

Search Strategies: Internet vs. Library Databases

Searching the Internet, Searching Library Databases

People are so accustomed to searching the Internet that they assume that searching library catalogs and library databases is going to work exactly the same. This couldn't be farther from the truth. Here's why:

- The Internet is actually a conglomeration of thousands of unrelated networks of computers that share one thing in common: they are all able to communicate with each other using a shared set of computer protocols. Other than this, there is no inherent organization to the Internet that makes it particularly easy to navigate. For this reason, we have search services like Google.
- Internet search services such as Google regularly pull content from the Internet and create indexes of the content that can then be searched much like a database can be searched. Web page designers who adhere to standards provide contextual clues in their documents that make finding their documents via a search service like Google easier. The problem is that much of the content on the Internet does not adhere to standards.
- Indexes of Internet content are created by computers and are largely unexamined for accuracy by people who use the indexes. Anyone who has searched for information using any Internet search service will notice all kinds of strange matches in their search results. This is because the indexes that help them find pages are created wholly by computers and have not been evaluated for relevancy by humans. Therefore, many results in any search listing will not be at all what the searcher expects.
- Library databases are designed by professionals to make finding information a more precise process. A library's catalog is based on specific description, following very well defined rules, of each and every item contained in a library's collection. A researcher who knows how the system operates will be able to eliminate non-relevant materials with relative ease. This goes also for research databases that provide access to articles from journals, magazines, and newspapers. In each of these types of systems, professional catalogers or indexers have examined materials covered in the databases and have provided descriptions that a searcher can take advantage of to bring more precision to a search.
- Library databases and research databases make use of "controlled vocabularies." Psychology has its own vocabulary. Sociology has its own vocabulary. Art, music, history, all have their own vocabularies. A database that focuses on identifying materials useful to history researchers will make use of the professional vocabulary for history, thus making it easier for a historian to find what's needed on a particular topic.
- Library databases may or may not include the ability to search the full text of their covered publications. Systems like *ProQuest* and *EBSCO* do provide full text search capabilities, while other research databases do not. Systems that do not search full text rely heavily on indexing and abstracting to provide researchers with results.



Even though there are fundamental differences in structure and coverage among databases and the open Internet, researchers using any search system will benefit from thinking carefully about what they're searching and about how they construct their search strategies.

Constructing Searches -- Google/Database Comparisons

Constructing a search in Google can be very different from constructing a search in a research database, although many databases are now moving to a Google-type interface for basic searching. Google is designed to return results based on any or all of the words entered into the Google search box. A Google user does not need to specify connections between the terms and doesn't have to follow any particular rules to come up with some kind of result listing. Many databases, on the other hand, work better using logical connectors between search words. The most commonly used connector is the Boolean connector AND. Most databases now assume that multiple terms entered into a search window should be matched, so the search strategy defaults to an AND search, even though it is not apparent to the searcher.

The following example compares a search executed in Google and a search executed in the UNF Library's OneSearch system.

The image displays two side-by-side search results for the query "solar water heaters costs".

Left Panel (Google): Shows a typical Google search interface. The search box contains "solar water heaters costs". Below the search box, it indicates "About 29,000,000 results (0.27 seconds)". The search results include a link to "How Much Will Solar Cost? | sunshineisfree.org" with a snippet: "Calculate cost, rebates, tax credit Pre-Screened Solar PV Contractors".

Right Panel (UNF OneSearch): Shows the UNF Thomas G. Carpenter Library's OneSearch interface. The search box contains the same query, but it is enclosed in quotation marks: "\"solar water heaters\" costs". Below the search box, there are radio buttons for "Keyword" (selected), "Title", and "Author". The search results are not visible in this panel.

Annotations:

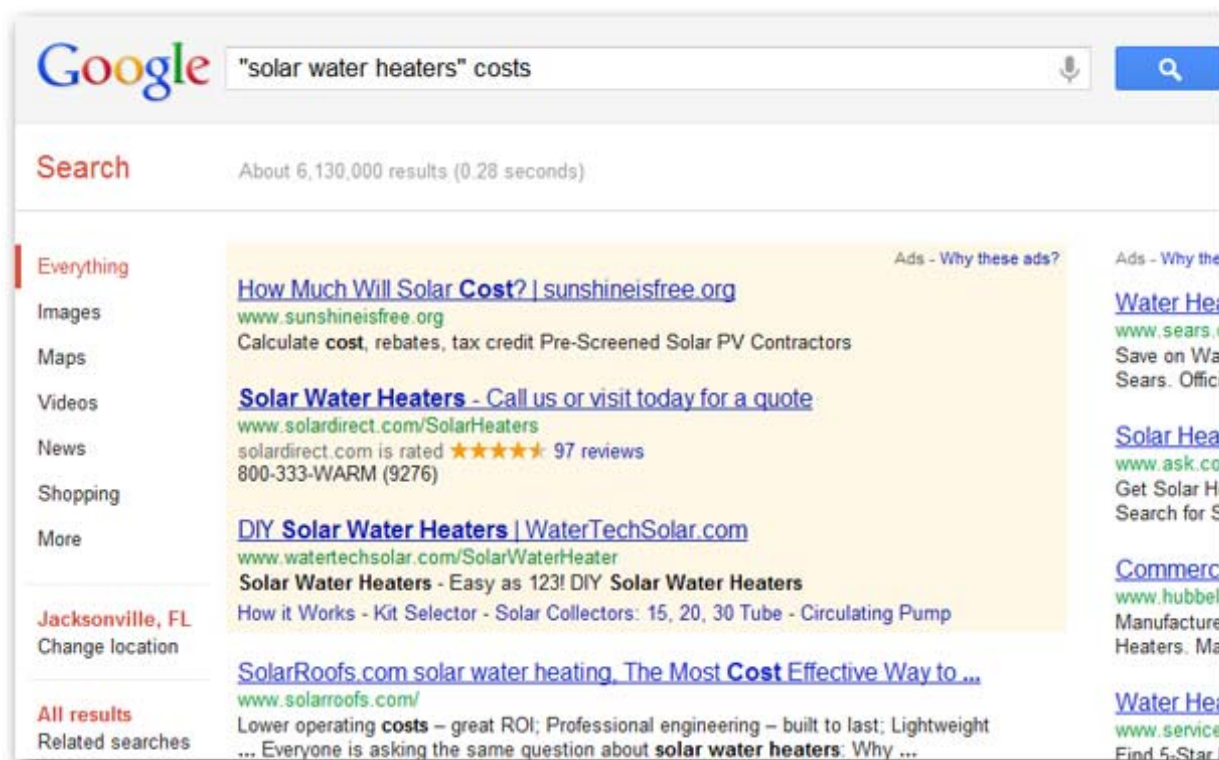
- A text box on the right side of the Google search results reads: "Typical Google search for costs associated with the use of solar water heaters."
- A text box on the left side of the UNF OneSearch search results reads: "OneSearch strategy for costs associated with the use of solar water heaters."

The Google search is entered as most people would normally enter a search in Google. In other words, a string of words has been entered to describe the main topics being searched. Entering the same strategy in OneSearch will also work, but the example shows the use of phrase searching to more precisely match the phrase "solar water heaters." Behind the scenes, OneSearch is constructing a search using the following search logic: "solar water heaters" AND costs.

The result of the Google search pictured above is a listing of 29,000,000 web pages. The result in OneSearch is 5500 items coming from a variety of publications, including newspapers, magazines, journals, and books. One of the major differences in results between Google and OneSearch comes from the fact that OneSearch focuses on published materials and not on web pages. Another difference comes from the use of phrase searching in OneSearch. Searching "solar water heaters" as a phrase results in fewer matches than searching the three words separately. Both searches result in way too many results to be practical and would benefit from some fine tuning. Most researchers are not going to want to examine 29,000,000 websites nor will they want to look through 5500 library items, so finding a strategy that works better in both systems will be of benefit.

