

Working With Search Engines

NOTE: The following section is excerpted from the *ServInt Webhosting User Manual*. Please see the manual for the full version of this and many other articles covering essential topics for webmasters.

Many of ServInt's clients are interested in the Web's major search engines, sites that provide a de facto index of what's out there on the web. This section is designed to review some basic information on the search engines, and how you can optimize your website to make use of them, as well as to state ServInt's policy on search engine use and abuse for our clients.

So, we're presenting the following as a quick overview on search engines - how they work, and how to use them - and how not to abuse them. There are many helpful Internet sites and resources out there (from which much of the information for this article was drawn) with excellent information on improving your search engine ranking. A list of these sites and services is given at the end of this article.

How Search Engines Work

Many people use the term "search engine" to describe a variety of directory sites that have different purposes and inner workings. Let's review those briefly, to cover the distinctions between them. There are three major types of sites which are generically called "search engines:" search engines, directories, and hybrid search engines.

Search Engines: True search engines crawl the web, then automatically generate their listings. If you change your web pages, search engine crawlers will eventually find these changes, and that can affect your listing. Page titles, body copy, meta tags (sometimes) and other elements all play a role in how each search engine evaluates the relevancy of your page (and hence its ranking). There are plenty of ways that you can cater to a search engine's crawlers and change your site to help improve its rankings.

A search engine has three "parts:" first is the crawler, or spider. The spider visits web pages, reads them, follows links into the site and transmits what it finds back to the search engine's master database, or index.

The index (or catalog) contains a copy of every web page found by its spiders/crawlers. If a crawler sends back new/updated information on a page, it updates the page's entry in the index. The third part is the search engine software, which is the engine's user interface. The software sifts through the index in response to user searches and returns a listing of sites, based on what it finds most relevant.

Directories: A directory (such as Yahoo!) works only by having sites submitted to it by humans. You enter a short description of your site, or the directory's editors write reviews of sites that attract attention. That description then stays there until you resubmit your site; just changing your site will have no effect on how the site is listed.

Furthermore, any search on your site in a directory will bring up only those keywords you chose to identify it with when you submitted it (or the editors did when they reviewed your site). Normal "search engine" tips for improving your site's ranking will have no effect on a directory's listing (although a well-designed site might make it a better candidate for a review).

Note, however, that Yahoo! is unique from all other directories in that you could be listed on Yahoo! without having submitted your URL. They make use of a robot that searches for new sites at certain Internet locations (i.e. announcement sites, etc.). Despite Yahoo!'s implementation of this search robot, mostly all of their new additions still come via their "Add URL" form.

Hybrid Search Engines: Some search engines maintain a directory that "works with" the information its crawlers retrieve. Being included in one of these directories is as much luck as it is skill, since you can submit your site, but there's no guarantee of it being included.

How Searches Work

Each search engine works differently. There is no standard way to determine search string relevance in a search. Some engines determine relevance only by the user-visible text on the page (**Lycos**). Some only use keywords (**Excite**). Some use a combination (**AltaVista/InfoSeek**). There are also intelligent agents out there that use who knows what to index on.

Several different factors influence how a search results page is built. These factors include:

- what word is being searched for
- whether Web pages containing the word were ever registered with the engine or spidered by the engine.
- whether meta keywords are used or ignored
- whether the contents of the Web page are processed or ignored
- how the particular search algorithm weighs the different aspects of a Web page in determining the relevancy of any word being search on.

Making Your Site Search Engine-Friendly

Each search engine's crawler looks at different elements of your page, so there's no "silver bullet" that will improve your listing on all search engines. However, following as many of the following tips as you can should help across the board.

<TITLE> Tags

Use keywords in the `<TITLE>` of your document, making it as descriptive as possible. When visiting your site, an agent will go first to the `<TITLE>` tag. For clarification purposes, the `<TITLE>` tag is what a browser will display in its title bar and is not simply the first line of HTML that shows up on your page. (Although your first words of introductory text should be descriptive as well). Search engines will display the text located between the `<TITLE>` tags when your web page is listed in a search.

Probably every search engine or directory uses the contents of the title tags – some use them a little, but some use them a lot. And this is your opening shot to add relevant content to your page. So write a brief, descriptive title. Keep the title reasonably short, but be sure it has something to do with the Web page itself. A title like

`Ed's Bait, Tackle and Dial-Up`

might be fine for a home page title, but if it's a business, add a more information. Consider:

`Ed's Bait, Tackle and Dial-Up:#1 Provider of Nightcrawlers and 56k Modem Access`

By adding a more information, you've given the search engine spiders a little more content to include in the catalog. By making your `<TITLE>` descriptive, you'll be better off than those who only have keywords within the text of their page. It will also be helpful when people bookmark your web site. If a more descriptive name appears in a person's hotlist, it will be easier to find your site at a later date.

