

TIPS ON SEARCHING THE INTERNET FOR MEDICAL INFORMATION

Ibrahim Mansoor, MBBS, Department of Histopathology, King Abdulaziz University Hospital, Jeddah, Saudi Arabia

البحث عن المراجع يعتبر جزء من العمل اليومي في مجال الطب، و نظراً لانتشار خدمة الإنترنت، فقد أتاح ذلك فرصة قيمة للأطباء للوصول إلى المعلومات بأسلوب دقيق و سهل و سريع. و لكن لوجود معلومات كثيفة و ضخمة على الإنترنت فقد أصبح الوصول إلى معلومة معينة عملية صعبة و مرهقة في هذه الأيام. هذه المقالة توضح الطريقة المثلى و الفعالة باستخدام برامج مساعدة و بعض طرق البحث المفصلة للوصول إلى المعلومات المطلوبة في مجال الطب. و كذلك يقدم بعض قوائم المعلومات المفيدة و مواقع قواعد البيانات التي تساعد كثيراً للوصول إلى المعلومات الصحيحة.

المصطلحات: إنترنت، معلومات طبية على صفحات الإنترنت، برامج البحث.

Searching for references is part of everyday life in medicine. Since the arrival of the Internet, it has provided great promise for clinicians because of its ready provision of access to large amounts of knowledge and information. But because of the overload of information, searching for particular information has now become a tedious time-consuming and frustrating task. This article describes effective ways, tips, tools, detailed search techniques and strategies for searching for medical information. It also lists some useful resource and database sites that can help in the search for accurate information.

Key Words: Internet Search, search tools, World Wide Web.

INTRODUCTION

We sometimes feel that our search on the Web is not as effective as it should be. Are there times when we never seem to find the information we want even though we know that it exists somewhere on the Web? I have tried to solve this problem by summarizing the internet search tools and the tips for searching information on the internet. Various specific medical information and search-related problems are briefly discussed at the end.

Finding information on the Web may be difficult, but not impossible. As an alternative to a central catalogue, the Web offers a choice of dozens of different Search Engines, each with its own database, command language, search capabilities, and method of displaying results.

The Search Engines, also known as Search Services, find documents to match your interests. Each search engine operates on its database of URLs (Uniform Resource Locators), texts and descriptions that point

Correspondence to:

Dr. Ibrahim Mansoor, 6112 North Harding, Chicago, IL 60659, USA.

to the actual documents on the World Wide Web.¹⁻² It must be pointed out that whenever a search is made with the help of a search tool, what is viewed is data extracted from the database of this search tool and not from the whole World Wide Web. Since none of these search tool databases includes the whole World Wide Web, different results are obtained from different search tools. All search tools provide the search results as lists of Web documents with hypertext links, which when clicked take us to that particular Web document from the search tool.

The search tools on the Web fall into two main categories: Subject Directories, which rely heavily on the human element as part of their indexing strategy, and Search Engines, which keep human/data interaction to a bare minimum. Both use software robots called "Spiders" that crawl the Web, newsgroups, and gopher, FTP (File Transfer Protocol) and WAIS (Wide Area Information System) sites, extracting URLs addresses and keywords to add to the search tool's database. Both of these search tools have benefits and drawbacks, depending on what you are willing to sacrifice.²

New MetaCrawlers have now emerged, as the best for querying multiple engines at once. They do not maintain their own database; instead, they act as middle agents passing queries to many major search engines.³⁻⁵

TIPS FOR SEARCHING

The following are brief and very effective search tips that apply to most of the search tools available on the internet:⁴⁻⁵ (1) The accuracy of the search depends on the greater number of search words used. (2) Good keyword selection is as much about excluding the irrelevant as it is about including the relevant. Being as specific as possible will yield the best results. (3) Since no single search tool will supply all your

needs, use at least two or three different tools regularly. Learn the features and the capabilities of these tools to use them effectively. (4) It is wise to check more than one search tool for any topic because search results vary widely from one to another. (5) If you are more interested in broad, general information, the first place to go to is a Subject Directory. If you are after narrow, specific information, a Web Search Engine is probably a better choice. (6) Most of the search engines return results with confidence or relevancy rankings. In other words, they list the hits according to how closely they think the results match the query. Consequently, it is often not necessary to browse through more than the first few pages of results, even when the total results number in the thousands. (7) Many search engines provide two different interfaces for searching the internet; the Basic and the Advanced. The Basic or Simple search interface is a good place to start your search but it lacks many of the search engine's features. If you are looking for information that is hard to find, you can search more effectively by using the advanced search capabilities of the search engines. (8) Wherever possible, use Boolean commands in your search query. Boolean commands are specific words or symbols that allow you to include, combine, or restrict the keywords of your search. Some search engines will allow you to use Boolean commands only from advanced search interface. (9) Some search engines do not support Boolean commands directly. They use the characters instead of Boolean operators to include and exclude terms. Most search engines will allow you to use these Implied Boolean commands. (10) Each engine catalogs information in a different way. Knowing how each engine works helps to use the right search engine for the job on hand.

