

Table of Contents

How Search Engines Work	1
How Search Engines Rank Web Pages.....	3
Search Engine Placement Tips	5
How To Use HTML Meta Tags	11
Checking Your URL.....	19
Measuring Link Popularity	24
Search Engines And Frames	26
What Is A Bridge or Doorway Page?	30
How Search Engines Regionalize.....	32

How Search Engines Work

By [Danny Sullivan](#)

Editor, SearchEngineWatch.com

Updated: October 14, 2002

The term "search engine" is often used generically to describe both crawler-based search engines and human-powered directories. These two types of search engines gather their listings in radically different ways.

Crawler-Based Search Engines

Crawler-based search engines, such as Google, create their listings automatically. They "crawl" or "spider" the web, then people search through what they have found.

If you change your web pages, crawler-based search engines eventually find these changes, and that can affect how you are listed. Page titles, body copy and other elements all play a role.

Human-Powered Directories

A human-powered directory, such as the Open Directory, depends on humans for its listings. You submit a short description to the directory for your entire site, or editors write one for sites they review. A search looks for matches only in the descriptions submitted.

Changing your web pages has no effect on your listing. Things that are useful for improving a listing with a search engine have nothing to do with improving a listing in a directory. The only exception is that a good site, with good content, might be more likely to get reviewed for free than a poor site.

"Hybrid Search Engines" Or Mixed Results

In the web's early days, it used to be that a search engine either presented crawler-based results or human-powered listings. Today, it is extremely common for both types of results to be presented. Usually, a hybrid search engine will favor one type of listings over another. For example, MSN Search is more likely to present human-powered listings from LookSmart. However, it does also present crawler-based results (as provided by Inktomi), especially for more obscure queries.

The Parts Of A Crawler-Based Search Engine

Crawler-based search engines have three major elements. First is the spider, also called the crawler. The spider visits a web page, reads it, and then follows links to other pages within the site. This is what it means when someone refers to a site being "spidered" or "crawled." The spider returns to the site on a regular basis, such as every month or two, to look for changes.

Everything the spider finds goes into the second part of the search engine, the index. The index, sometimes called the catalog, is like a giant book containing a copy of every web page that the spider finds. If a web page changes, then this book is updated with new information.

Sometimes it can take a while for new pages or changes that the spider finds to be added to the index. Thus, a web page may have been "spidered" but not yet "indexed." Until it is indexed -- added to the index -- it is not available to those searching with the search engine.

Search engine software is the third part of a search engine. This is the program that sifts through the millions of pages recorded in the index to find matches to a search and rank them in order of what it believes is most relevant. You can learn more about how search engine software ranks web pages on the aptly-named [How Search Engines Rank Web Pages](#) page.

Major Search Engines: The Same, But Different

All crawler-based search engines have the basic parts described above, but there are differences in how these parts are tuned. That is why the same search on different search engines often produces different results. Some of the significant differences between the major crawler-based search engines are summarized on the [Search Engine Features Page](#). Information on this page has been drawn from the help pages of each search engine, along

with knowledge gained from articles, reviews, books, independent research, tips from others and additional information received directly from the various search engines.

Now let's look more about how crawler-based search engine rank the listings that they gather.

How Search Engines Rank Web Pages

By [Danny Sullivan](#)

Editor, [SearchEngineWatch.com](#)

Updated: October 14, 2002

Search for anything using your favorite crawler-based search engine. Nearly instantly, the search engine will sort through the millions of pages it knows about and present you with ones that match your topic. The matches will even be ranked, so that the most relevant ones come first.

Of course, the search engines don't always get it right. Non-relevant pages make it through, and sometimes it may take a little more digging to find what you are looking for. But, by and large, search engines do an amazing job.

As WebCrawler founder Brian Pinkerton puts it, "Imagine walking up to a librarian and saying, ' travel.' They' re going to look at you with a blank face."

OK -- a librarian's not really going to stare at you with a vacant expression. Instead, they're going to ask you questions to better understand what you are looking for.

Unfortunately, search engines don't have the ability to ask a few questions to focus your search, as a librarian can. They also can't rely on judgment and past experience to rank web pages, in the way humans can.

So, how do crawler-based search engines go about determining relevancy, when confronted with hundreds of millions of web pages to sort through? They follow a set of rules, known as an algorithm. Exactly how a particular search engine's algorithm works is a closely-kept trade secret. However, all major search engines follow the general rules below.

Location, Location, Location...and Frequency

One of the the main rules in a ranking algorithm involves the location and frequency of keywords on a web page. Call it the location/frequency method, for short.

Remember the librarian mentioned above? They need to find books to match your request of "travel," so it makes sense that they first look at books with travel in the title.

Search engines operate the same way. Pages with the search terms appearing in the HTML title tag are often assumed to be more relevant than others to the topic.

Search engines will also check to see if the search keywords appear near the top of a web page, such as in the headline or in the first few paragraphs of text. They assume that any page relevant to the topic will mention those words right from the beginning.

Frequency is the other major factor in how search engines determine relevancy. A search engine will analyze how often keywords appear in relation to other words in a web page. Those with a higher frequency are often deemed more relevant than other web pages.

Spice In The Recipe

Now it's time to qualify the location/frequency method described above. All the major search engines follow it to some degree, in the same way cooks may follow a standard chili recipe. But cooks like to add their own secret ingredients. In the same way, search engines add spice to the location/frequency method. Nobody does it exactly the same, which is one reason why the same search on different search engines produces different results.

To begin with, some search engines index more web pages than others. Some search engines also index web pages more often than others. The result is that no search engine has the exact same collection of web pages to search through. That naturally produces differences, when comparing their results.

Search engines may also penalize pages or exclude them from the index, if they detect search engine "spamming." An example is when a word is repeated hundreds of times on a page, to increase the frequency and propel the page higher in the listings. Search engines watch for common spamming methods in a variety of ways, including following up on complaints from their users.

Off The Page Factors

Crawler-based search engines have plenty of experience now with webmasters who constantly rewrite their web pages in an attempt to gain better rankings. Some sophisticated webmasters may even go to great lengths to "reverse engineer" the location/frequency systems used by a particular search engine. Because of this, all major search engines now also make use of "off the page" ranking criteria.

Off the page factors are those that a webmasters cannot easily influence. Chief among these is link analysis. By analyzing how pages link to each other, a search engine can both determine what a page is about and whether that page is deemed to be "important" and thus

deserving of a ranking boost. In addition, sophisticated techniques are used to screen out attempts by webmasters to build "artificial" links designed to boost their rankings.

Another off the page factor is clickthrough measurement. In short, this means that a search engine may watch what results someone selects for a particular search, then eventually drop high-ranking pages that aren't attracting clicks, while promoting lower-ranking pages that do pull in visitors. As with link analysis, systems are used to compensate for artificial links generated by eager webmasters.

Search Engine Placement Tips

By [Danny Sullivan](#)

Editor, [SearchEngineWatch.com](#)

Updated: October 14, 2002

A query on a crawler-based search engine often turns up thousands or even millions of matching web pages. In many cases, only the 10 most "relevant" matches are displayed on the first page.

