



Internet Search Engine Guide

Welcome!

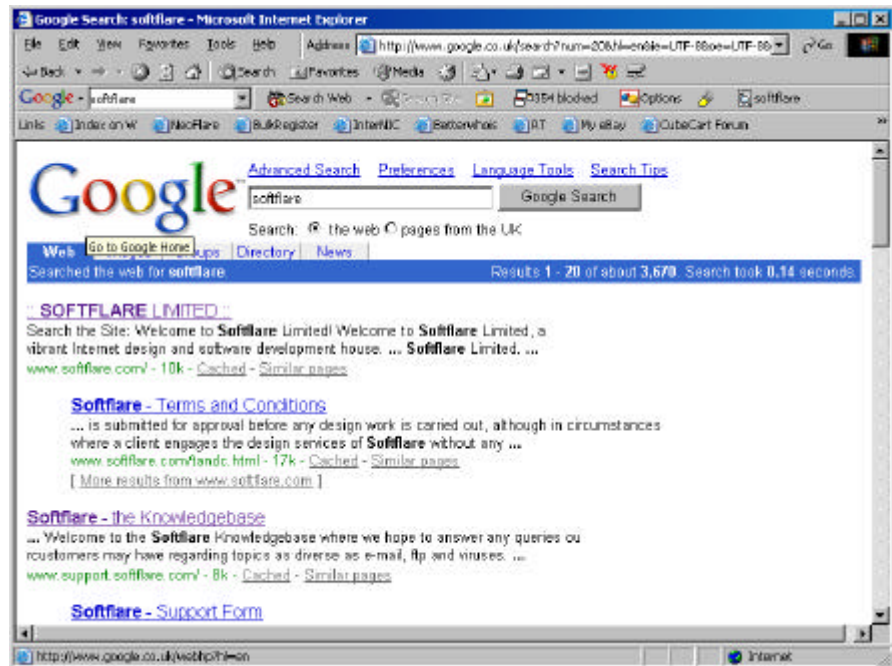
We hope this document will help to explain at least some of the subtle mysteries of how the search engines work. We set the highest standards for our software and services and this guide is intended to give you the information you need to determine whether a professional search engine optimisation programme is going to be cost-effective or whether you could just do it in-house!

Contents

How do Internet Search Engines Work?	3
Location and Frequency	3
The 'Secret' Ingredient	4
Off the page factors	4
Deep crawl	5
Sites that use Frames	5
Improving your Search Engine placement	5
Pick your target Keywords	5
Position your Keywords	5
Have relevant Content	6
Avoid Search Engine stumbling blocks	6
Have HTML links	7
Frames can kill	7
Dynamic doorblocks	7
Build links	7
Say No to search engine spamming	8
Submit your Key pages	8
Verifying your listing	8
Beyond Search Engines	9
Do Gateway / Doorway Pages work?	10
Paid For Listings: Google AdWords	11
Appendix 1 - Using HTML Tags	12
Meta Tag overview	12
The Title Tag	12
The Meta Description Tag	13
The Meta Keywords Tag	13
Meta Robots Tag	14
Other Meta Tags	15
Appendix 2 - Submitting to Google	16
Appendix 3 - Yahoo!	17
Getting Listed	17
Appendix 4 - Verifying your Listing	19
Alta Vista	19
AllTheWeb/FAST Search	19
Google	20
Appendix 5 - Resources	20
Appendix 6 - Softflare Products	21

IMPORTANT!

If increasing the rankings of your website on the Internet's major search engines is a serious goal then pursue it seriously. It will not be achieved in 10 minutes, and be under no illusions - it will require commitment, persistence and consistency.

**How do Internet Search Engines work?**

Note: Most search engines are 'crawler-based' in that they crawl around your website looking for appropriate content: e.g. Google, Ask Jeeves, Lycos. Strictly speaking, Yahoo does not fit into this pattern.

Please see the Appendices for more information specific to Yahoo.

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 a pretty good job.