AN EXAMPLE OF A PRACTICAL EFFECTIVE SEARCH TECHNIQUE

Here is a list of easy to follow techniques and strategies explained with an example that can definitely boost your search engine's performance. (1) State What You Want to Find: In one or two sentences, state what you want to find on the internet. For example: What are the recent findings about a new drug therapy discovered for cancer treatment? Using the information "contained" in these statements, you can see how an effective query can be built by following the guidelines. (2) Identify Keywords: Break down the topic into key concepts and underline the main concepts in the statement. What are the recent findings about a new drug therapy discovered for cancer treatment? (3) Use nouns as query keywords: When conducting a search, the central keywords in your queries will be nouns. Though sometimes adverbs and adjectives can help refine your search, the key pivot point is a noun, or series of nouns. In our example, the noun is Drug, drugs. Actions (verbs) and modifiers (adjectives, adverbs, predicate subjects) are very diverse, easily substitutable, and generally not universally applied in any given description. Search engines either return too many "hits" for these words that are not very useful or "throw them away". As a general rule, try to avoid using action terms and modifiers in your queries. (4) Use sufficient number of keywords in query: One of the major mistakes usually made in preparing a query is not providing enough keywords. On average, most users submit 1.5 keywords per query. This number is insufficient to find accurately the information you are seeking. Thus, a central task in query formulation is for you to identify a sufficient number of appropriate keywords. In our example, the possible keywords are new, drug, drugs, therapy, treatment, cancer, and neoplasm.

(5) Truncate words to pick up singular and plural versions: One of the mistakes in query formulation is inadequate use of word stemming, or truncation. By using either only singular or plural version of a word, we would eliminate about half of the potential documents that we would like to use as our search basis. The better way to handle this problem is through truncation. Truncation is applying a wildcard character after the first few letters in a term (the "stem"). The asterisk (*) is the most universally accepted truncation wildcard. This wild card means any word or letter after this. Generally, you must also have a minimum of three characters at the beginning of the word as your stem basis. Once marked for truncation, any matching characters after that will be picked up in the search query. In our examples, the keywords that can be truncated are as follows: Drug*, discover*, cancer*, treatment*. (6) Use synonyms and variant word forms: Another way of increasing your search effectiveness is to be as specific as possible; that is include as many terms and synonyms as you can think of to describe your topic fully. The best synonyms provide relatively complete coverage for the subject at hand and are "pitched" for the right informational objective. In our first example, the possible synonyms, alternate spellings, and variant word forms of each keyword are as follows: drug*: therapy, treatment; cancer*: neoplasm, malignant; discover*. A thesaurus, and a dictionary, are both worthwhile sources of synonyms for the major subject(s) in your query. (7) Combine keywords into phrases where possible: A very effective way to increase the relevance or precision of "hits" is to search as a phrase. Phrases are combinations of words that must be found in the search documents in the EXACT order as shown. You denote phrases within closed quotes (""). Phrases should be used where the constituent terms

are naturally married; like "lipo-protein" or "lipoprotein*". It is a powerful search technique for significantly narrowing your search results, and it should be used as often as possible. Some other examples are "Ischemic heart disease", "Diabetes Mellitus", "Downs syndrome" etc. When using phrases, it is important to consider nuances of the phrase that wouldn't normally be of concern. For example, the spaces between words are as important as characters. Some search tools provide specific options for phrases; some do not allow them at all, but almost all allow you to enter a phrase in quotes, ignoring the quotations if not supported. (8) Combine Synonyms with Boolean or Use Boolean or to string together synonyms. For example discover* or find; popular or common or favorite; treatment* or therapy*. (9) Combine 2 to 3 "concepts" in query: Triangulating on multiple query concepts narrows and targets results, generally by more than 100-to-1. For example, the concepts in our examples could be as follows: "Cancer therapy"; "new drugs*"; discover* or find; method* or way* or technique*. (10) Distinguish "concepts" with parentheses: Nest single query "concepts" with parentheses. A simple way to ensure that the search engines evaluate your query in the way you want, from left to right: ("Cancer therapy") ("new drugs*"). (11) Order "concepts" with subject first: Put main subject first. Engines tend to rank documents more highly that match first terms or phrases evaluated: ("new drugs*") (discover* or find) ("Cancer therapy"). (12) Link "concepts" with the AND operator: Combine keywords with Boolean AND. AND glues the query together. The resulting query is not overly complicated nor nested, and proper left-to-right evaluation order is ensured: ("new drugs*") and (discover* or find) and ("cancer therapy").