Text at the Top

Search engines assign greater relevancy to text located at the top of a page than to text located in the middle or at the bottom of the page. The search engines assume that web page authors will present their most important information first. If your page has a main graphic at the top, you should place some descriptive text either underneath or beside the image (see "Header Tags"). The search engines will index this text and assign it a high level of relevancy.

<H1> and Other "Header" Tags

Be sure to add headings to the file, using the HTML "header tags" (`<H1> ... </H1>` and so on) rather than `` tags. Most search engines are looking for `<H1>` and `<H2>` as a clue to what concepts are "most important" on the page. Use them appropriately in your file.

Use `<H1> ... </H1>` near the top of your home page to restate the the page's title. It's fine to put an eye-catching logo at

the top of the page, but a logo means nothing to a search engine. Consider using both a logo and the `<H1>` heading.

<META> Tags

Use `<META>` tags that allow you to provide even more detail about your Web pages and thereby gain greater control over how your pages are indexed. Not all search engines make use of `<META>` tags, but adding these tags to your pages will make them more accessible to the search engines that do. `<META>` tag codes are inserted within the `<HEAD> ... </HEAD>` tags. The basic syntax is:

```
<META name="description" content="a nightcrawlers and Internet access provider in Virginia">
```

This will control what appears as the summary of your Web page and will be displayed after the title of your document in the index listing. The content of the description should clearly convey what one can expect to find when linking to your site.

```
<META name="keywords" content="fishing worms, dry flies, bass lures, unlimited Internet">
```

This will allow you to provide extra information about your page to the search engines without it being visible to the reader. While search engines do take these keywords into account when indexing your page, they are still going to index the entire contents of your page as many sites do not include `<META>` tags. Since this is the case, there is no need to be redundant. Include keywords that will not necessarily be derived when a robot visits your site. In other words, "night-crawlers" and "56k modem access" need not be included in your list of keywords as it is part of your `<TITLE>` .

Remember, HTML limits you to 1024 characters in your `<META>` tag, so don't waste any by using words that are Internet-related. The reason? Most of the engines disregard these because there are just too many, resulting in millions of matches on those words. Besides – it's a web site. Why bother using a keyword of "World Wide Web," anyway?

If you use a page-layout tool (like Adobe PageMill or Microsoft FrontPage) to create your web pages, be on the lookout for META tags inserted by your page generator. For example, a page created with PageMill will automatically include a tag like this:

```
<meta name= "generator" content="Adobe PageMill 2.0 Mac">
```

This META tag may be picked up by search engines, and be distracting. You can safely remove this tag completely; but if you try to leave it there but only alter what's in it, your page-generator program will either erase your changes, or may even crash. Check the documentation for your page-layout program for more information.

Using Text

If your home page has lots of graphics or Java scripts, if you use page redirection, or if most of your site is contained in databases, Perl and CGI, consider creating a text-only Web page that describes your company/organization/Web site. A text-only page is a huge help to people searching for specific information, for the visually impaired who use ASCII-to-voice systems and for people who use text-based browsers like lynx. Put this text-only page in your top-level directory, link it to the home page and submit an ADD SITE request for the text-based page.

General Tips

Organized Text Helps

Document structure does seem to have some merit with the search engines and thinking logically from their point of view the reasons are probably twofold. Good structure results in documents that are easy to spot the meat in and secondly only good authors or those that really know what they're doing can produce them. This results in higher ratings for good documents.

Pick Strategic Keywords

The best way to identify good keywords for use in `<META>` tags and page descriptions is to try to visualize what words your site's visitors will be searching for. With these in mind, you can try to incorporate the most "strategic" keywords to prominently display on your pages, or in your site's directory listing.

To help with search engine crawlers, each page in your website should have different strategic keywords in its `<TITLE>` and `<META>` tags, to ensure that the broadest range of possible keywords are covered somewhere in your site.

Your strategic keywords should always be at least two or more words long, since any one-word keyword is probably too broad and will return literally thousands or millions of matches. Reduce your level of competition by being as specific as possible with your keywords.

