



Loss Control TIPS

Technical Information Paper Series

Innovative Safety and Health SolutionsSM

Is A “Comprehensive” Internet Search Possible?

As we become more familiar with the Internet and the search tools available to find information on the World Wide Web, it's tempting to think that because we can find so much information so quickly and easily, we are conducting comprehensive searches. Not only is this belief untrue, it could cause problems for you—and your customers—if you rely on the Internet for all your information needs.

To understand this, let's take a look at the overall content of the Internet and the various search tools we use to find our way around.

What's Actually In The Internet?

First, consider the nature of the Internet and how it has evolved. While the Internet was once the domain of the military and university establishments and a means for them to exchange technical information, it has since become a primary venue for the various entertainment, news, and sports industries, commercial enterprises, and individuals exchanging personal or recreational information (note the many sites devoted to genealogy, parenting, cooking, etc., as well as the numerous “personal home pages”). The proportion of educational and scientific information has shrunk. A recent study shows that 83% of Internet sites contain commercial content and only 6% contain scientific or educational content.

Of these categories, which are most useful to you in your work as a safety professional? Relatively few: educational and scientific institutions and government agencies, certainly; commercial sites, sometimes, to the extent that you can locate specific technical information and evaluate its objectivity; networks and newsgroups, yes, to the extent that you can evaluate the contributors and their credentials and assess the accuracy and relevance of the advice they offer.

(A note of caution: Keep in mind that while there may be some very good information in usegroups and online discussions, *anybody* can participate and offer opinions or “advice.” Identify the source of the information, and carefully evaluate its quality before you use it, and especially before you pass it on to colleagues or customers, or incorporate it in your reports or recommendations.)

But How Much Information Is Really There?

Everyone talks about how vast the Internet has become, and “everyone” is right—it is big. The study cited above also indicates that the “publicly indexable Web contains [as of February 1999] an estimated 800 million pages, encompassing about 15 terabytes of information after removing HTML tags, comments,



and extra white space, or about *six terabytes of text*.” That means that only about 40% of the Web’s content is actual text.

While six terabytes of text sounds enormous, compare it, for example, to the *amount of textual information* available from The Dialog Corporation, a major vendor of fee-based online research services. Dialog’s 900 databases includes over *9 terabytes of information*, contained in more than six billion pages of text and over three million image documents (patent and chemical structure images). There is no advertising.

Consider, too, the *indexing* available in these two resources (web vs. Dialog). Only 34% of web sites include keywords and descriptive tags which would facilitate precise indexing by search engines. By comparison, all of the records in all of Dialog’s databases are fully indexed by professional librarians and indexers, according to established standards and using standardized, controlled vocabularies. This means that in Dialog, it is much easier to know how and where to search for information, and the results are more predictable and more precise.

Finally, compare the *types of information* available in these two resources. Recall that 83% of Internet sites contain commercial content and only 6% contain scientific or educational content. (The remainder, presumably, comprises governmental and personal sites.) By contrast, Dialog’s databases provide three decades’ coverage of the medical, scientific, engineering, and technical literature; business and management journals; governmental and regulatory news, including recalls; comprehensive newswire services; full texts of hundreds of newspapers and magazines; patents; the arts and humanities, education, and more.

How Much Information Do The Internet Search Tools Actually Index?

None of the search tools indexes the entire Web, and the degree of coverage, both by individual search tools, and their combined coverage, is declining as the Web continues to increase in size and as more search tools are deployed. A 1998 study (published in the April 3, 1998 issue of *Science*) revealed that leading search engines cover surprisingly small portions of the Web. An updated study by the same research team (published in *Nature* in July 1999) demonstrated two important trends: first, an increase in the *number* of search tools; and second, a marked decrease in the *proportion* of the Web covered by each search tool.

Search Tool	Percentage (%) of the searchable Web covered by each search tool	
	1998	1999
HotBot	34	11.3
AltaVista	28	15.5
Excite	14	5.6
InfoSeek	10	8.0
Lycos	3	2.5
Northern Lights	*	16
Snap	*	15.5
Microsoft	*	8.5
Google	*	7.8
Yahoo	*	7.4
Euroseek	*	2.2

* not included in 1998 survey



Since no individual search tool indexes more than 16% of the Web, and since the search tools' *combined* coverage is only about 60%, there may be no efficient and effective way to gain access to the remaining 40% of the Web, which contains thousands or millions of potentially *useful* Web sites.