Naturally, anyone who runs a web site wants to be in the "top ten" results. This is because most users will find a result they like in the top ten. Being listed 11 or beyond means that many people may miss your web site.

The tips below will help you come closer to this goal, both for the keywords you think are important and for phrases you may not even be anticipating.

Pick Your Target Keywords

How do you think people will search for your web page? The words you imagine them typing into the search box are your target keywords.

For example, say you have a page devoted to stamp collecting. Anytime someone types "stamp collecting," you want your page to be in the top ten results. Then those are your target keywords for that page.

Each page in your web site will have different target keywords that reflect the page's content. For example, say you have another page about the history of stamps. Then "stamp history" might be your keywords for that page.

Your target keywords should always be at least two or more words long. Usually, too many sites will be relevant for a single word, such as "stamps." This "competition" means

your odds of success are lower. Don't waste your time fighting the odds. Pick phrases of two or more words, and you'll have a better shot at success.

Position Your Keywords

Make sure your target keywords appear in the crucial locations on your web pages. The page's HTML title tag is most important. Failure to put target keywords in the title tag is the main reason why perfectly relevant web pages may be poorly ranked. More about the title tag can be found on the [How HTML Meta Tags Work](#) page.

Build your titles around the top two or three phrases that you would like the page to be found for. The titles should be relatively short and attractive. Think of newspaper headlines. With a few words, they make you want to read a story. Similarly, your page titles are like headlines for your pages. They appear in search engine listings, and a short, attractive title may help make users click through to your site.

Search engines also like pages where keywords appear "high" on the page, as described more fully on the [Search Engine Ranking](#) page. To accommodate them, use your target keywords for your page headline, if possible. Have them also appear in the first paragraphs of your web page.

Keep in mind that tables can "push" your text further down the page, making keywords less relevant because they appear lower on the page. This is because tables break apart when search engines read them. For example, picture a typical two-column page, where the first column has navigational links, while the second column has the keyword loaded text.

Humans see that page like this:

[Home](#) **Stamp Collecting**

[Page 1](#)

[Page 2](#) Stamp collection is worldwide experience.

[Page 3](#) Thousands enjoy it everyday, and millions

[Page 4](#) can be made from this hobby/business.

Search engines (and those with old browsers) see the page like this:

[Home](#)

[Page 1](#)

[Page 2](#)

[Page 3](#)

[Page 4](#)

Stamp Collecting

Stamp collection is worldwide experience.

Thousands enjoy it everyday, and millions

can be made from this hobby/business.

See how the keywords have moved down the page? There's no easy way around this, other than to simplifying your table structure. Consider how tables might affect your page, but don't necessarily stop using them. I like tables, and I'll continue to use them.

Large sections of JavaScript can also have the same effect as tables. The search engine reads this information first, which causes the normal HTML text to appear lower on the page. Place your script further down on the page, if possible.

Have Relevant Content

Changing your page titles is not necessarily going to help your page do well for your target keywords if the page has nothing to do with the topic. Your keywords need to be reflected in the page's content.

In particular, that means you need HTML text on your page. Sometimes sites present large sections of copy via graphics. It looks pretty, but search engines can't read those graphics. That means they miss out on text that might make your site more relevant. Some of the search engines will index ALT text and comment information. But to be safe, use HTML text whenever possible. Some of your human visitors will appreciate it, also.

Be sure that your HTML text is "visible." Some designers try to spam search engines by repeating keywords in a tiny font or in the same color at the background color to make the text invisible to browsers. Search engines are catching on to these and other tricks. Expect that if the text is not visible in a browser, then it may not be indexed by a search engine.

Finally, consider "expanding" your text references, where appropriate. For example, a stamp collecting page might have references to "collectors" and "collecting." Expanding these references to "stamp collectors" and "stamp collecting" reinforces your strategic keywords in a legitimate and natural manner. Your page really is about stamp collecting, but edits may have reduced its relevancy unintentionally.

An excellent resource for more about writing copy that naturally pleases search engines is the free [High Rankings Advisor](#) newsletter. Consider signing up for it.

Avoid Search Engine Stumbling Blocks

Some search engines see the web the way someone using a very old browser might. They may not read image maps. They may not read frames. You need to anticipate these problems, or a search engine may not index any or all your web pages.

Have HTML links

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 that lead to major inside pages or sections of your web site. This is 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.

Finally, be sure you do a good job of linking internally between your pages. If you naturally point to different pages from within your site, you increase the odds that search engines will follow links and find more of your web site.

Frames can kill

Some of the major search engines cannot follow frame links. Make sure there is an alternative method for them to enter and index your site, either through meta tags or smart design. For more information, see the tips on using [frames](#).

Dynamic Doorblocks

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.

Build Links

Every major search engine uses link analysis as part of their ranking algorithms. This is done because its very difficult for webmasters to "fake" good links, in the way they might try to spam search engines by manipulating the words on their web pages. As a result, link analysis gives search engines a useful means of determining which pages are good for particular topics.

By building links, you can help improve how well your pages do in link analysis systems. The key is understanding that link analysis is not about "popularity." In other words, it's not an issue of getting lots of links from anywhere. Instead, you want links from good web pages that are related to the topics you want to be found for.

Here's the simple means to find those good links. Go to the major search engines. Search for your target keywords. Look at the pages that appear in the top results. Now visit those pages and ask the site owners if they will link to you. Not everyone will, especially sites that are extremely competitive with you. However, there will be non-competitive sites that will link to you -- especially if you offer to link back.

Why is this system good? By searching for your target keywords, you'll find the pages that the search engines themselves are telling you are good, as evidenced by the fact that they rank well. Hence, links from these pages are more important -- and important for the terms you are interested in -- than links from other pages. In addition, if these pages are top ranked, then they are likely to be receiving many visitors. Thus, if you can gain links from them, you might receive some visitors who initially go to those pages.

Just Say No To Search Engine Spamming

For one thing, spamming doesn't always work with search engines. It can also backfire. Search engines may detect your spamming attempt and penalize or ban your page from their listings.

Also, search engine spamming attempts usually center around being top ranked for extremely popular keywords. You can try and fight that battle against other sites, but then be prepared to spend a lot of time each week, if not each day, defending your ranking. That effort usually would be better spent on networking and alternative forms of publicity, described below.

If those practical reasons aren't enough, how about some ethical ones? The content of most web pages ought to be enough for search engines to determine relevancy without webmasters having to resort to repeating keywords for no reason other than to try and "beat" other web pages. The stakes will simply keep rising, and users will also begin to hate sites that undertake these measures.

Consider search engine spamming against spam mail. No one likes spam mail, and sites that use spam mail services often face a backlash from those on the receiving end. Sites that spam search engines degrade the value of search engine listings. As the problem grows, these sites may face the same backlash that spam mail generates.

Submit Your Key Pages

Most search engines will index the other pages from your web site by following links from a page you submit to them. But sometimes they miss, so it's good to submit the top two or three pages that best summarize your web site.

Don't trust the submission process to automated programs and services. Some of them are excellent, but the major search engines are too important. There aren't that many, so submit manually, so that you can see if there are any problems reported.

