Guaranteed hits

How to make your library's Web site stand out in Web search engines

by Jeffrey Beall

Creating and publishing a site on the World Wide Web is only the first step in making your library information available to the world at large: if you want to make sure it gets used, you have to publicize it. *Publicizing* in the context of the Web means ensuring that your site gets adequately indexed in the appropriate search engines. (*Search engines* are Web sites with search platforms that allow users to retrieve lists of links to Web sites matching their search requests.) Getting the search engines to index your site is a process that requires patience, but the indexing can pay high dividends if it results in having your site become a highly relevant "hit" for information-seekers looking for materials on the subject of your Web site.

If you create a new Web site and make no attempts to publicize it on the Web, there is, admittedly, still a fair chance that it will become indexed eventually by at least a few of the search engines that index the Web. This happens because, one way or another, people are going to find out about your site and make links to it. For instance, search engines may have the ability to find and index every site on a given server. Some search engines *crawl* the Web. This means that they start on a single Web site, index it, and follow all the links on the page and index them as well, and so on. This process is also referred to as *spidering*. When a search engine indexes a site, it downloads all the textual material from it and excludes all the HTML coding. So what the search engine sees at each site is a body of language material, or simply text.

However, relying on Web crawling to get your site indexed in the search engines is not an efficient means of getting your site included. It is possible that few other sites will link to yours, or the spidering process can be slow or inefficient. The best way to get your site indexed is to be proactive about it: in short, to do it yourself.

Include me in

The most successful way to get your Web site included in the search engines is to visit their home pages, look for the "Add URL" button (or equivalent), click on it, and follow the instructions. Usually the process is very simple; in fact, for some engines all you have to do is enter the URL or URLs you wish to be indexed. A few sites may ask for more information, such as your e-mail address and some keywords to describe the site. Some may even offer you a set of classification terms, which you can select to describe your site.

Smaller search engines can come and go, but there are seven major engines that are permanent and comprehensive, and when you begin to publicize your site you should begin with these engines. These major en-

About the author

Jeffrey Beall is senior cataloger at Widener Library at Harvard University; e-mail: jbeall@fas.harvard.edu

gines include AltaVista, Excite, HotBot, Infoseek, Lycos, Open Text, and WebCrawler. For a complete list of search engines on the Web, look at one of these two sites:

http://www.isleuth.com/webs.html

or

http://www.yahoo.com/Computers_ a n d _ I n t e r n e t / I n t e r n e t / World_Wide_Web/ Searching_the_Web/ Search_Engines/.

For a price . . .

A number of Web-based companies have sprung up that, for a fee, will take the information you provide, visit the "Add URL" pages of each search engine, and get your site indexed for you. Some even purport to do it, on a limited basis, for free. One major disadvantage of using such services is that they are mainly set up to serve Web-based businesses, the primary source of their revenue, and therefore a library might spend money for something that it can do more efficiently and more accurately itself. It is probably much better for Web site creators to publicize themselves—what looks like a legitimate Web-promotion company (based on its Web appearance) can turn out literally to be two guys with a Mac and good graphic skills in a garage somewhere.

Indexing by design

There are a number of things you can do in terms of the design of your Web site that can increase the likelihood of its being indexed accurately. First, don't include vital information exclusively in graphics, because they may not get "read" by the search engines. Put the most vital information—the name of your library or department, for example—in a larger-than-average font at the top of the Web site. Some search engines determine relevance by comparing the users' search queries with the size and position of the information on a Web page.

Use the Dublin Core

Second, incorporate the elements of the Dublin Core metadata set into your site. This is a block of data in a standard form that provides information about your site in a way that is recognized by the search engines. Several sites exist on the Web that

will create this vital HTML information for you—all you have to do is fill out a form on a Web page, submit it, and then copy/paste the HTML data the site returns to you directly into your Web site.

By having a conventional set of metadata describing your Web site, you will increase the likelihood of its being accurately indexed and retrieved by the search engines. Moreover, the Dublin Core allows you to control how the search engines describe your site in their search results pages.