In addition, remember that most search engines exist as commercial enterprises. Although most are “free” to users, they are supported by advertising revenue. Therefore, it is in the best interest of the search engine companies—and that of their sponsors—to include web sites that will draw users who respond to advertising, make purchases online, or seek information about products and services. They are under no obligation to provide comprehensive coverage of the entire Web, or even make an attempt to do so. They are much more likely to spend their resources identifying and indexing popular sites (such as CNN or ESPN) rather than scientific or educational sites (such as a university department of occupational safety and health studies).

How Can You Search More Effectively?

First, *understand the tools* available to you:

- ❑ Get to know how the various search engines and directories work. Read the online help documents and the Frequently Asked Questions (FAQs). Find out the shortcuts and “best practices” for each one. However, keep in mind that search tools are updated and “improved” frequently, so check for new capabilities often.
- ❑ Practice searching; try different things; try the same searches using different search engines. Get to know what to expect—and what *not* to expect—from the various search tools.
- ❑ Understand how “metadata” (information about the web sites, such as subject headings) can affect your search strategy and results. First, recognize that some web creators pack their sites with “spam” (excess or misleading metadata) to draw your search. This can lead to unexpected results. Second, understand that only about one third of web sites include subject descriptions; this means that search engines will index only the text of the site, and will not have the added benefit of subject indexing.
- ❑ Allow yourself plenty of time to learn how to search effectively. The Web is a dynamic, complex entity. It is not easy.

Second, plan *and execute your searches carefully*:

- ❑ First, consider whether the Internet is the best resource to answer your question. This understanding will come with experience.
- ❑ Know what you are looking for. Check spellings, alternative terminology, and variant terms.
- ❑ Plan your search. Choose the best search tool(s) for your inquiry, understand how the tools work, and set up your strateg(ies) to suit the various tools.
- ❑ Start your search with one or two keywords; take care not to use words that are too broad, like “safety” or “health,” unless you combine it with another word or two to indicate a specific concept. When possible, use concise phrases like “safety incentives” or “seismic isolators.”
- ❑ Take advantage of features to help you define your search, such as the ability to specify phrase vs. word search, or to use truncation or word stemming (the ability to retrieve all variations of a word). Many search engines allow you to limit your search by language, geographic region, date, etc. Others will allow you to specify how “deep” into the Web sites you want to direct your search; for example, you can indicate “first pages only” or “entire sites.” A very useful feature is the ability to specify the domain(s) of the Web site(s) in which you wish to conduct your search (.com, .edu, .gov, .org, etc.)
- ❑ Know when to stop. If you don't find an answer in the first 30 or so hits of a search, try a different strategy or a different search tool. It's very unlikely that you will find good answers to your question past the first 30 to 50 hits.

- ❑ Understand that to conduct a thorough search, you will need to run your search in several different search tools, and perhaps use two or three different search strategies or approaches in each one. Remember, though, that you will never actually search the entire web, and that the various search engines only cover relatively small portions of the Web. You can never conduct a truly comprehensive search of the Web.
- ❑ Don't feel bad if you can't find the information you need on the Internet. There's a good chance that it hasn't been indexed in a search engine, or that it just isn't there.

Third, *don't rely only on the major, most popular search engines*. Seek out useful sites with real information. Here are some tips for finding good web sites:

- ❑ Follow links from credible sites. Good web sites, such as those of many universities or government agencies, are rich in well-organized links to authoritative sites.
- ❑ Seek out megasites, where subject specialists have organized links to web sites on particular subjects.
- ❑ Explore web sites noted in journal articles, specialty newsletters, etc.
- ❑ Sign up for e-mail lists that will keep you up to date on specific topics. A note of caution: Choose your lists carefully; sign up only with authoritative organizations (such as government agencies or universities). Be selective; you don't want to be overwhelmed with email.
- ❑ Seek out search tools that are created by scientific institutions, library groups, universities, and others who have a vested interest in providing access to the educational and scientific content of the Web (as opposed to search tools created by commercial entities and publishers).

When Is It Most Appropriate To Use The Internet?

When is it *appropriate* to use the Web to look for information?