Also, don't bother submitting more than the top two or three pages. It doesn't speed up the process. Submitting alternative pages is only insurance. In case the search engine has trouble reaching one of the pages, you've covered yourself by giving it another page from which to begin its crawl of your site.

Be patient. It can take up to a month to two months for your "non-submitted" pages to appear in a search engine, and some search engines may not list every page from your site.

Verify And Maintain Your Listing

Check on your pages and ensure they get listed, in the ways described on the [Check URL](#) page. Once your pages are listed in a search engine, monitor your listing every week or two. Strange things happen. Pages disappear from catalogs. Links go screwy. Watch for trouble, and resubmit if you spot it.

Resubmit your site any time you make significant changes. Search engines should revisit on a regular schedule. However, some search engines have grown smart enough to realize some sites only change content once or twice a year, so they may visit less often. Resubmitting after major changes will help ensure that your site's content is kept current.

Beyond Search Engines

It's worth taking the time to make your site more search engine friendly, because some simple changes may pay off with big results. Even if you don't come up in the top ten for your target keywords, you may find an improvement for target keywords you aren't anticipating. The addition of just one extra word can suddenly make a site appear more relevant, and it can be impossible to guess what that word will be.

Also, remember that while search engines *are* a primary way people look for web sites, but they are not the *only* way. People also find sites through word-of-mouth, traditional advertising, the traditional media, newsgroup postings, web directories and links from other sites. Many times, these alternative forms are far more effective draws than are search engines.

Finally, know when it's time to call it quits. A few changes may be enough to make you tops in one or two search engines. But that's not enough for some people, and they will invest days creating special pages and changing their sites to try and do better. This time could usually be put to better use pursuing non-search engine publicity methods.

Don't obsess over your ranking. Even if you follow every tip and find no improvement, you still have gained something. You will know that search engines are not the way you'll be attracting traffic. You can concentrate your efforts in more productive areas, rather than wasting your valuable time.

How To Use HTML Meta Tags

By [Danny Sullivan](#)

Editor, [SearchEngineWatch.com](#)

Updated: October 14, 2002

Want to get a top ranking in search engines? No problem! All you need to do is add a few magical "meta tags" to your web pages, and you'll skyrocket to the top of the listings.

If only it were so easy. Let's make it clear:

- *Meta tags are not a magic solution.*
- *Meta tags are not a magic solution.*
- *Meta tags are not a magic solution.*

Meta tags have never been a guaranteed way to gain a top ranking on crawler-based search engines. Today, the most valuable feature they offer the web site owner is the ability to control to some degree how their web pages are described by some search engines. They also offer the ability to prevent pages from being indexed at all. This page explores these and other meta tag-related features in more depth.

Meta Tag Overview

What are meta tags? They are information inserted into the "head" area of your web pages. Other than the [title tag](#) (explained below), information in the head area of your web pages is not seen by those viewing your pages in browsers. Instead, meta information in this area is used to communicate information that a human visitor may not be concerned with. Meta tags, for example, can tell a browser what "character set" to use or whether a web page has self-rated itself in terms of adult content.

Let's see two common types of meta tags, then we'll discuss exactly how they are used in more depth:

```
<HEAD>
<TITLE>Stamp Collecting World</TITLE>
<META name="description" content="Everything you wanted to know
about stamps, from prices to history.">
<META name="keywords" content="stamps, stamp collecting,
stamp history, prices, stamps for sale">
</HEAD>
```

In the example above, you can see the beginning of the page's "head" area as noted by the HEAD tag -- it ends at the portion shown as /HEAD.

Meta tags go in between the "opening" and "closing" HEAD tags. Shown in the example is a TITLE tag, then a META DESCRIPTION tag, then a META KEYWORDS tag. Let's talk about what these do.

The Title Tag

The HTML title tag isn't really a meta tag, but it's worth discussing in relation to them. Whatever text you place in the title tag (between the TITLE and /TITLE portions as shown in the example) will appear in the reverse bar of someone's browser when they view the web page. For instance, within the title tag of this page that you are reading is this text:

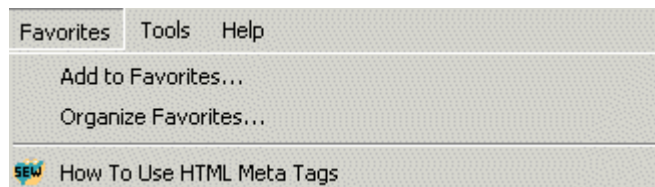
How To Use HTML Meta Tags

If you look at the reverse bar in your browser, then you should see that text being used, similar to this:



Some browsers also supplement whatever you put in the title tag by adding their own name, as you can see Microsoft's Internet Explorer doing in the picture above.

The title tag is also used as the words to describe your page when someone adds it to their "Favorites" or "Bookmarks" lists. For instance, if you added this page to your Favorites in Internet Explorer, it would show up like this:



How did that little Search Engine Watch logo also show up? Everyone always asks. The article below provides more help:

Creating Your Own Favicon.ico Icon For IE5

Web Developer's Journal, March 7, 2000

<http://www.webdevelopersjournal.com/articles/favicon.html>

But what about search engines! The title tag is crucial for them. The text you use in the title tag is one of the most important factors in how a search engine may decide to rank your web page (see the [Search Engine Placement Tips](#) section for more details). In addition, all major crawlers will use the text of your title tag as the text they use for the title of your page in your listings.

For example, this is how Teoma lists the page you are reading:



You can see that the text "How To Use HTML Meta Tags" is used as the hyperlinked title of this page's listed in Teoma's results.

In review, think about the key terms you'd like your page to be found for in crawler-based search engines, then incorporate those terms into your title tag in a short, descriptive fashion. That text will then be used as your title in crawler-based search engines, as well as the title in bookmarks and in browser reverse bars.

The Meta Description Tag

The meta description tag allows you to influence the description of your page in the crawlers that support the tag (these are listed on the [Search Engine Features](#) page).

Look back at the [example](#) of a meta tag. See the first meta tag shown, the one that says "name=description"? That's the meta description tag. The text you want to be shown as your description goes between the quotation marks after the "content=" portion of the tag (generally, 200 to 250 characters may be indexed, though only a smaller portion of this amount may be displayed).

For this page you are reading, I would like it described in a search engine's listings like this:

This tutorial explains how to use HTML meta tags, with links to meta tag generators and builders. From SearchEngineWatch.com, a guide to search engine submission and registration.

Will this happen? Not with every search engine. For example, Google ignores the meta description tag and instead will automatically generate its own description for this page. Others may support it partially. For instance, let's see again how this page is listed in Teoma:

Top 200 of about 569 (Showing 1-10):

[How To Use HTML Meta Tags](#)

This tutorial explains how to use HTML meta tags, with links to meta tag generators and ... How To Use HTML Meta Tags By Danny Sullivan Editor, ...
searchenginewatch.com/webmasters/meta.html

You can see that the first portion of the page's description comes from the meta description tag, then there's an ellipse (...), and the remaining portion is drawn from the body copy of the page itself.

In review, it is worthwhile to use the meta description tag for your pages, because it gives you some degree of control with various crawlers. An easy way to do this often is to take the first sentence or two of body copy from your web page and use that for the meta description content.