Google Search Using Phrase Searching

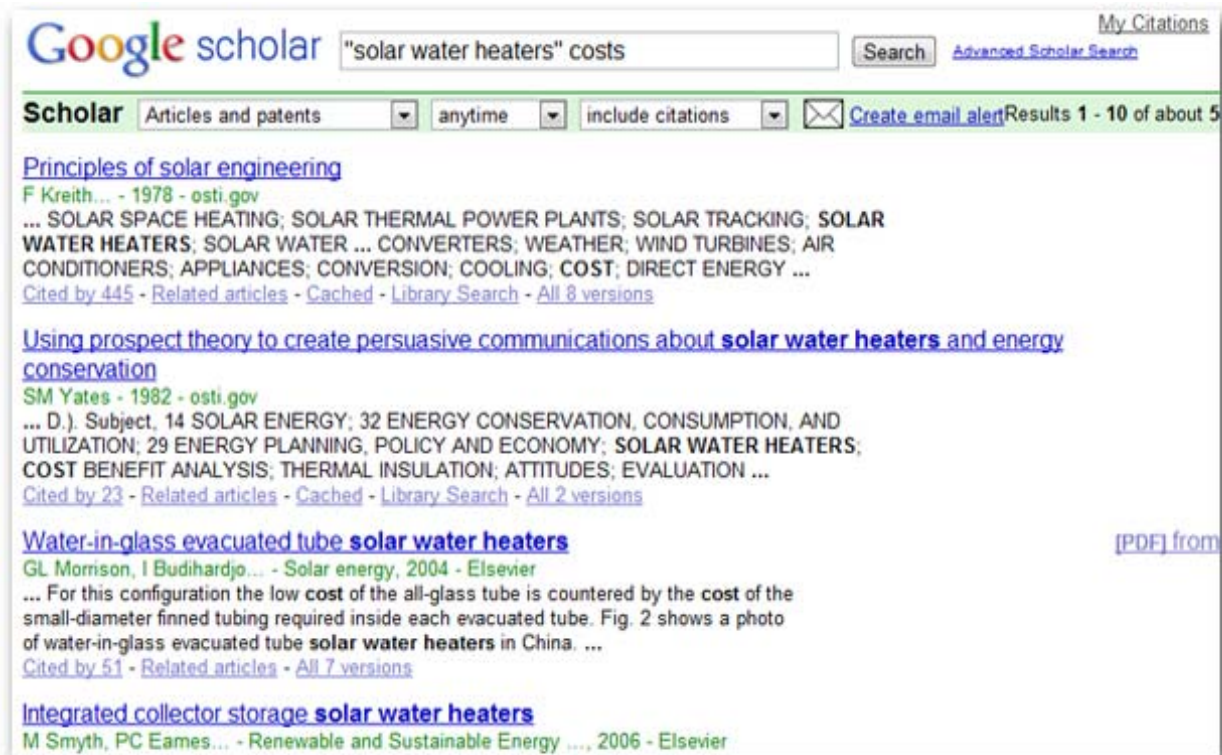
Google searching also accommodates the use of phrases. Just as in library databases, search results can be trimmed by enclosing phrases inside double quotation marks (" "). Picture below is a revised Google search specifying that solar water heaters be searched as a phrase.



Revised Google search strategy -- phrase searching

Notice that the search results now number 6,130,000 instead of 29,000,000, still a huge number, but considerably fewer items to go through. Getting to this point was as simple as enclosing the phrase solar water heaters in quotation marks. This indicates to Google that it should only bring back web pages that contain this exact phrase and the word costs, if possible.

Further reductions are possible by using Google's Advanced Search panel. A searcher can limit by date and even by language if need be. Further reduction can be gained by using Google Scholar instead of Google. Scholar searches only scholarly websites and includes articles that can be located in some database systems. The same search of Google Scholar results in 2,050 matching web pages, clearly a huge reduction in the number of materials a researcher will need to review.



Google Scholar search using phrase searching.

The Google examples above demonstrate the importance of thinking about how to construct a search. Database searches also benefit from careful wording of the search. With both Google and library databases, the better the choice of words used to describe what a researcher is looking for, the better the results of the search. The huge difference in results brought back by Google and by OneSearch can be explained simply: Google searches the full text of all searchable Internet content; databases search descriptions and sometimes full text of journals, magazines, and newspapers selected for inclusion by the individual databases. While OneSearch covers a majority of the library's databases, the content that it searches differs from what Google searches. OneSearch specializes in searching published materials instead of web pages.

Constructing Searches -- Using Database Organization to Construct Better Searches

The UNF Library subscribes to numerous databases and offers researchers the means for selecting databases that focus on a very specific discipline. For example, engineers might be interested primarily in searching only the engineering literature and might not want to search all of the library's databases for information. The *ProQuest* system, for example, offers a number of science and technology databases for searching and specifically accommodates researching just engineering, if that is what the researcher wants to do. The *ProQuest Engineering Research Database* will target specific systems that focus on engineering resources.