- ❑ When you *know* that the information you need is on the Web.
- ❑ When you need government and regulatory information from federal agencies such as OSHA, BLS, FDA, CPSC, or DOT, or from state or local governments.
- ❑ When you are looking for an explanation or definition of a specific term (for example, a specific manufacturing process), especially if it is a relatively new technology.
- ❑ When you need information from professional and technical organizations such as NFPA or AIHA.
- ❑ When you need news and current events, information and news about the insurance industry, information and news about the Internet and personal computers.
- ❑ When you are looking for information about specific products or companies. Many firms maintain Web sites; however, remember that the information you find there will be self-promoting and may not give the whole picture (for example, if you are trying to find out about recalls).
- ❑ When you are looking for a map, product illustration, graphic, photograph, etc.
- ❑ When you need information on an unusual topic.
- ❑ When you have exhausted all other resources.

When is it *not appropriate* to use the Web to look for information?

- ❑ When you need a quick fact or other piece of information that is readily available, and easier to find, in a print source (library, dictionary, telephone book, etc.)
- ❑ When you need comprehensive, objective information from a variety of sources in order to get a complete picture of an issue or situation, or when you need concise, factual information from objective sources (e.g., research published in well-established safety magazines or in peer-reviewed scholarly journals), contact your corporate, university, or public library, or consult a professional librarian to do this research for you

Summary

So—is a comprehensive Internet search possible, No—at least not at present. Perhaps future enhancements to Internet structure and search tools will make this possible. However, you can maximize your chances of success—and minimize your frustration—by understanding these basic principles:

- Understand what information is—and is *not*—likely to be available
- Understand how that information is—and is *not*—likely to be organized
- Understand when—and when *not*—a Web search is likely to be productive in answering a particular question
- Understand what Internet search tools are available: the different types, what they can and cannot do, and how to use them effectively
- Know when to quit!

Experienced Internet users have found that:

- The Web lacks the “rich, refined traditions of the print world.” It is up to users to evaluate the information you find and to determine what is, and what is not, reliable.
- The Web is chaotic. There is no overall organization. There are no universal standards or guidelines for creation, presentation, or organization of information on the Web.
- The Web provides open access, both for creators and users. Anyone can put anything on the Web. While this guarantees a wide variety of information, it makes it difficult to find and evaluate *useful* information.
- The Web is a dynamic entity which changes constantly. A Web site or a portion of a site you find today may look different tomorrow, and may even disappear! Web creators are constantly “improving” the content, look, and functionality of their Web sites. Be prepared for surprises, both delightful and disappointing.
- Effective use of the Web requires individual discipline, both in learning to use it correctly and effectively, and in learning to use your search time efficiently.
- You will encounter steep learning and efficiency curves, but with regular use you can quickly develop familiarity and a degree of expertise.
- Many books, articles, newspaper features, and (especially) web sites discuss Internet use. Be advised, however, that information about the Internet becomes dated very quickly.

References

1. *Evaluating Information from the Internet*. The Hartford Loss Control Department, c2002. Technical Information Paper Series (S 100.100)
2. "How Search Engines Work." Search Engine Watch home page. searchenginewatch.com (August 5, 1999)
3. "Information Services Division Fact Sheet." Dialog Corporation home page. www.dialog.com (August 5, 1999)
4. *Introduction to the Internet and the World Wide Web: The Loss Control Perspective*. The Hartford Loss Control Department, April 1999 (L 100.102)
5. Lawrence, Steve and C. Lee Giles. "Accessibility of Information on the Web." *Nature*, 400: 107-109, 1999.
6. Lawrence, Steve, and Lee Giles. "Accessibility and Distribution of Information on the Web" www.wwwmetrics.com (August 5, 1999)
7. Lawrence, Steve and C. Lee Giles. "Searching the World Wide Web." *Science*, 280 (5360): 98, April 3, 1998
8. "Search Engine Coverage Study Published." Search Engine Watch home page. searchenginewatch.com (August 5, 1999)
9. "Search Engines Map Just A Piece Of Web" (Knight Ridder Newspapers Wire Service) *Hartford Courant*, Thursday, July 8, 1999, p. A15.
10. *Understanding Copyright and the Internet*. The Hartford Loss Control Department, c2002. Technical Information Paper Series (S 100.101)

For more information, contact your local Hartford agent or your Hartford Loss Control Consultant.
Visit The Hartford's Loss Control web site at <http://www.thehartford.com/corporate/losscontrol/>

This document is provided for information purposes only. It is not intended to be a substitute for individual legal counsel or advice on issues discussed within. Readers seeking resolution of specific legal issues or business concerns related to the captioned topic should consult their attorneys and/or insurance representatives.