The Meta Keywords Tag

The meta keywords tag allows you to provide additional text for crawler-based search engines to index along with your body copy. How does this help you? Well, for most major crawlers, it doesn't. That's because most crawlers now ignore the tag. The few supporting it can be found on the [Search Engine Features](#) page).

The meta keywords tag is *sometimes* useful as a way to reinforce the terms you think a page is important for ON THE FEW CRAWLERS THAT SUPPORT IT. For instance, if you had a page about stamp collecting -- AND you say the words stamp collecting at various places in your body copy -- then mentioning the words "stamp collecting" in the meta keywords tag MIGHT help boost your page a bit higher for those words.

Remember, if you don't use the words "stamp collecting" on the page at all, then just adding them to the meta keywords tag is extremely unlikely to help the page do well for the term. The text in the meta keywords tag, FOR THE FEW CRAWLERS THAT SUPPORT IT, works in conjunction with the text in your body copy.

The meta keyword tag is also sometimes useful as a way to help your page come up for synonyms or unusual words that don't appear on the page itself. For instance, let's say you had a page all about the "Penny Black" stamp. You never actually say the word "collecting" on this page. By having the word in your meta keywords tag, then you may help increase the odds of coming up if someone searched for "penny black stamp collecting." Of course you would greater increase the odds if you just used the word "collecting" in the body copy of the page itself.

Here's another example. Let's say you have a page about horseback riding, and you've written your page using "horseback" as a single word. You realize that some people may instead search for "horse back riding," with "horse back" in their searches being two separate words. If you listed these words separately in your meta keywords tag, THEN **MAYBE FOR THE FEW CRAWLERS THAT SUPPORT IT**, your page might rank better for "horse back" riding. Sadly, the best way to ensure this would be to write your pages using both "horseback riding" and "horse back riding" in the text -- or perhaps on some of your pages, use the single word version and on others, the two word version.

I'm using all these capital letters on purpose. Far too many people new to search engine optimization obsess with the meta keywords tag. **FEW** crawlers support it. For those that do, it **MIGHT! MAYBE! PERHAPS! POSSIBLY! BUT WITH NO GUARANTEE!** help improve the ranking of your page. It also may very well do nothing for your page at all. In fact, repeat a particular word too often in a meta keywords tag and you could actually harm your page's chances of ranking well. Because of this, I strongly suggest that those new to search engine optimization not even worry about the tag at all.

Even those who are experienced in search engine optimization may decide it is no longer worth using the tags. Search Engine Watch doesn't. Any meta keywords tags you find in the site were written in the past, when the keywords tag was more important. There's no harm in leaving up existing tags you may have written, but going forward, writing new tags probably isn't worth the trouble. The article below explores this in more detail:

[Death Of A Meta Tag](#)

The Search Engine Report, Oct. 1, 2002

Still want to use the meta keywords tag? OK. Look back at the [opening example](#). See the second meta tag shown, the one that says "name=keywords"? That's the meta keywords tag. The keywords you want associated with your page go between the quotation marks after the "content=" portion of the tag. Generally, up to 1,000 characters may be indexed.

In the example, you'll see that for organizational purposes, keywords and multikeyword phrases are separated by commas. Commas are not required in the tag, however. If you're really interested in trying to understand why you might leave out commas and keep keywords in a particular order, then read the [More About The Meta Keywords Tag](#) article, if you are a Search Engine Watch [member](#).

Meta Robots Tag

One other meta tag worth mentioning is the robots tag. This lets you specify that a particular page should NOT be indexed by a search engine. To keep spiders out, simply add this text between your head tags on each page you don't want indexed. The format is shown below (click on the picture if you want to copy and past the HTML for your own use):

```
<HEAD>  
<TITLE>Page I Don't Want In Search Engines</TITLE>  
<META NAME="ROBOTS" CONTENT="NOINDEX">  
</HEAD>
```

You do NOT need to use variations of the meta robots tag to help your pages get indexed. They are unnecessary. By default, a crawler will try to index all your web pages and will try to follow links from one page to another.

Most major search engines support the meta robots tag. However, the robots.txt convention of blocking indexing is more efficient, as you don't need to add tags to each and every page. See the [Search Engines Features](#) page for more about the robots.txt file. If you use do a robots.txt file to block indexing, there is no need to also use meta robots tags.

The meta robots tag also has some extensions offered by particular search engines to prevent indexing of multimedia content. The article below talks about this in more depth and provides some links to help files. Search Engine Watch members should follow the link from the article to the members-only edition for extended help on the subject.

Image Search Faces Renewed Legal Challenge

The Search Engine Report, August 22, 2001

<http://searchenginewatch.com/sereport/01/08-images.html>

Other Meta Tags

There are many other meta tags that exist beyond those explored in this article. For example, if you were to view the source code of this web page, you would find "author," "channel" and "date" meta tags. These mean nothing to web-wide crawlers such as Google. They are specifically for an internal search engine used by Search Engine Watch to index its own content.

There are also "Dublin Core" meta tags. The intent is that these can be used for both "internal" search engines and web-wide ones. However, no major web-wide search engine supports these tags. More about them can be found below:

Dublin Core Metadata Initiative

<http://dublincore.org/>

Dublin Core - Tagging the Web for better search and retrieval

WebReference.com, Nov. 5, 2000

<http://webreference.com/xml/column24/index.html>

An introduction to the fifteen Dublin Core meta tags.

How about the meta revisit tag? This tag is not recognized by the major search engines as a method of telling them how often to automatically return. They have never supported it.

In Conclusion

Overall, just remember this. Of all the meta tags you may see out there:

- **Meta Robots:** This tag enjoys full support, but you only need it if you DO NOT want your pages indexed.
- **Meta Description:** This tag enjoys much support, and it is well worth using.
- **Meta Keywords:** This tag is only supported by some major crawlers and probably isn't worth the time to implement.
- **Meta Everything Else:** Any other meta tag you see is ignored by the major crawlers, though they may be used by specialized search engines.

More Resources

At the bottom of this page are more resources about meta tags, including tutorials and meta tag building applications. But first...

If you've been following the "Next" buttons to read the numbered sections of the [Search Engine Submission Tips](#) guide in order, you've now reached the last page.

Congratulations!

There's still more information you might find helpful, however. Please review the pages listed under the "[Optional But Helpful](#)" section for additional assistance with search engine marketing issues.

In addition, do consider becoming a Search Engine Watch [member](#), for access to even more information on search engine marketing issues.

Just started learning from this page? Don't worry -- [click here](#) to go to the beginning of the guide.

Now, here are those additional meta tag resources and articles...

Meta Tag Generators, Builders and Evaluators

Site Announce Meta Tag Generator

<http://www.siteannounce.com/tools/meta-tag-generator/>

Simple, free online form that creates basic meta tags for your web pages.

SiteUp's Meta-Tag Generator

<http://www.siteup.com/meta.html>

This is a software-based package for Windows that creates meta tags. It is a freeware package -- no registration fee required.