Include Relevant Content

Changing page titles or adding `<META>` tags won't do you a lick of good if your site doesn't have anything to do with the keywords.

Particularly, you should have HTML text on your page which matches the keywords. As previously mentioned, if much of your site is in graphics, be sure to include `<ALT>` tags which provide a suitable description for the crawlers. It can work to your advantage – some search engines will index `<ALT>` tags and comment information along with the `<META>` tags. Even if they don't, there's no way they can index pure graphics, so you really have nothing to lose.

Submit Your Key Pages

Generally, search engine crawlers will follow all of the available links through your site and index all of your pages. However, the process isn't perfect, and they sometimes "miss." With this in mind, it's a good idea to separately submit the top two or three (but not much more!) pages that best describe your site.

Avoid Common Mistakes

Bad Keywords

When selecting keywords, try to use a unique combination of words, or combine a word and a number. Avoid using words with punctuation in them, especially hyphens (-) and question marks (?). Don't start a word with a numeral. Some search engines have a real problem with terms such as 3M, C++, or AT&T.

If you want your Web page to be found using certain search words, be sure to include those words near the beginning of the Web page. If you don't want your site to be found using certain search words, make sure they are not included on the Web page. The spider makes determinations about relevant words based on how the words are used on the page.

It's very important to understand that the more frequently words are found in Web pages, the more difficult it is to find any particular page containing those frequently-used words. If your company's name is "Computer Software Products", it's tough for any search engine to return your page in response to those words being entered because those words appear in tens of thousands of other Web pages. Your page might appear as number 1,939 on the search results list of 3,000 Web pages that all three of those words appear in.

Frames and Tables

The use of tables to layout a document can and will in some cases change your placement because when you have a table for say a left border margin you force the meat of the document where most of your keywords will be found further down in the document lowering your ranking. Keep in mind that some engines' crawlers only index *part* of your site, ranging from 10-50%. A rule of thumb could be to put keywords in the left margin or in the top table of your document.

In some cases, the use of frames will limit the robots' ability to even index your site. If you're using frames, you should have the robot index the `<NOFRAMES>` version of your site, as some can't index frames or will just index the frameset which isn't much to index and will result in a low ranking.

Design Problems

Often, designers create only image map links from the home page to inside pages. A search engine that can't follow these links won't be able to get "inside" the site. Unfortunately, the most descriptive, relevant pages are often inside pages rather than the home page. Solve this problem by adding some HTML hyperlinks to the home page, something that will help some of your human visitors, also. Put them down at the bottom of the page. The search engine will find them and

follow them. Also consider making a site map page with text links to everything in your web site. You can submit this page, which will help the search engines locate pages within your web site.

Generating pages via CGI or database-delivery? Expect that some of the search engines won't be able to index them. Consider creating static pages whenever possible, perhaps using the database to update the pages, not to generate them on the fly. Also, avoid symbols in your URLs, especially the ? symbol. Search engines tend to choke on it.

Password Protected-Sites

Be forewarned that many search engines can't crawl pages which are in password-protected directories (i.e., with .htaccess). Make sure that you aren't relying on the contents of your password-protected pages to get you into search directories. Place your important keywords and content on pages which are *not* password-protected. Currently, the engines Excite, InfoSeek and Lycos *will* crawl through password-protected areas. However, AltaVista, HotBot and WebCrawler will *not*.

Engine-Specific Indexing Traits

HotBot / Inktomi

1. HotBot supports both the keywords and description `<META>` tags.
2. If you strongly believe that your site was not ranked as high as you thought it deserved in a search query, HotBot allows you to send them an E-mail to bugs@hotbot.com. Be sure to include the URL of the search page.

InfoSeek

1. InfoSeek supports both the keywords and description `<META>` tags. Your description can include up to 200 characters of text and the keywords can include up to 1000 characters of text. Do not repeat versions of a keyword more than seven times. If you do, InfoSeek will disregard the entire keyword list.
2. If you do not make use of the description `<META>` tag, InfoSeek's agent will simply insert the first 200 characters after the `<BODY>` tag as the web page description. Hence, if your Web document does not contain `<META>` tags, at least try to make your first 200 words accurately describe your Web site.
3. InfoSeek also indexes the ALT attribute in the `` tag. If the majority of your home page consists of graphics, you can describe your page with the ALT attribute. You can include this attribute in your image tags with the following syntax: ``

Excite

At the present time, Excite does not make use of tags. Since keywords and summaries are automatically generated by Excite, you have less control over their creation. However, there are still a few things you can do. Excite's software looks for common words or themes within a page. It then selects sentences for the summary that either contain these words or convey the overall theme. The words within these sentences are also used as keywords for which the site can be searched.

