniqueness and number of hits	Top 10 were unique hits
	Best of Breed-Ask	# not listed on screen	Checking for uniqueness and number of hits	9 of top 10 were unique
Advanced Search	All of These Words: reference distance academic This Exact Phrase: instant messaging text messaging Any of These Words: library libraries	62 62	Would not allow a comma between two sets of phrases Moved the word "messaging" to All of These Words and eliminated "text" and "instant"	Opened to so many uses of "academic" and "reference" because I lost the nesting. Moving the word "messaging" up to the All of These Words made no difference

In the excellent recall of 62 hits, there was a surprising amount of “dogs” in this search of Dogpile using the entire search string with the nested Boolean strategy! Complete off-topic entries appeared at hit #21 (free wallpapers,) and #24 (free ringtones for your cell phone). Examination of these hits revealed no search terms whatsoever. The first hit in the list was a resource from 2001, hardly current, and about chat reference. The second hit was from the Instructional Technology Council with the terms spread out over many links off their page. The information was not directed to academic libraries or to reference services in higher education. The third reference, an article by Jody Fagan and Michele Calloway “Creating an Instant Messaging Reference System” was a familiar website, having seen in all the lists but much lower than third. The next hit was a website from the Atkins Library at the University of North Carolina using a paragraph to announce their use of instant messaging. The University of Wisconsin made a showing and then a scholarly doctoral thesis from Syracuse University made its way forward. Almost all of the first twenty hits were unique to all the search engines I had used previously. This uniqueness did not contribute significantly to the content I was seeking. The precision ratio was not acceptable.

I know that the Boolean strategy of nesting mystifies Dogpile so I was curious to see Dogpile’s offering to interpret my search in the hyperlink at the top of the page: Did you mean (instant messaging or text messaging) and (academic library or academic? The result list broadened to 84 hits and they were blanketed with commercial sites for ringtones, telecom companies, and cell phone services. Most of the sites had no search terms present. The syntax that Dogpile used to refashion my search string gave me an indication of what to expect.

In the Best of Breed breakouts of each search engine covered by Dogpile, there were surprising results. There were no results from Google which did not surprise me given that there is no nesting strategy in Google. However, the Yahoo! Search yielded no hits either and they had previously served up interesting and current content when I searched it by itself. I can’t account for the poor showing with Yahoo! in the meta search. Almost all of the sites found in MSN were unique to the search I had performed previously. The results included a 2002 bibliography I had not seen before, several pages from the Indiana University Distance Education Student Primer and related materials, and a poorly-designed staff support web page of unknown institutional origin. The top ten delivered from the Ask search engine did a bit better with solid content in the top nine results, almost all unique.

Before I finished with Dogpile, I tried the Advanced Search feature and placed the same search terms in the All of These Words, This Exact Phrase, and Any of These Words I had used in the three search engines accessed previously. I got the same number of results compared to the original search string but the web sites that were found used the words “academic” and “reference” in separate and not useful ways (example, “in reference to a previous conversation... or that argument is purely academic....,” not providing me with the content I needed. I tried to simplify the search by removing the words “instant” and “text” and moved “messaging” to All of These Words but it made no difference in the delivery of suitable content.

Recommendation/Evaluation of the Searches

Remembering some of the sources I found in journals using available Boolean logic (truncation and not nesting,) I had an idea of what kind of information was available in the library literature. I had previewed the juried and peer-reviewed material from professional journals, research, and survey materials. I also knew that I would expect to find casual writings and more on-the-ground reports on the internet. I appreciate having the ability to find the best of both venues and to also check our webpac for monographs. If I had relied on one search pool or another, I would have missed some energy of the content and some thoughtful, research-based information. I hope our university students learn what they are missing when they eschew one venue of searching over another. This exercise has taught me about that balance and to tailor the character of information necessary to research a topic of interest.

In respect to narrowing and widening the search string, I found that with all three engines, narrowing the search to include AND “distance learning” corrupted the content of the search to the slant of distance education that did not illuminate the information I wanted on virtual reference. My recall lowered using each engine but precision went awry. When I narrowed each search with AND “virtual reference” I got closer to the content I desired but was still overwhelmed with the recall quotient. I might have chosen not to include the distance learning string as, without truncation abilities, was completely arbitrary in its helpfulness. It could have read “distance learning,” “distance students,” or “distance education.” I learned to live with the conference proceedings-type hit because some of them were hyperlinked to presentations.

I got the results I expected from Altavista. The content was staid but moderately solid. I did not get much of the grassroots experience that I was looking for that made the content feel fresh. I have used Altavista years before and had forgotten about it since Google became a verb. The material seemed less edgy and current than what I found in MSN or Yahoo! I was pleasantly surprised to find more current information through blogs and recent bibliographies in Yahoo! Search. I found more audio and visual media there as well. There were webcast archives that I did not find in Altavista. I was still disappointed that I could not narrow my search enough to improve the recall at the sake of losing the precision. MSN Search shocked me most in its performance. I was initially dissuaded to use it because of the glitz of the home page and the ponderous amount of popular culture links there. I could have easily dismissed it as being not helpful by the look and feel alone. The content I found in the entire string search was fruitful with the first twenty sites reflecting excellent content. Narrowing the topic gleaned more research that was scholarly and current with a jaunty mix of blogs and “real-world” experiences. It felt less commercial than the other engines and more reflective of trends in academic librarianship.

Dogpile was like jamming a round peg into a square hole. The search strategy that I had chosen was not adaptive to the engine protocol. The results were weird and uneven. When analyzing the breakdown of the results list by search engine, the results were not

consistent from searches conducted with the search engine alone. The results lists bore little resemblance to searches conducted before but represented much unique content. That would have been great if the content had been rich and useful. Sadly, it was not. I did not understand the alternate string I was offered by Dogpile, resulting in a commercial glut of websites. Maybe that was the strategy-offer the searcher another way to search and then inundate the searcher with paid advertisement sites. Those sites often contained none of the search terms.

I will be using MSN Search more often now but will be curious to see if the trendy research I found there is consistent while searching for other topics. I will also not be using a meta search engine with a complicated nesting search strategy that I was using. I will keep in mind that using the Internet as a sole source of information is folly in gathering academic resources. I've learned that isolating all the variables in a search would be so much easier if librarians ruled the 'Net! (Schneiderman, 1996)

Resources

- Best Search Tools Chart [Electronic Version]. Retrieved October 2, 2005 from <http://www.infopeople.org/search/chart.html>.
- Notess, G. (2005). Review of Yahoo! Search. Retrieved October 13, 2005, from <http://www.searchengineshowdown.com/features/yahoo/review.html>
- Schneiderman, R. A. (1996). Why Librarians Should Rule the Net. Retrieved October 17, 2005, from <http://hackvan.com/pub/stig/news/ENODE--Why-Librarians-Should-Rule-The-Net.html>
- Sullivan, D. (2001). Search Features Chart [Electronic Version]. Retrieved October 2, 2005 from <http://searchenginewatch.com/facts/article.php/2155981>.
- Sullivan, D. (2004). Major Search Engines and Directories [Electronic Version]. Retrieved September 28, 2005 from <Http://searchenginewatch.com/links/article.php/2156221>.