Meta Tag Builder

<http://vancouver-webpages.com/META/mk-metas.html>

This form allows you to create very complicated meta tags using much more than the keywords and description tags, if you wish. Note that it will place a commented credit line into the tag. This can easily be removed, if you wish.

Articles About Meta Tags

[Death Of A Meta Tag](#)

The Search Engine Report, Oct. 1, 2002

Now supported by only one major crawler-based search engine, the value of adding meta keywords tags to pages seems little worth the time. A look at how we gained and lost the meta keywords tag.

Are search engines dead?

WDVL, June 26, 2000

http://www.wdvl.com/Internet/Dead_SearchEngines/

A look at the RDF meta data structure and how search engines aren't using it. Why not? Experience has taught them that meta data often cannot be trusted.

[The New Meta Tags Are Coming -- Or Are They?](#)

The Search Engine Report, Dec. 4, 1997

The proposed Resource Description Framework, or RDF, would provide a new way of describing web pages via meta data. There are high hopes for what it may accomplish, but support by the search engines isn't certain. Also learn more about the Dublin Core meta tags, which may be incorporated into the system.

[What Is Meta Content Framework](#)

Search Engine Watch, June 1997

Summary of a Netscape-backed meta data proposal now outdated by the rise of RDF (see above).

Checking Your URL

By [Danny Sullivan](#)

Editor, [SearchEngineWatch.com](#)

Updated: Oct. 26, 2001

Some crawler-based search engines make it easy to confirm that your web page is in their index. With others, it can be more difficult. Below are the best ways to find your web pages in the major crawler-based search engines.

Please note that these commands can also be useful for web searchers who wish to refine their queries, as explained more in the [Site Search](#) section of the Power Searching For Anyone page.

The Search Engine Alliances page explains where some search engines get or give their listings to others. Also, Search Engine Watch members have access pages that explain in depth the different data sources that each major search engine users and how they display information.

AltaVista

AltaVista has commands that can be used to easily narrow your search to a single URL or to pages within a particular web site. These commands can also be combined with query terms by those who wish to refine their search results.

URL Search

To find a single page listed in AltaVista's crawler-based index, you can use the "url:" command. Simply preface the URL you wish to locate with this command, such as:

url:http://searchenginewatch.com/webmasters/meta.html

If the URL is in the index, it will be displayed. You can also use this command to find pages within a particular section of a web site. For example, this:

url:http://searchenginewatch.com/webmasters/

would list all the pages from within the /webmasters/ area of Search Engine Watch. This can be a useful way to find all the pages from your web site, if it resides within someone else's domain.

Site Search

To locate all the URLs listed from a particular web site, use the "host:" command, such as:

host:searchenginewatch.com

Use only the actual domain name. Omit the http:// prefix. Also, be aware that using the www prefix can make a difference. For instance, the query below:

host:www.searchenginewatch.com

would bring back only the pages AltaVista has found from Search Engine Watch with the www prefix. However, Search Engine Watch can also be reached without the www prefix. In fact, this is the more common way that people come to Search Engine Watch. Consequently, AltaVista has actually indexed many more pages from the site without the www prefix. To see these pages, the first example shown would have to be used.

AllTheWeb.com/FAST Search

At FAST Search, commands can be used to find a single URL or multiple web pages from a particular site, as explained below:

URL Search

To find a single page listed in FAST's crawler-based index, you can use the "url.all:" command:

url.all:searchenginewatch.com/webmasters/meta.html

This command will also work to bring up a single URL that is listed in the FAST-powered results used by Lycos.

Site Search

To locate all the URLs listed from a particular web site, use the "url.host:" command, such as **url.host:searchenginewatch.com**

Use only the actual domain name. Omit the http:// prefix. Also, be aware that using the www prefix can make a difference, as described with AltaVista.

Google

At Google, commands can be used to find a single URL or multiple web pages from a particular site, as explained below:

URL Search

To find a single page listed in Google's crawler-based index, you can use the "allinurl:" command, such as:

allinurl:searchenginewatch.com/webmasters/meta.html

The allinurl command works the same as with AltaVista, which means you can also use it to find pages within a particular section of a web site. Be sure to omit the http:// prefix.

Please note that if you are trying to find web pages with both words in the URL and in the document itself, you'll need to use the special "inurl" command. This is explained more in the [URL Search](#) section of the Power Searching For Anyone page.

Site Search

To locate all the URLs listed from a particular web site, use the "site:" command in combination with a word or words that you know appear on all the pages. For example:
site:searchenginewatch.com searchenginewatch
would bring up all (or nearly all) of the pages Google lists from Search Engine Watch, because all the pages should have the word "searchenginewatch" on them as part of the footer text.

You must use the site command in combination with a search term. It will not work, otherwise.

Inktomi

Inktomi powers some of the results used by a variety of different search engines. Below is how to locate a single or multiple URLs within Inktomi powered-listings.

URL Search

To find a particular URL listed in Inktomi's crawler-based index, you can use the "originurl:" command. Simply preface the URL you wish to locate with this command, such as:

originurl:http://searchenginewatch.com/webmasters/meta.html

If the URL is in the index, it will be displayed. This command has been tested to work on the following Inktomi-powered services:

- AOL Search
- GoTo
- HotBot

The originurl command will bring up an individual URL listed in the Inktomi-powered results of these services. It does not work at iWon, LookSmart or MSN Search.

Keep in mind that not all Inktomi partners tap into the entire Inktomi database. That's one reason why you may find a URL at one service but not at another.

Site Search

To locate all the URLs listed from a particular web site, use the "domain:" command, such as:
domain:searchenginewatch.com

Use only the actual domain name. Omit the *http://* prefix. As explained [above](#) for AltaVista, using the www prefix can also make a difference.

Unfortunately, the domain command works inconsistently at different Inktomi-powered services. Here's a rundown:

HotBot: Use the command. If you have any listings in the Open Directory, these will be shown first. You'll know these are Open Directory listings because they will have a "More like this" link underneath them. By pass these and find a listing for your site that instead has a "See results from this site only" link. Select that link, and you'll see all the pages listed in Inktomi from your web site.

iWon & LookSmart: Using the command will list all pages from a web site.

At the following services, the command fails to operate because "clustering" (as explained on the [Search Assistance Features](#) page) prevents you from seeing more than a few pages from your site.

- AOL Search
- GoTo
- MSN Search

Directory Search

Inktomi has a special command that lets you find pages within a specific area of a web site. This is the "originurlpath:" command, and you use it in combination with the domain command, such as:

domain:searchenginewatch.com originurlpath:webmasters

This would find pages from within the /webmasters/ area of Search Engine Watch. In other words, everything within this area would be listed:

http://searchenginewatch.com/reports/ekgs/

Directories: Yahoo, LookSmart & The Open Directory

Directories are search engines that are powered by human beings, rather than by crawling the web. Because humans are involved, directories tend to list only a few pages per web site. This means that you probably won't be needing to make use of special site or URL commands to locate your listings. In fact, of the three major directories, only Yahoo has any specific command like this. At Yahoo, you can use the "u:" command to locate specific URLs, like this:

u:searchenginewatch.com

That would bring up any pages from Yahoo's human-compiled listings that contain "searchenginewatch.com" within the URL, if it is done from the [Yahoo Directory](#) page, as opposed to the regular Yahoo home page (which brings back Google results).