1. Especially at the beginning of your page, be as concise as possible and limit non-descriptive sentences. If the Excite robot comes across a number of ambiguous phrases, it will have to look deeper and deeper into your site to determine its theme and site summary. Along the same line, too little text will also force the robot to travel further into the site for more information in order to establish a theme.
2. Excite's indexing software places preference on complete, punctuated sentences. If you have content, such as a quote, at the top of your page that you do not want Excite to include in your site summary, do not display it as a complete sentence. This will lessen the chance that the quote will be included, but will not guarantee its exclusion.

Lycos

Lycos creates Web site titles and descriptions from the text of your Web page. Lycos' search agent selects a portion of the site that well represents its theme. It then displays this section as the site's description.

The keywords are also chosen via artificial intelligence by Lycos' spider. With this in mind, do not open your page with an image map. If you do, Lycos will not be able to take an abstract for your document.

WebCrawler

WebCrawler relies on the statement within the <TITLE> tag to use for the name of your page. While other search engines will derive a summary from the <BODY> text of the document, Web Crawler will default to the URL if you fail to include a title.

General Advice for Directories

Since directories are hierarchical databases organized by subject, subject category selection would logically be the most critical aspect of linking your site to a directory. Give some thought to which category or categories your URL should be placed as this location will ensure the right people (those interested in your site) will be visiting your link. URL registrations can quickly become complex process when every directory contains a different sub-categorization scheme.

There unfortunately is no formula or exact method for correctly placing one's site into a subject-oriented guide. However, here are a few tips:

1. Do not assume a category (i.e. "Entertainment") means the same thing on every directory.
2. Take the time to drill down within the directory until you come to the subcategory most appropriate for your site. If you find many organizations similar to your own listed, you're probably in the right place.
3. Do not choose a less fitting category because it's either alphabetically desirable or doesn't contain many other links. If it is not appropriate, there is no reason to believe anyone clicking that category would be interested in visiting your site.
4. If you cannot find an appropriate category, do not hesitate to suggest a new one. This advice even applies to Yahoo!! Do you think the Yahoo! staff thought of all 22,000 headings all by themselves?

Yahoo Made Easy

For Internet users, Yahoo! is one of the more notoriously difficult submission forms to successfully complete. In actuality, the Yahoo! submission form is far from a mind-numbing experience. We have outlined the necessary steps below. If you follow them, it will not only simplify the process for you, but also enable you to gain the greatest advantage possible through your submission.

1. Find your Yahoo category.

From Yahoo's home page, drill down and locate up to 2 Yahoo categories that best describe your Web site. A good indicator that you have found an appropriate sub-category is when you scroll down and see sites already listed that are similar to your own, or, even more so, find a direct competitor! If after searching through a number of Yahoo categories, you still are not sure where to place your site, try searching by a keyword you would want your site to be found under. Existing Yahoo categories will be displayed with the search results.

TIP: Yahoo requires commercial sites to be placed within a **Business and Economy** subcategory - either under **Companies** or **Products and Services**. Personal Home pages must go in the **Entertainment/People** category. If your site is regionally specific, add it to the appropriate subcategory under the Regional heading category.

TIP: If suggesting more than one category, press the ADD URL icon when on the category of your first choice. Then enter the string of the second category into the **Additional Categories** field. To avoid making a typo to a long string, copy the URL located within the Address or Location field of your browser. To do this, highlight the text after <http://www.yahoo.com/> with your mouse. Right mouse click and release the mouse over "copy." For example:

```
Business_and_Economy/Companies/Music/  
CDs_Records_and_Tapes/Online_Shopping/
```

2. Click the Add URL icon at the top of the screen

After finding the category you wish your site to be placed within, click the Add URL icon at the top of the screen. This will take you to the Yahoo submission form with the Category fields already pre-filled.

3. Fill out the ADD URL Form

Two particularly important fields are the Title and Comments fields. Yahoo does not allow you to suggest keywords in their submission form. Instead, it will use both the Title and Comments fields to derive search words for your site. Therefore, think carefully about what keywords you want Yahoo to list you under and construct your Title and Comments fields around them. Do not, however, just submit keywords separated by commas. Enter complete, coherent phrases/sentences.

