DATABASE SEARCHING TIPS: PART 1

By Marvin Hunn

The Atla and WorldCat exercises are designed to make you think about how to search a structured, controlled-vocabulary metadata-based database using explicit operators and search syntax.[1] Practical points related to those exercises are summarized here.

Generalities

1. Link to databases and reference works via the library website (http://library.dts.edu). The home page includes a search box for a few of the most frequently used databases. It also has a dropdown list of all databases. Not sure which database to use? The Databases Listed by Topic or Function (http://library.dts.edu/guide-topic) may help. It is organized to help you find resources by topic. It contains descriptions of many databases and reference works, sometimes with links to tutorials.

2. We provide access to scores of databases. But you don't have to learn how to use scores of search engines. Most of the databases are hosted by a few vendors (like EBSCO and ProQuest). Often all the databases from a given vendor use the same search engine and the same interface. Some databases from that vendor may have special features like a thesaurus, but all databases from that vendor will offer the same basics. Further, search engines from different vendors tend to have the same basic features implemented in similar ways. So your task is easy: learn the common core.

3. At this point in time, most library search engines do not support natural language queries; you can't just talk to the software like you would talk to a friend. You do need to know how to formulate search statements using a syntax the software can process. You need to know how to use operators and how to group terms. This skill is easy to acquire. "Basic Concepts for Database Searching" includes an introduction to operators (https://library.dts.edu/basic-concepts).

4. Picking good search terms is far more important than using complex search statements. This is the key to success. Below we summarize some practical ways to discover good search terms. Many databases use standardized terminology for names and subjects. This is the "controlled vocabulary" approach to expressing access points. Often these controlled vocabulary terms are good search terms. Sometimes other terms are more helpful. See Library of Congress Subject Headings for an example: http://id.loc.gov/authorities/subjects.html

5. Initial searches can often be improved. Evaluate the results of every search for possible improvements and then search again. Refine. Persist. We develop this idea below.

6. Precision is inversely related to recall (completeness). A very precise search will not be comprehensive or complete; a very complete search will not be precise—it will include many, many irrelevant citations. Because of this inverse relation, you must sometimes choose between either high precision or high recall, and formulate your search statement accordingly.

---

[1] Confused by terminology like "metadata" and "controlled vocabulary?" For help, see https://library.dts.edu/basic-concepts.

7. If you are looking for very specific information, a brief citation or abstract may not suffice to tell you whether a particular work will be helpful. Reading is often the only way to discover whether a source is helpful. Reading takes time and patience. Budget time accordingly.

8. If you are engaged in long-term research (such as thesis research) then expect to search many databases repeatedly. Over the course of your research you will discover new topics you need to investigate. So keep a list of databases you have searched, what terms you used, and when you searched.

How to Find Good/Better Search Terms

Selecting good search terms is extremely important. Sometimes you will need to find additional synonyms in order to formulate a more comprehensive search. At other times you will need to change terminology to formulate a more precise search. Listed below are some tactics to help you with both goals when you encounter problems. If you are searching a controlled vocabulary database, then focus on finding the correct standardized terminology for that particular database. If, on the other hand, you are searching a natural language database that lacks vocabulary control, then synonyms and use of phrases will be more important.

1. Execute successive searches to improve searches. Conduct preliminary searches in the database. At this preliminary stage, you may search for "obvious" subject terms or you may search for relevant specific titles you already know about. Review results of every search for relevant new terminology. Examine many records in detail, looking for patterns. Note words used in titles. Note terminology used in subject headings. Execute new searches with the new terminology.[2] For example, an initial search for "pastor" retrieves some records. Examine those records and notice many of them use the LCSH term "clergy." Include "clergy" as a search term. The successive searches approach can be used in any searchable database, even those that lack subject headings. This approach is just common sense, but it is commonly neglected because searchers are not willing to patiently and thoughtfully review results. Evaluate terms one by one. You will discover some terms are unhelpful and should be discarded. When an exploratory search surfaces a term (word or phrase) that seems like it might be helpful, do a search just for that term to see if it really is helpful. Extremely rare terms or highly ambiguous terms are usually not helpful, for example. (But you may be forced to use them if you can't identify any good terms. That will be rare.)