At the web's two other major directories, LookSmart and the Open Directory, you'll find that searching for your domain or a portion of your domain should bring up many or all of your listings.

For example, by entering "searchenginewatch.com" or "searchenginewatch," I would be able to find most of my human-compiled listings in both places.

LookSmart also provides a detailed guide to locating your URL within its service and the listings it provides to partners:

How to Find Your Listing in the LookSmart Network

<http://submit.looksmart.com/info.jhtml?page=find>

Outsourcers

All the search engines and directories mentioned above produce their own listings. However, there remain some major search engines that simply outsource to produce the results they provide. For example, MSN Search uses information from both LookSmart and Inktomi. Trying to check your listings at such a place is difficult, because two or more data sets are involved. Below is a guide to what happens as such places

AOL Search

Use tips described above for the Open Directory to find your Open Directory listings and for Inktomi to find your Inktomi listings.

HotBot

Using the Inktomi site search command described above will bring up both your Open Directory and Inktomi listings. There is no command to bring up all your pages included in HotBot's Direct Hit powered "Top Ten" listings.

Lycos

Use tips described above for FAST to find your FAST listings. There is no way to find all your listings from the Open Directory or Direct Hit information the service uses.

MSN Search

Use tips described above for Inktomi to find your Inktomi listings. There is no way to find all your listings from LookSmart.

Netscape

Use tips described above for the Open Directory to find your Open Directory listings and for Google to find your Google listings.

Other Resources

There are services that can check search engines for your URL automatically, including checking on how they appear in relation to particular keyword phrases. A list of these is maintained for [Search Engine Watch members](#) in the [Position Checking/Tracking section](#) of the Search Engine Optimization Toolbox page.

Also, this page gives you just a taste of some of the powerful searches that can be done with search engines. See the [Power Searching](#) page for an at-a-glance guide to other types of searches.

Measuring Link Popularity

By [Danny Sullivan](#)

Editor, [SearchEngineWatch.com](#)

Updated: Oct. 26, 2001

The best way to discover how people are finding your web site is to analyze your site's activity logs, a topic which is covered in depth on the [Keywords Used To Find Your Web Site](#) page available to Search Engine Watch [members](#).

Those unable to analyze their logs can instead use search engines to track down referral links. In particular, this method gives you an idea of how "popular" a search engine believes your site to be. That's important for those search engines that rank sites in part by the site's link popularity.

Below is described how to measure link popularity at popular crawler-based search engines.

Be aware that "popularity" is only one part of the link analysis systems that search engines such as Google use to rank web pages. The quality and context of links is also taken into account, rather than sheer numbers.

AltaVista, Google & Northern Light

To search for pages linking to your site, simply enter your domain this way:

link:searchenginewatch.com

This will return all pages that have hyperlinks linking to the searchenginewatch.com web site.

This would include links such as:

http://searchenginewatch.com

http://searchenginewatch.com/webmasters/

http://searchenginewatch.com/webmasters/popularity.html

You can narrow the search to a particular URL by being more specific. For example:

link:searchenginewatch.com/webmasters/popularity.html

Some pages from within your site probably link to each other. To eliminate these, use the `-url` command (see the [Checking Your Listings](#) page for more about this command), such as:

`link:searchenginewatch.com -url:searchenginewatch.com`

This only works at AltaVista and Northern Light. There's no way to count links to a page and also simultaneously subtract links from within your own site to a page with Google.

Note that the URLs entered after `link:` don't include the `www` prefix. It's not necessary. The format shown will find links with or without the prefix. The `http://` prefix is not necessary, either.

AllTheWeb.com

To search for pages linking to your URLs, use the `link.all:` command, such as:

`link.all:searchenginewatch.com`

Inktomi

Inktomi has a `linkdomain:` command that can be used to measure link popularity within its listings. However, because different Inktomi-partners may not implement this command correctly, you may find it doesn't always work in places where Inktomi results are used. How it works for major Inktomi partners is covered below.

To find all the pages linking to a particular domain, use `linkdomain:`, such as:

`linkdomain:searchenginewatch.com`

To eliminate your own pages, use `-domain` (see the [Checking Your Listings](#) page for more about this command), such as:

`linkdomain:searchenginewatch.com -domain:searchenginewatch.com`

The command works above as shown for Inktomi results at:

- AOL Search
- HotBot
- iWon
- MSN Search

It does not work properly at LookSmart, because LookSmart will fail to show the entire count of links to a particular domain. Instead, it will max out at 1,000.

The `linkdomain:` command only works to measure the link popularity within an entire web site. In contrast, if you need to find just the links to a particular page, you'll need to use the advanced search pages offered by HotBot and MSN Search. Enter the full URL you wish to locate, including the `http://` prefix. Then look for the "links to URL" or similar option. Conduct the search, and you'll be shown the links to that particular page.

Other Resources

Marketleap Visibility Index

<http://linkpop.marketleap.com>

Enter your URL, along with up to three comparison URLs, then see at a glance how you measure up in terms of sheer numbers of links to some major sites on the web. Free, easy to understand and use! You can then click on the report to see link count listings from the major search engines that are queried.

Link Popularity Check

<http://www.linkpopularitycheck.com>

Enter your URL, along with up to three comparison URLs, then see with bar charts how these particular sites compare to each other. You can also see the link count listings from the major search engines that are queried. Free, easy to use and you can have results emailed to you on a regular basis.

LinkPopularity

<http://www.linkpopularity.com/>

Free and simple service that helps you generate link lists from three major search engines. Results can also be emailed each month.

Compute Your Own Web Traffic Rank

<http://www.useit.com/alertbox/relativeranking.html>

This isn't really about measuring link popularity, but it is an interesting method of determining the overall popularity of your site. In short, Jakob Nielsen explains that dividing your site's page views by the page views of the web's most popular site tells you how popular your site is.

Search Engines And Frames

By [Danny Sullivan](#)

Editor, SearchEngineWatch.com

Updated: December 5, 2000

Note: This tutorial assumes you are familiar with frames. If not, there are links below to general frames information.

Search engines have a tough time with frames. Using frames either prevents them from finding pages within a web site, or it causes them to send visitors into a site without the proper frame "context" being established. Both problems can be corrected, with a little foresight by webmasters.

Say More Than Sorry To Search Engines

Many sites use frames for navigation, and the fictional "Wonderful World of Stamp Collecting" site in this tutorial is a typical example of this. It has three frames: one for navigation, one for the site's title and one as the main content window.

View the example, then return to this page (use your back button or click on the big, ugly link that says "BACK TO THE TUTORIAL."

[Example 1](#)

You saw a single page with three frames appear. The content for these frames actually came from three different pages and was blended together according to the instructions of a fourth "master" page, also called the "frameset" page.

In contrast, most search engine spiders will only see the master page. Just like an old browser, they don't understand the instructions on how to produce the frame layout. These are ignored, and only information within the *noframes* tags are read (information which a frames-capable browser will ignore).