SPECIFIC MEDICAL INFORMATION AND REFERENCE COLLECTION TIPS

Now in discussing the search for medical information specifically, I have tried to discuss and solve a few problems which may arise while looking for references and articles on Web-based medical data-bases and Medline's. One of the most useful and frequently used Medline databases is the PubMed Medline. It is sometimes necessary to limit the results of a PubMed Medline search to those articles that are freely available in full-text on the Web. It is now possible since the Hardin Meta Directory of Internet Health Resources <<http://www.lib.uiowa.edu/hardin/md/>> has compiled a list of full-text medical journals (with an emphasis on those that are indexed by Medline) that are freely available on the internet. In addition to simply providing an A-Z list of these titles, however, the teams at Hardin have added all the relevant ISSNs (unique journal numbers) into a pre-formatted Medline search. On clicking on this link, <<http://www.lib.uiowa.edu/hardin/md/ej.html>> a Medline search which rings together in one result set all Medline articles that are available in full-text is executed. All that is required is to add-in the search term(s) you are interested in and combine both sets of results. Using this method, to search for papers discussed on the internet, I was directed to articles in titles such as the BMJ, Emerging Infectious Diseases and the Proceedings of the National Academy of Science.

It is sometimes necessary, on the internet, to exclude 'personal' home pages from the results-set to make the list concise and more useful. Probably the easiest way to achieve this is to use the Power Search option at the Northern Light search site <<http://www.northernlight.com/power.html>>. Through the use of simple check boxes, you can, for example, specify that only pages that reside

in the education or government domains are identified in the search results. The search page in this site also allows you to limit your results by date and language.

One method in use at present by researchers is to find high-quality papers somewhere and to make use of cited references in that article – so that this good paper leads to another set of good papers on the same subject in an effective way. Something similar can be done on the Web, though it is not as reliable or as finely tuned as the cited reference approach. If Netscape is used (version 4 and above) a ‘What’s Related’ button on the toolbar will be seen. On clicking on this icon, Netscape will identify sites that are in some ways related to the one currently on view. This feature, developed by Alexa, works by examining the links on the page to identify related sites, and by looking at how users move from site to site. For example, relationship is likely if thousands of users go directly from site A to site B. Users of the Internet Explorer can also make use of this feature, but only if they available at: <http://www.alexa.com/> From the PubMed Medline site, the ‘What’s Related’ link points to the BMJ, NEJM, Medscape and the Centers for Disease Control and Prevention.

Now-a-days, with so many meta-search tools available, it is unwise to use any single

search service such as Altavista or Google. Meta-search services such as Dogpile <<http://www.dogpile.com>> and Ixquick <<http://www.ixquick.com/>> allow users to simultaneously search multiple search services. Ixquick, for example, runs any given search against 14 separate search engines including AltaVista, Hotbot and Yahoo. However, though such services are useful - especially if you are looking for something that is relatively uncommon - they are unable to exploit the full power of advanced Boolean search that many search engines offer. Consequently, if a complex search that uses Boolean logic is required, nesting of terms etc, and meta-search tools should be avoided.

REFERENCES

1. S. Lawrence and C. Lee Giles, "Searching the World Wide Web," *Science Magazine* 1998;280:98-100
2. A Helpful Guide To Web Search Engines, Spider's Apprentice, [<http://www.monash.com/spidap.html>], pp 3-4.
3. Cohen, Laura, Searching the Internet, [<http://www.albany.edu/library/intemet/search.html>]
4. David Haskin, Power search, IW Labs, Internet World, Dec 1997 [<http://www.iw.com/print/monthly/1997/12/rwllabs.html>] pp 1-2
5. Elkody, Angela, Web Searching, Sleuthing and Sifting, [<http://www.angelfire.com/in/virtual/librarian/ismain.html>] pp 3-4.