To create an abstract of Web sites, many search engines merely copy the first 25 words or so of the text of your Web document. By filling out a brief description and including it within your Dublin Core metadata, you can control what the abstract of your site says. Because the Dublin Core standard has been promoted by the library community, using it means you are setting a good example and helping promote a standard for metadata. For Dublin Core metadata generators (or templates) see the following sites:

http://www.ukoln.ac.uk/metadata/dcdot/

and

http://www.ub.lu.se/metadata/ DC_creator.html.

Catalog it!

Another way for a library to promote its Web site is to catalog it. This can be done exclusively in the library's online catalog, but better yet, each library should catalog its homepage in at least one of the bibliographic utilities. A catalog record may increase awareness of the site, and other libraries may be more willing to include the bibliographic record in their local OPACs if someone has already cataloged it.

Some libraries may wish that certain Web documents not be included in Web search engines. There is a way to prevent your site from being included, and it is done by including some data in the meta tag of a HTML document. Simply include the following data, called the robots exclusion tag, in the head portion of your HTML document:

<META NAME="ROBOTS" CONTENT ="NOINDEX, NOFOLLOW">

For more information on the robot exclusion tag, see http://info.webcrawler.com/ mak/projects/robots/exclusion.html#meta.

The presence of the robots exclusion tag means the search engines will not include the document that contains the tag among items returned by searches done on the search engines. Note that some search engines may be less sophisticated and may not be programmed to allow for this feature.

Accentuate what's special

Is there some special topical aspect of your site that may allow for its inclusion in a specialized search engine? Because the general search engines have gotten so big that they are often unreliable, many smaller, specialized search engines have sprung up. So, if your library deals with a special topic, say architecture, you may want to look for specialized search engines on that topic and get your site indexed in them.

Perhaps the largest collection of search engines—specialized and general—is located on the Web page called the "Internet Sleuth." This site has a neatly classed collection of hundreds of Web search engines. If the site you wish to promote has some special subject emphasis, you can go to the Internet Sleuth and find all the engines that correspond to your site's topic. Then you can visit each site and look for the "Add URL" button, and follow each site's instructions for getting indexed. The Internet Sleuth is located at: http://www.isleuth.com/.

Be patient and time it right

It is important to note that not all Web search engines index instantly. Some may take one or two weeks or more to include a site. Often this time lag information is stated when you get the confirmation screen after submitting your site's URL. Also, it is better to submit your sites to the search engines in the morning, when the Web is less crowded. URL submission sites have a tendency to be crowded, so accessing them in the morning will be more time saving.

Generally it is not possible to do multiple URL submissions via mega-search engines. *Mega-search engines* (also known as *meta-search engines* or *parallel search engines*) are Web sites that search two or more Web search engines simultaneously and then in one way or another collate and return the search results to the user.

Although one can search several search engines at a time, one cannot use this method to submit URLs for multiple indexing. Moreover, do not be confused by the expression "Link to us," a statement often seen on various Web sites. This means that the owners of the site want you to include a link to their site from within your site. This is not a means of publicizing your site.

Take shortcuts when possible

Finally, there are several Web sites that bring together all of the "Add URL" pages of the major search engines. A good one is located at:

> http://www.tiac.net/users/seeker/
> searchenginesub.html

and the corresponding Yahoo! page is located at:

> http://www.yahoo.com/Computers_
> and_Internet/Internet/World_Wide_
> Web/Announcement_Services/Indices/.

By using these sites, you can gain easy and generally quick and direct access to precisely the pages you need to submit your URL to the search engines.

Try it yourself

After you have submitted the URL of your Web page to the various search engines and after you wait some time for them to get included in the engines' indexes, you should search them to make sure your site is indeed included. A good way to do this is to search some of the less common words from your site in the search engines' main search area. Using less common words will create a smaller retrieval set and make the retrievals easier to analyze.

Alternatively, enter the title of your Web site in the search box, and see how highly ranked it is among the retrieval set. If it is not at the top, you may want to redesign your site to ensure that Web surfers, including your local library's users, have access to it.

The whole point of putting together a Web page is to provide access to information: you "build it so they will come." This effort can be lost if no one knows that the information is there for the taking. So take the next logical step and use the aforementioned techniques to ensure others know it *is* built. ■