How I Use It: Search Engines

The strength and at the same time the challenge of the Internet is the vast amount of free information available, which continues to expand exponentially. The surface web is the part of the web most people browse and although large, covers only a fraction of the available information in the World Wide Web. Even just using the surface web, it is impossible to find information without either knowing the Uniform Resource Locator (URL) for a specific web page or using a search engine.

Search engines such as Google or Alta Vista are programs designed to retrieve data published on the web. They work by sending out spiders or webcrawlers to fetch as many documents relating to keywords. The information on the Internet sites is then coded by a subsequent program and indexes it in a database (like filing books or journals in a library). These programs using metatags created using HyperText Markup Language (html) related to keywords to ensure the data is retrievable when a specific search is requested by someone using the search engine. Algorithms are used to direct the search and rank the information with the assumption that increased frequency and location of keywords have higher relevance in a search than those with keywords being used less frequently or not in a prime location on the website.

Metasearch engines such as Dogpile and Metacrawler are a special type of search engine that don’t create their own database but are programmed to search other search engine databases to provide broader searches. It should be noted that some search engines such as Google and Northern Light do not allow metasearching of their databases. When using search engines, it is useful to know what databases they are searching to ensure that the relevant data you require is accessible through that search engine.

The deep web refers to the content that is contained in searchable databases but can’t be accessed by the spiders used by normal search engines. These include library databases, commercial and specialized technical databases that may be indexed by people attaching specific metadata tags. Some metasearch engines such as Complete Planet and Pipl (which specifically searches for people on the deep web) are now overcoming these technological barriers to provide deeper searches in response to a query.

Cost

Search engine databases used for searching the surface web are usually free and are financed through advertising. Advertisers use a marketing strategy and bid on keyword phrases that are most relevant to their target audience. Metaengines and search engines for the deep web are also usually free of charge. Commercial or specialized databases may require a subscription or charge a pay to view fee for access. Ovid is a leading online medical and health sciences database which charges for access to medical journals.

Other associated costs for Internet searching are related to owners of web pages. Many companies position their web pages to be picked up by search engines and placed high on the ranking as part of their marketing strategy. Search engine optimization (SEO) is an Internet marketing service provided on a commercial basis that will make pages readable to search engines through relevant keyword use and emphasizing key topics related to the web pages’ content.

Uses

The first step of an effective search is to choose the right search engine for the type of information you want. For general information and with good keyword choice, Google, Alta Vista and Yahoo are effective. A broader and deeper search capacity can be conducted using Meta-search engines; however, more information is not necessarily efficient and can produce large amounts of information to be reviewed. Wider searches are often used when there is a requirement to broaden a search, for example, as part of an evidence-based literature review. Specialist search engines are often used for searching specific subject-related databases; these are regularly used in medicine and other professions, e.g. MEDLINE, EMBASE. This can be highly effective for identifying relevant material in response to a search query.

By clever use of keyword searches and Boolean algebra, you can construct effective search strategies for obtaining relevant information. Exact phrases can be used in searches by putting the keywords in quotation marks. Boolean logic using the operators AND, OR, NOT and NEAR can be used in some search engines to expand or reduce searches. You will often see Boolean operators used in search strategies for literature reviews of information indexed in specialized databases such as MEDLINE and EMBASE.

What are the risks?

The differences in programming and algorithms between search engines will result in the delivery of potentially different results in response to a keyword.
search. This is why searching two or more databases are recommended when performing a systematic review of the literature. Poor use of keywords, too specific or too wide a specific search strategy can result in an ineffective search that can waste time and fail to identify the data that is needed to answer the query. It is also important to verify the information (as discussed in previous articles in this series) and to know the source of the information so that you can rate its reliability. Alternatively you can cross reference using an alternative database but again this can be time consuming.

Links identified by the search engine in some cases might not be active as it is dependent on when the data was last verified in the database. In some cases, hackers may use the search engine technology to promote sites infected with their malware. This includes a technique known as ‘Blackhat SEO’. It should be noted that there is no correlation between search engine rating and security and that the use of sponsored links increases the risk of visiting unsafe websites. It is therefore important that you ensure that you have adequate security software on your computer. It has also been suggested that wherever possible you use a confirmed URL rather than a link identified by a search engine for added security.

Nicola Cordell
e-mail: nikkicordell@gmail.com