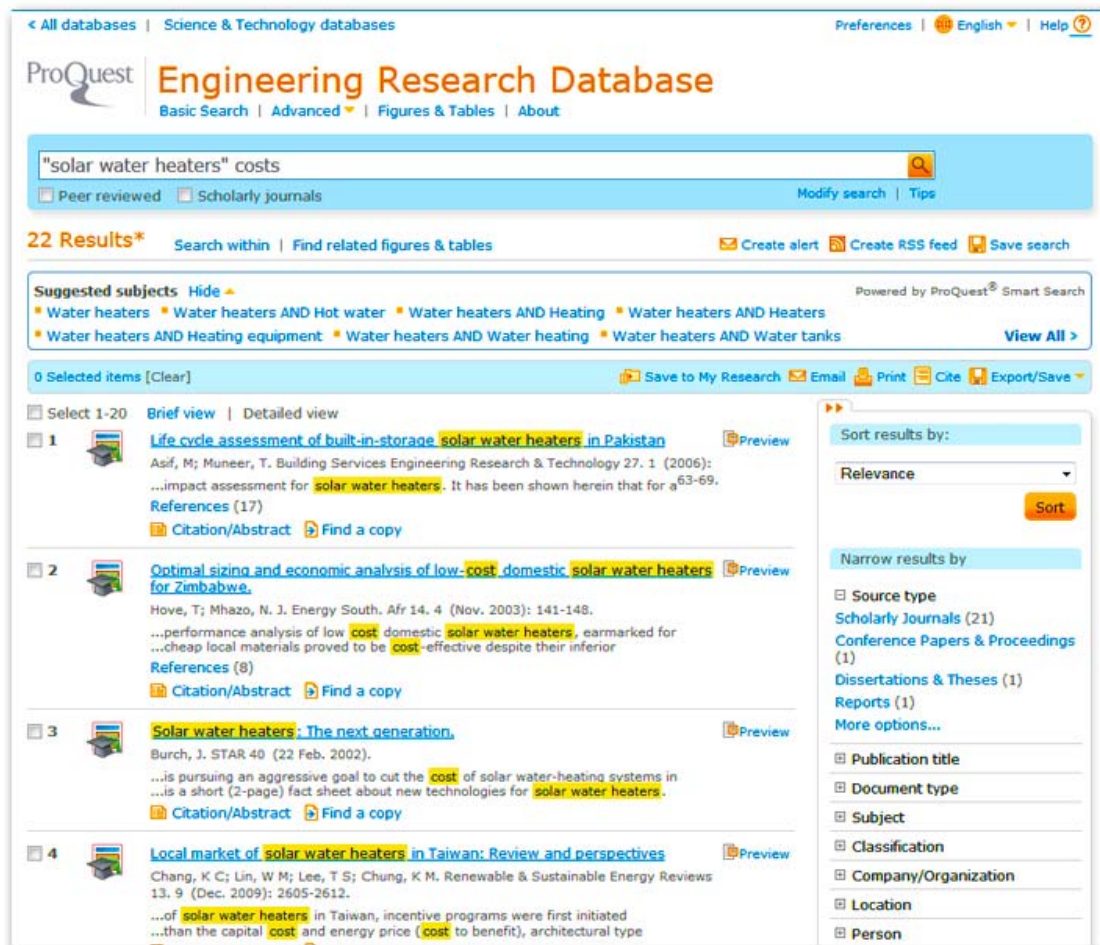
Further examination of the *ProQuest Engineering Research Database* will help to illustrate some of the methods that can be used in databases to get more precise results. First of all, knowing what is covered in a database will help drive a researcher's choice of that database over another. The *ProQuest Engineering Research Database* covers professional journals and conference papers that focus on engineering and includes access to the following systems:

Abstracts in New Technology & Engineering
Biotechnology and BioEngineering Abstracts

Biotechnology Research Abstracts
Civil Engineering Abstracts
Earthquake Engineering Abstracts
Environmental Engineering Abstracts
Mechanical & Transportation Engineering Abstracts
ProQuest Deep Indexing: Engineering

Note that these databases are labelled with the term "abstracts." An abstract is an article summary or overview. While these systems, themselves, do not directly provide full text, they do provide article summaries and other evaluative data that will help a researcher decide what articles are best. The library provides a linking system (LinkSource) that will lead the researcher to the full article if it is available.

The previously used search ("solar water heaters" costs) results in a listing of 22 articles in *ProQuest Engineering*, most of the sources coming from journals. The illustration below shows the search strategy and the result screen in the database.