2. Browse headings. Some databases will let you browse term lists. Unclear on the distinction between searching and browsing? Searching allows you to look for any arbitrary word or combination of words anywhere in a record. Browsing is fundamentally different. It allows you to scan a sorted list of fields in various categories (author, title, subject, etc.). It is a two step process. Ordinarily you supply a word or words you expect to match the start of a field (e.g., the first two or three words of a

---

[2] The successive searches approach is sometimes called "recursive searching" or "using feedback" or even "pearling" (where the record you start with is the pearl, and you grow a string of pearls with successive searches based on subject headings in those records) or "following the tracings" (where subject headings and other access points in an antique printed catalog record were known as tracings).

title). The database then displays a sorted list with the closest match at the start of the list. Then you pick one or more specific entries from the list, and matching records are displayed. So browsing lets you see a list of potentially useful terms.

3. Use cross references. But many databases lack explicit cross references.

4. Consult thesauri. Check the Library of Congress subject heading guide or some other relevant list of subject headings. Atla, ERIC, Psycinfo, and some other databases have their own thesauri similar to LCSH. Consulting LC is helpful even if the particular database you are searching uses something else as its list of standardized terminology. The most comprehensive multi-lingual, multi-disciplinary authority file is http://viaf.org .

5. Brain storm. Still don't have the magic terms? Step back and define what you need. List the distinct concepts. List synonyms for each concept. Remember to include central Bible passages. Record all of this in a master list of terms. At this point also take a minute to ask yourself what your goal is. What do you want: high precision or high recall? (See below discussion on how to retrieve more or less, which parallels how to increase recall or increase precision.)

6. Read before you search any more. Read an overview of your topic in an encyclopedia or textbook. Note additional terminology. Note bibliography at end of article. Read any highly relevant works you have already identified and think about what is really relevant. Continue to follow footnotes.[3] Update your master list of search terms based on anything you read. Try your searches again.

7. Try a different database or specialized bibliography.

8. Ask library reference staff for help. Ask faculty to recommend specific works.

How to Retrieve Less and/or Increase Precision

Some databases (like WorldCat) are so enormous that they will yield thousands of records for a general search. You can usually limit the results to recent publications in English or to publications available full-

---

[3] This process of using footnotes/references in one known work to find other useful works is sometimes called citation chaining. It is very common and very useful. But be sure you read what you cite!!! Citation indexes are a related topic. Citation indexes use citations (footnotes and other references) to associate or link documents together. You are probably more familiar with the concept in Internet search engines that let you search for "pages that link to this page." Citation indexing is old. In the 1870s Frank Shepard devised a means of tracking legal precedent now named for him. To shepardize a case is to identify other cases and authorities that have discussed the case being shepardized. This will tell you, for example, if the case has been overturned or sustained. In 1961 Garfield applied this idea to scientific journal articles and created the Science Citation Index. Why would you want to use a citation index? Suppose the only source you have for your paper is an old article. You can't find anything else on the topic. Wouldn't it be nice to have a list of all the other articles that have interacted with the one article you have? Or suppose there are thousands of articles on your topic but one article has some very unusual arguments. Wouldn't it be nice to have a list of all the other articles that have interacted with the one article? There are no citation indexes for theology but several for the social sciences. Counseling majors at DTS will recognize a citation indexing function in the Psycinfo database.

text online. But you may need to increase precision, not just reduce the number retrieved. And increasing precision can be a difficult task.

1. If your initial set is the result of AND-ing terms, consider using proximity operators or searching for exact phrases to increase precision.

2. If your initial search used keywords, and you discover a subject heading that closely matches what you need, then re-execute the search using the exact subject heading with field set to subject. (In some databases, a subject search may include words from both subject headings and titles.) No subject heading expressed what you need? Sometimes there is a stock phrase used in titles or abstracts or in the full text of the articles that does accurately reflect the concept. In such a case, a title search for that phrase would be better than a search of the subject heading.

3. If there are still too many citations resulting from a subject search, then further reduce the set by requiring some of the same key terms appear as both subject and title words. This will probably also increase precision. If you are searching a free-text database that indexes every word in the documents, then limiting to title words almost always increases precision while decreasing recall.

4. If this set is still too large, and you can think of additional precise terms, then require that some of these terms also appear in the final set. In other words, AND another term or set to your previous set. Repeatedly add additional terms until it is a manageable size. This is sometimes called the "successive fractions" approach.