Example of a poorly constructed title: Brian's Music Store

Example of an optimized title: Brian's Music Store, CDs for sale

TIP: Your title should not exceed 50 characters (including spacing).

Example of a poorly constructed Comment:

Click to this page to listen to a wide variety of music sure to fit your interest.

Example of an optimized Comment:

Pop, Rock, New Age and Alternative CDs. Many artists, your favorite music, CDs on sale.

TIP: Your comment should not exceed 20 words or 200 characters (including spacing).

Yahoo has employees visit the URL of submissions to verify that the category you selected is appropriate and that the title and comments are accurate. As a result, do not exaggerate what is located on your site. In the end, your site will simply not get listed.

4. Click the Submit button

Once you have completed the form, click the Submit button to send your Web site information to Yahoo. If the form has been filled out properly, Yahoo will display a screen that thanks you for the submission. If there is an error in the submission form, Yahoo will display a screen informing you that the submission has failed and, depending on the error, tells you why the submission failed. In the case of a failure, press the Back key on your browser, make the necessary changes and resubmit.

5. Check your listing

Yahoo states that it should take two weeks for your site to be listed. They recommend you do a search for your URL. If your site is not listed, then resubmit your entry using the Add URL function in Yahoo. For more information, check out Where's My Site on Yahoo.

Updating Your Listing

Here are some general rules and indicators for knowing when you need to make updates to your listings. For more detailed information on the process involved in making updates to each directory, consult the Submission Hints and Tips accessible within the Submit It! Gold and Pro versions. You should make an update if:

For search engines:

- If your URL changes, you will need to resubmit. Some search engines have Dead Link forms for you to fill out. Those that do not will drop the old URL from its records the next time it tries to visit your site at the old address and is unable to find it.
- If the content of your site changes so extensively that the abstract no longer adequately describes your site. These changes, however, will be taken into account when the indexing agent next visits your site to refresh its information. Resubmitting will hasten this process.

For directories:

- If your Web site address (URL), description, or category classification changes, you will need to inform the webmaster of that directory.
- If your category classification has changed and you are now inaccurately located within the directory's index.

Do *not*, however, make updates to your listing:

- If there have been minor changes made within the site which do not *alter the site's description*. If the content or purpose of the site remains the same, you are not misleading anyone who chooses to visit your site based on the original description.

Conclusions

Do not make it your goal to appear in the top ten list of every search engine. This would not only be a very difficult task, but would most likely end in disappointment. You are virtually guaranteed to have varying success rates with different search engines due to the collection of variables that play a role in every search result. These variables include size of database, method used for determining relevancy, policy on spamming, use of <META> tags and more.

Optimize your listings by capitalizing on the indexing criteria *shared* by search engines. If you have a favorite or being yielded as high as possible in a search, customize your Web pages accordingly. Just keep in mind that this might lessen your perceived relevancy on another.

Also, remember that search engine spamming may not just get you bounced from the search engine – severe cases may result in your ServInt account being suspended completely. So don't take chances with something that offers little to gain, and plenty to lose.

ServInt Policies

Just as with Unsolicited Commercial E-mail and Usenet abuse, ServInt also has policies on the abuse of search engines. If ServInt receives complaints about a site which is abusing a search engine, we will make our best efforts to get in contact with the client via phone and E-mail. If no response is received, or if the abusive practices are not stopped, ServInt reserves the right to take any necessary action to protect its other network users, up to or including suspension of account. If you have any questions about ServInt's policies on search engine use and/or abuse, send E-mail to admin@servint.com.

Additional Information

If you're really interested in search engines, an excellent first step to learning more is by getting a membership with a search engine review site – we highly recommend Mecklermedia's Search Engine Watch, at <http://www.searchenginewatch.com>.

A quick and easy way to have your site submitted to most search engines is through the Submit-It! service (<http://www.submit-it.com>). This site also contains a great deal of valuable information about search engines.

Two sites with excellent beginner-to-medium tutorials on search engines and making the most of them are Webmaster T's Search Engine Tips (<http://www.globalserve.net/~iwb/archives/fa2.htm>) and the Search Engine Tutorial (<http://www.webreference.com/content/search/>).

A site with a general review of search engine technology, featuring some charts and a set of tools for webmasters is at Northern Webs Services (<http://northernwebs.com/set>).

For webmasters interested in tracking their search engine performance, the Web Position Analyzer software (<http://www.webposition.com>) is a probable best choice. You can read a review of the product (from a very happy user) at <http://www.globalserve.net/~iwb/WebPosition/review.htm>.