One of the things that databases are really good about is showing researchers what types of published materials are available. In the result screen above, the ProQuest system divides results into scholarly journals, conference papers & proceedings, dissertations & theses, and reports. This makes it easier for a researcher to review the types of materials available at a glance. Many databases also include newspaper articles, magazine articles, trade journal articles, and other types of sources. If a researcher is primarily interested in retrieving journal articles, having the articles sorted by types, like in ProQuest, makes it easier to select the most appropriate types of resources.

Life cycle assessment of built-in-storage solar water heaters in Pakistan

Asif, M; Muneer, T. *Building Services Engineering Research & Technology* 27. 1 (2006): 63-69.

[Hide highlighting](#)

Abstract (summary) [Translate](#)

Sustainability of building construction and its services is increasingly drawing attention. In this respect CIBSE has issued a number of calls to industry. Solar water heating systems are now economically competitive and are being used throughout the world. Following earlier work of the present authors on the life cycle assessment (LCA) for building components, this work presents a monetary, energy usage and environmental impact assessment for **solar water heaters**. It has been shown herein that for a developing country like Pakistan where large amounts of hot water are required within industry, solar water heating offers favorable payback periods on each of the above three counts. It has been estimated that under the tested conditions, the embodied energy payback period for the plain and finned heater is equivalent to 185 days and 169 days respectively. It has also been found that the amount of carbon released during the construction process of plain and finned water heaters is 27.9 kg and 29.7 kg respectively. Annual carbon saving for the plain and finned water heaters has been calculated to be equivalent to 59.5 kg and 69.6 kg respectively, which in turn gives a carbon payback period of 156 days and 171 days. The monetary payback period for the plain and finned heater has been found to be equivalent to 6.7 years and 6.1 years respectively.

Indexing (details) [Cite](#)

Subject

- Environmental impact assessments;
- Equivalence;
- Life cycle engineering;
- Life cycle assessment;
- Construction;
- Construction **costs**;
- Hot water;
- UK;
- Counting;
- Sustainability;
- Mathematical analysis;
- Water heaters;
- Carbon;
- Heaters;
- Economics;
- Heating equipment;
- Construction industry;
- Water heating;
- Drawing

Another feature of databases is the use of "controlled vocabularies" that rely on terminology that makes sense to professionals who work in particular fields. For example, engineers will have a firm grasp of the terminology used in engineering and will likely rely on that terminology in conducting reviews of published literature. The ProQuest system shows terminology assigned to each article in the full article description. Many databases call these terms descriptors. In the example above, the article description provides a number of subject terms that could be useful in a search. Useful terms in this article would include "construction costs," "water heaters," and "water heating." A researcher might also be interested in researching "sustainability" and "environmental impact assessments." A researcher looking specifically for construction costs related to the installation of solar hot water heaters in residential construction can make use of the descriptors to find articles that target his or her particular interests.

"Advanced" Database Searching

Most databases offer researchers more options for searching than the single line basic search. OneSearch offers an "advanced search." The *ProQuest* databases also offer advanced searching options. One thing that a researcher should keep in mind about the so-called advanced searching available in

most databases: ADVANCED does NOT mean more difficult; it means that the database provides additional options for getting better search results.

The Advanced Search in ProQuest accommodates entering multiple search terms, the use of Boolean connectors (AND, OR, and NOT), the selection of specific types of resources (journals, conference reports, etc.), and a number of other means for further limiting the search results. A researcher can begin with a basic keyword search, review relevant articles, take note of important subject terms (or descriptors) used in the articles, and then use the advanced search to come up with better results. The following illustration demonstrates how to conduct a new search using subjects discovered in a relevant article.

The full article description in ProQuest includes all the information necessary to document the article in a bibliography: author's name, article title, journal title, volume, issue, date, and pages. Additionally, ProQuest often also provides contact information for the author(s), author affiliation(s), subject descriptors used to categorize the article, an abstract of the article, and any additional information that might be useful to a researcher. The abstracts and subject headings can be particularly useful in helping a researcher identify terminology that is used in the field and in constructing a new search strategy that yields better results.

The example search strategy below makes use of subjects discovered in the basic search and includes the terms "water heating," "construction costs," and solar. These could have been entered as subjects by using the selection boxes to the right of each term and switching from "All fields" to "Subject heading." The OR options on the second and third search lines would allow plugging in synonyms or similar terms. For example, if the researcher was interested in either "construction costs" OR sustainability, the term sustainability could be plugged in on the same line as "construction costs" following the OR and the search would pick up articles that had either term.