So what do frame-challenged search engines see in our example? It's pretty typical of most frames-based sites:

Sorry! You need a frames-browser to view this site.

Obviously, we need to provide search engines with a much better description of the site than this. There's also another problem. There are no links within the noframes area to pages within the site. That means the search engines won't crawl past our master page. We could have hundreds of information-packed pages inside the site, but this simple mistake essentially makes them invisible to many search engines.

Your Friend, The NOFRAMES Tag

One solution to the description problem is to add meta tags to the master page (read more about these on the [Meta Tags](#) page). However, meta tags are only a partial solution, because not all the search engines support them. Meta tags also don't help human beings that want to view your site but lack frames capability.

We can help both search engines and humans with some smart design. Take a look at the next example, then come back:

[Example 2](#)

The example looks just like our original page, unless you are a search engine or are using an old browser. For them, we have added this additional information in the noframes area:

The Wonderful World of Stamp Collecting

Many people consider stamp collection to be a hobby, but it can be more than that. Stamp collecting can be a business, an art form or even a topic of conversation. Stamp history is absolutely fascinating. It's just one of the many topics you'll find in this site.

If you are viewing this text, your browser lacks the ability to read frames. Don't worry, you can still enjoy our site. All the pages can be viewed from the contents page. Please come inside!

[Contents](#)

Excellent. Now we have some descriptive text that any search engine can read, not just those that support meta tags. Furthermore, we've created a way for them (and humans) to get inside the site. That's via the contents link at the bottom of the description. Go ahead and click on it.

What appears is the information normally placed in the navigation window. From here, search engines can crawl through the rest of the site.

The noframes information can go immediately after the first frameset tag, if you want the text to be as high as possible on the page. Do not place it above the first frameset tag. This will disable the frame information in Netscape, although Internet Explorer is unaffected and will display the frames correctly.

Including *body* tags within the noframes tags works with either browser. It helps ensure that the body tags exist for any browser or search engine that might require them.

Also remember to give your frame pages a title, even though it won't appear when the pages are viewed in their proper context. Titles are the most important element that a search engine indexes, so you want all your pages to have them.

Reestablishing The Frame Context

We're almost finished. There's only one downside to the work we've done so far. Now that search engines can index the individual pages within the site, visitors will come to them without the proper "context" being established. In other words, they won't see the page within one of the three normal frames. Instead, the page will stand alone.

This can be a big problem. Some webmasters never anticipate that pages may be viewed outside the frames context. These pages may lack links back into the site, essentially "trapping" inexperienced web surfers on the page.

This situation will happen even if you ignore all the advice above about making the site more accessible to search engines. That's because some, such as AltaVista, are frames-capable. They will crawl your site and index each individual page, and thus visitors will enter the site out of context.

Fortunately, an easy answer is to include a "home" link at the bottom of all your pages. View the example, then click on the "home" link at the bottom. Then return here.

[Example 3](#)

If all went well, you got an unframed page just as someone might who found it via a search engine. By following the only link present, they are brought back to the home page, where the frames can be drawn.

If you do this, make sure the link is created using the target `_top` command, such as:

```
<a href="index.html" TARGET = "_top">Home</a>
```

Without the command, those entering your site the "normal" way and then clicking on the home link will get a new set of three frames drawn within the main content frame.

Another method is to use JavaScript. Here's one such JavaScript solution (there are many). Add this to your frame pages (anywhere between either the head or the body tags will work):

```
<SCRIPT LANGUAGE="JavaScript">
<!--
if (top == self) self.location.href = "FRAMESET PAGE NAME HERE";
// -->
</SCRIPT>
```

You need to replace the section that says FRAMESET PAGE NAME HERE with the name of the master page that has the appropriate frameset information. In our example, that's the index.html page, so the script reads as so:

```
<SCRIPT LANGUAGE="JavaScript">
<!--
if (top == self) self.location.href = "index.html";
// -->
</SCRIPT>
```

To see it in action, view the example below. It's the same as the out-of-context page as in Example 3, but the script forces the frames to be redrawn.

[Example 4](#)

Be aware that if you use this JavaScript tag, people will be unable to use their browser's "Back" button to return to the preceding page unless they hit it quickly twice in succession. However, once they navigate to another page within your site, the Back button will work normally.

Naturally, the tips above may not be for everyone. However, with a little trial and error, you can make your frame sites accessible to both search engines and non-frame visitors. Moreover, the above examples require a bare minimum of work and no additional pages to essentially create a frames and non-frames site.

Frames Resources

Framing the Web

<http://webreference.com/dev/frames/>

An excellent introduction and tutorial on using frames

W3C Frames in HTML documents

<http://www.w3.org/TR/REC-html40/present/frames.html>

The official HTML 4.0 specs on frames, though they fail to use the body tags within noframes examples, which is incorrect, in my opinion. It doesn't hurt to have them and probably helps, so use them.

Frames and Framesets: Netscape HTML Tag Reference

<http://developer.netscape.com/docs/manuals/htmlguid/tags11.htm>

Information from Netscape regarding frames tags.

A Compendium of HTML Elements

<http://www.htmlcompendium.org/>

The ABCs of HTML tags, including frame-related tags.

WebDeveloper.com

<http://www.webdeveloper.com/>

The JavaScript frames solution described above came from WebDeveloper's [Dr. Website feature](#). See what other solutions to your HTML problems the site has to offer.

About HTML Guide - Help With Frames

<http://html.about.com/msubframes.htm>

A cornucopia of links to advice on dealing with frames.

NCSA Mosaic

<ftp://ftp.ncsa.uiuc.edu/Web/Mosaic/Windows/Archive/>

Want to see the web the way a search engine does? Download an old version of NCSA Mosaic -- the older the better. Any version listed via this link are incapable of displaying frames. Pick one and use it to see what you're saying in your noframes area.

What Is A Bridge or Doorway Page?

By [Danny Sullivan](#), Editor

SearchEngineWatch.com

Page updated: June 2, 2001

Webmasters are sometimes told to submit "bridge" pages or "doorway" pages to search engines to improve their traffic. Doorway pages are created to do well for particular phrases. They are also known as portal pages, jump pages, gateway pages, entry pages and by other names.

Doorway pages are easy to identify in that they have been designed primarily for search engines, not for human beings. This page explains how these pages are delivered technically, and some of the problems they pose.

Low Tech Delivery

There are various ways to deliver doorway pages. The low-tech way is to create and submit a page that is targeted toward a particular phrase. Some people take this a step further and create a page for each phrase and for each search engine.

One problem with this is that these pages tend to be very generic. It's easy for people to copy them, make minor changes, and submit the revised page from their own site in hopes of mimicking any success. Also, the pages may be so similar to each other that they are considered duplicates and automatically excluded by the search engine from its listings.

Another problem is that users don't arrive at the goal page. Say they did a search for "golf clubs," and the doorway page appears. They click through, but that page probably lacks detail about the clubs you sell. To get them to that content, webmasters usually propel visitors forward with a prominent "Click Here" link or with a fast meta refresh command.

By the way, this gap between the entry and the goal page is where the names "bridge pages" and "jump pages" come from. These pages either "bridge" or "jump" visitors across the gap.