5. Abandon that database for the moment. Find some other source that will provide a short list of "best books" on the topic. How? Consult the bibliography in a textbook or the bibliography in an encyclopedia article or other reference book.

How to Retrieve More

When your search retrieves insufficient material, take the following steps.

1. First look for minor errors. Check spelling/typing and operator syntax. Spelling errors are the main cause of "zero hit" searches (searches that retrieve nothing).

2. Specify additional synonyms. Think broadly. They may be subject headings, title phrases, or call number ranges (in WorldCat, for example). What you do will depend on the specific situation. Sometimes there is an obvious solution; often there is not. The section on "How to find good/better search terms" tells you how to find synonyms. "OR" synonyms together.

3. Be less restrictive. As needed, relax logical, proximity, and field constraints in your search statement.

4. If your database searches metadata (author, title, subject) by default, but has an option to search the full-text of documents, try that option.

5. Consider using a different database or a print bibliography. Use Google Scholar and Google Books.

Do You Need Precision or Completeness?

How you search depends on how thorough you need to be.

Ordinary assignments require you to examine a relatively small number of good sources (e.g., between 5 and 10). You will often use only three main databases (Atla Religion Database, WorldCat and Google). For each database you will execute an initial exploratory search, examine results, discover helpful search terms, and then execute a second better search which uses controlled vocabulary terms or quoted phrases. Your search statements will be simple (just a few terms). You won't make much use of operators, truncation, etc. Your focus will be on the right search terms.

Some assignments (e.g., a thesis) require you to examine a much larger number of sources. You may use many special purpose databases. You may examine many hundreds of citations to find hidden treasures. For each database you will execute several exploratory searches, examine results, discover terms, test terms, and eventually execute improved searches. You will use both controlled vocabulary terms and natural language words and phrases. You will "OR" synonyms. You will use parentheses to group terms.

Here is an attempt to be both precise and somewhat thorough at the same time. We want articles about designing curriculum for STEM (science) degree programs in a higher educational context.

> ("curriculum design" OR "curriculum development" OR "curriculum planning")
>
> AND
>
> (stem OR science OR technology OR engineering OR math*)
>
> AND
>
> ("higher education" OR "postsecondary education" OR "graduate study" OR "undergraduate study" OR college OR university)

It attempts precision by using phrases (e.g., curriculum design") and especially by combining three concepts together (curriculum design; STEM; higher ed). It attempts to be thorough by using multiple ways of expressing each concept. This is a reasonable compromise. However, a serious attempt to be through would involve many more terms. And that would sharply increase the number of irrelevant items retrieved.

A Warning about Automatic Processing

Sometimes a search engine will automatically do things we didn't ask (or expect) it to do. For example, the software might automatically perform pluralization (looking for both singular and plural forms) or stemming (looking for words that match the root or stem of a search term) or synonym expansion or substitution (automatically looking for additional synonyms). These automatic features are sometimes un documented. And they may produce unexpected results.

On balance, you will probably benefit from various automatic features. But such features should never operate in secret, and they should always be optional.

- A search engine should always tell you what it has done. If the system does not tell you what it has done, you cannot easily understand why you got the results you got, and it can be difficult or even impossible to revise and improve the search. Unfortunately, some vendors consider such automatic features to be proprietary secrets that give them some advantage over competitors. They provide little or no documentation of the features so competitors cannot copy them.

- A search engine should always allow you to take control so you can help it or bypass it as needed. Few vendors offer a way to turn off automatic features. They have much confidence in their own software and little confidence in searchers.

Miscellaneous

Most systems treat punctuation as spaces/blanks. So the abbreviation N.T. (as in N.T. Wright = Nicholas Thomas Wright, or as in N.T. = new testament) might be treated as 'N T' for example.

Most systems ignore case. However, some systems require operators or field codes to be uppercase.

Bible searches may require special fields or special notation. This varies from database to database. Bible searches in WorldCat can be complex because WorldCat mixes records from many sources that follow different conventions for citing bible passages. The tutorials on the Atla Religion Database include much space on this issue.

Pay special attention to how the search engine deals with a zero hit (nothing found) situation. Some systems will tell you they found nothing. Other systems will reformulate the search and try again for an approximate match.

Know the default operator of the search engine you are using.

Don't hesitate to ask for help. Library reference staff are available by email, phone and in person.