The screenshot shows the ProQuest Engineering Research Database Advanced Search interface. At the top, there are navigation links for "All databases" and "Science & Technology databases", along with "Preferences", "English", and "Help". The main header includes the ProQuest logo and "Engineering Research Database" with sub-links for "Basic Search", "Advanced", "Figures & Tables", and "About". The "Advanced Search" section features a search form with three rows of input fields. The first row contains the text "water heating" in a search box, followed by "in" and a dropdown menu set to "All fields (no full text) - ALL". The second row starts with "AND", followed by a search box containing "construction costs", "OR", and another search box. The third row starts with "AND", followed by a search box containing "solar", "OR", and another search box. Below the search boxes are links for "Add a row" and "Remove a row". To the right of the search boxes are links for "Thesaurus", "Field codes", and "Tips". A "Search" button and a "Clear form" link are at the bottom right of the search area. Below the search form is the "Search options" section. It includes "Limit to:" with checkboxes for "Peer reviewed", "Scholarly journals", and "Latest update". "Date range:" is set to "All dates". "Source type:" has a "Select all" checkbox and a list of source types: "Books", "Conference Papers & Proceedings", "Dissertations & Theses", "Government & Official Publications", "Other Sources", "Reports", and "Scholarly Journals". On the right side of the "Search options" section is a "Search subject areas" section with the text "Use search forms customized for each subject." and four categories: "The Arts", "Business", "Dissertations & Theses", and "Health & Medicine", each with a small representative image.

As is, the above strategy results in 7 articles that all focus on "water heating," "construction costs," and solar. If this search is too limiting, the researcher might try other subject terms discovered in the abstracts and subject headings to construct a new search strategy.

Depending on the strategy selected, results might increase or decrease. For example, a search on "water heaters" alone will likely result in more items retrieved. A search on "water heaters" AND "solar energy" will result in fewer retrieved items. As a researcher works with a database, he or she will find it useful to keep track of key terminology that keeps cropping up in article descriptions as a means of identifying additional search strategies for finding more information or for further refining the search topic. The better armed a searcher is with vocabulary to describe the topic of a search, the more likely the searches will be successful.

Shaping Questions Into Search Strategies

All of the above discussion has revolved around isolating key phrases and words that would adequately describe a topic in a search. Note that none of the strategies was phrased in the form of a question. This was quite deliberate. While database designers are constantly working toward building search interfaces that work with natural language, realistically, most databases rely on logically combining words and phrases rather than expressing ideas in complete sentences. Researchers should spend some time thinking carefully about their topics and how to phrase their questions about the topics before launching into database searches. A good strategy to use is to start with a research question or questions and to make a list of all of the key concepts that the question(s) might touch upon.

Research Question

Possible Search Key Words and Phrases

Do celebrity endorsements of products influence consumer purchases?

"celebrity endorsement"
 consumers
 "consumer behavior"
 "product endorsement"
 "consumer influences"
 "buying habits"
 advertising
 celebrities
 "movie stars"
 "rock stars"
 "professional athletes"
 "purchasing decisions"

How effective are municipal recycling programs in reducing landfill waste?

recycling
 landfills
 "municipal recycling programs"
 "municipal waste"
 effectiveness
 "citizen participation"
 "recycling programs"
 "citizen buy-in"
 "solid waste recycling"
 management

Beginning with the list of key words and phrases, think about logical connections between the ideas and how to phrase these for use in a database. Most databases make use of the Boolean connector AND as a means of stringing together two or more concepts in one search. A few examples should help to clarify.

Possible Search Strategies for Celebrity Endorsements Question

"celebrity endorsement" AND "consumer behavior"

"celebrity endorsement" AND "purchasing decisions"

"professional athletes" AND "purchasing decisions"

Possible Search Strategies for Municipal Recycling Question

"solid waste recycling" AND effectiveness

"municipal recycling programs" AND "citizen participation"

landfills AND recycling AND management

A researcher should try a number of strategies, carefully reviewing results along the way. As relevant results come up, the researcher should review key terminology used in the articles, in the article abstracts, and in the database subject headings or descriptors with an eye toward identifying other ideas for constructing related searches. Not every search strategy is going to work, so the better the researcher's grasp of the topic and its related terminology, the more likely he or she will be in eventually finding materials suitable for the topic being scrutinized.

Reading by Jim Alderman. Updated March 2014.