Some search engines no longer accept pages using fast meta refresh, to curb abuses of doorway pages. To get around that, some webmasters submit a page, then swap it on the server with the "real" page once a position has been achieved.

This is "code-swapping," which is also sometimes done to keep others from learning exactly how the page ranked well. It's also called "bait-and-switch." The downside is that a search engine may revisit at any time, and if it indexes the "real" page, the position may drop.

Another note here: simply taking meta tags from a page ("meta jacking," as Infoseek calls it), does not guarantee a page will do well. In fact, sometimes resubmitting the exact page from another location does not gain the same position as the original page.

There are various reasons why this occurs which go beyond this article, but the key thing to understand is that you aren't necessarily finding any "secrets" by viewing source code, nor are you necessarily giving any away.

Agent Delivery

The next step up is to deliver a doorway page that only the search engine sees. Each search engine reports an "agent" name, just as each browser reports a name.

The advantage to agent name delivery is that you can send the search engine to a tailored page yet direct users to the actual content you want them to see. This eliminates the entire "bridge" problem altogether. It also has the added benefit of "cloaking" your code from prying eyes.

Well, not quite. Someone can telnet to your web server and report their agent name as being from a particular search engine. Then they see exactly what you are delivering. Additionally, some search engines may not always report the exact same agent name, specifically to help keep people honest.

IP Delivery / Page Cloaking

Time for one more step up. Instead of delivering by agent name, you can also deliver pages to the search engines by IP address, assuming you've compiled a list of them and maintain it.

Everyone and everything that accesses a site reports an IP address, which is often resolved into a host name. For example, I might come into a site while connected to AOL, which in turn reports an IP of 199.204.222.123 (FYI, that's not real, just an example). The web server may resolve the IP address into an address: ww-tb03.proxy.aol.com, for example.

If you deliver via IP address, you guarantee that only something coming from that exact address sees your page. Another term for this is page cloaking, with the idea that you have cloaked your page from being seen by anyone but the search engine spiders.

More Resources

[More About Doorway Pages](#)

Provides more information about doorway pages, including tips on avoiding problems with search engines. Available to Search Engine Watch [members](#).

[Page Cloaking](#)

Explains issues and some basic technical details of enabling a page "cloaking" system, which often goes hand-in-hand with doorway page efforts. Available to Search Engine Watch [members](#).

[Search Engines & Legal Issues](#)

See the pagejacking section for articles involving doorway pages that have resulting in legal action.

How Search Engines Regionalize

By [Danny Sullivan](#)

Editor, [SearchEngineWatch.com](#)

May 8, 2000

None of the major search engines have "United States" editions. They make no attempt to only list pages from the United States. Nonetheless, the major search engines tend to be dominated by US-oriented content. Furthermore, their search interfaces are in English and often highlight news or other events of interest to a US audience.

In contrast, regional editions aim to serve those in particular countries or regions of the world. They are created in a variety of ways, from minor cosmetic changes to full-blown content development. The common methods are discussed below.

Regional Interface

Creating a regional interface can be as simple as presenting the same search engine look-and-feel in the appropriate language for a particular country. For example, all the instruction text might be changed to French to create a version for France.

Changing the interface does not affect the results. Unless more modifications are done, a search on a regional edition will provide the same results as using the main service.

Domain Filtering

In domain filtering, a search engine's "world" index is filtered so that only sites from a particular country's domain will appear. For example, the United Kingdom's domain is .uk, so UK-specific web sites should end in that domain, such as:

companyname.co.uk

governmentagency.gov.uk

universityname.ac.uk

By filtering out all pages except those from .uk domains, a search engine can create a passable UK-specific edition.

In some cases, the filtering is expanded. For example, a search with Excite Germany also returns matches from sites in Austrian and Swiss domains.

The problem with domain filtering is that many companies outside the US have registered .com domains. That means these companies must be manually included in the results, otherwise they may be accidentally excluded from the regional search engine's listings.

A bigger problem will occur if new top level domains such as .web come into existence. These will be completely country-independent. Domain filtering will then become difficult to maintain. Additionally, some country-specific domains have already been devalued by those using them for other purposes. See the [Goodbye Domain Names, Hello RealNames?](#) article for more about this issue.

Domain Crawling

In domain crawling, a search engine maintains both a "world" index plus a "country" index which provides greater regional coverage. Some of the same pages may be listed in both, but the country index may have greater depth.

Human Categorization

Search engines like Yahoo, LookSmart and the Open Directory that use human beings create regional listings through classification. Sites relevant to a particular country are listed within a country-specific edition of the directory.

Language Specific

Many of the major crawler-based search engines now have their spiders check pages for common words and markers of specific languages. If found, a page is "tagged" internally as being from that language. When users search, they get only these pages

More Information

[More About Countries And Languages](#)

Provides more information about how to build web sites that can please country or language-specific search engines. Available to Search Engine Watch members.

[Click here to learn more about becoming a member](#)

[Search Links: Regional Search Engines](#)

Lists major regional search engines around the world.

Related Articles

[Google Speaks Languages, WAP, Adds Other Features](#)

The Search Engine Report, May 3, 2000

Covers new language support offered by Google.

[GeoSearch Comes To Northern Light](#)

The Search Engine Report, May 3, 2000

Information on a new geographical search option at Northern Light.

[Goodbye Domain Names, Hello RealNames?](#)

The Search Engine Report, May 3, 2000

Discusses how the country domain codes are being devalued, which has an impact on how well search engines can produce country-specific editions.

[AltaVista Unveils New Search Centers](#)

The Search Engine Report, March 3, 2000

Covers new country-specific guides being rolled out by AltaVista, plus mentions new regional partnerships gained by FAST Search.

[Voila Covers Europe, Reaches To The US](#)

The Search Engine Report, Feb. 3, 2000

Discusses how European-based search engine Voila produces editions for several countries.

[FAST Upgrades Site](#)

The Search Engine Report, Jan. 4, 2000

Brief on FAST Search adding language support.

[Inktomi Launches European Search Center](#)

The Search Engine Report, June 2, 1999

Information about Inktomi's domain crawling index for Europe.

[UKMax: New Inktomi-Powered Search Engine for the UK](#)

The Search Engine Report, January 5, 1999

Information about a new Inktomi-powered service for the United Kingdom.

[AltaVista Northern Europe Looks Gone For Good](#)

The Search Engine Report, November 4, 1998

About the closure of the Northern European mirror site that AltaVista used to operate.

[Northern Light Adds Search Functions, Freshens Index](#)

The Search Engine Report, August 4, 1998

Mentions Northern Light adding a language search option.

[AltaVista Canada Expands](#)

The Search Engine Report, August 4, 1998

Details about the AltaVista Canada site, produced through domain crawling.

[AltaVista Launches New Look, Larger Size, More Languages](#)

The Search Engine Report, June 3, 1998

Discusses how AltaVista stores information from pages in different languages in one index, so that separate services are not necessary.

[Infoseek Launches Advanced Search](#)

The Search Engine Report, March 31, 1998

Information on searching geographically at Infoseek.

[AltaVista's Search By Language Feature](#)

The Search Engine Report, June 4, 1997