Unfortunately, search engines don't have the ability to ask the few simple questions that might help focus your search. 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 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

Note: If you find things like "title tag" confusing then please see the Appendices for an illustrated Glossary of Terms.

short.

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 analyse 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.

The 'Secret' Ingredient

Now it's time to qualify the location/frequency method described above. All the major search engines follow it to some degree, they add a little 'extra' 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 webmaster 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.

Deep Crawl

All crawlers will eventually find pages to add to their web page indexes, even if those pages have never been submitted to them. However, some crawlers are better than others. Google and Inktomi do the "deep crawl" and gather many pages from your web site, even if these pages were never submitted, AltaVista doesn't. In general, the larger a search engine's index is, the more likely it will list many pages per site. See the Search Engine Sizes page for the latest index sizes at the major search engines.

Sites that use Frames

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. This subject requires more than a beginner's knowledge of HTML and, possibly, Javascript, and is therefore beyond the scope of this guide.

Improving your Search Engine Placement

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 and HTML Meta Tags in general is explained in the Appendices.

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.

Most websites these days are constructed using a HTML layout method called "tables". Keep in mind that tables can "push" your text further down the page, making keywords less relevant because they appear lower on the physical page. This, without getting into too much technical jargon, is because HTML tables are viewed vertically by search engines and horizontally by humans. There's no easy way around this, other than simplifying your table structure. Consider how tables might affect your page, but don't necessarily stop using them. Tables are what allow web designers to create interesting and navigable websites - the search engine issue on its own is not a reason to dispense with them.

Large sections of JavaScript (for example, embedded scripts that create the rollover effects we see so often) 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. Placing scripts further down on the page, if possible, will help alleviate this problem.

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 as 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.

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 or frames. You need to anticipate these problems, or a search engine may not index any or all of your 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

As already mentioned, some 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.

Dynamic Doorblocks

Is your site generating pages via PHP, 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.

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 centre 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 six months for your "non-submitted" pages to appear in a search engine, and some search engines may not list every page from your site.

Verifying your Listing

Check on your pages and ensure they get listed, in the ways described in the Appendices. Once your pages are listed in a search engine, monitor your listing every week or two. Strange things happen. Pages disappear from catalogues. 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, 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.

Do Gateway / Doorway Pages work?

Years ago a search engine optimization technique was developed to influence or manipulate the positioning of websites in the search engines. This technique is referred to as the creation of doorway pages, gateways or bridge pages. The concept is to build pages that rank highly for specific keyword sets. This is done by calculating the search engines algorithm and then building a page that ranks highly in that particular search engine based on the calculations.

Machine Generated Pages

There are a variety of programs that are marketed that will automatically create these doorways, gateways or bridge pages for you (e.g. <http://www.gatewaygenerator.com/>). You enter specific information into predefined form fields relative to the page that you want to optimise, and the program generates a page that is optimised for the selected search engine and keyword sets. These pages are referred to as machine generated pages.

Unfortunately, this technique has been abused over the years by many consumers and search engine consultants alike. We've seen websites that had 5 or 6 pages of core content and then hundreds of gateway pages that were developed to target just about every keyword set the site was optimised for. Typically, only a few of those pages may have ranked well, the rest were ineffective. Today, this technique is frowned upon by the major search engines and directories. When you see numbers pertaining to billions of web pages being indexed, you can be assured that a large percentage of those are gateway, doorway or bridge pages. The search engine spiders or robots are more intelligent now than they were a few years ago.

These machine generated pages are being dropped from the search engine indexes on a regular basis, or they are buried so far down in the SERP's (Search Engine Results Pages), that they would never be found. At times, the entire domain that the pages reside on may receive a ranking penalty and/or be banned from the search engines index. Do you want to take that chance?

How do search engines know about these pages?

If the pages are machine generated and untouched after generation, there is a specific footprint that the page has (the structure of that web page). When you look at the footprint of thousands of these pages, they are easy for search engine spiders to detect. Most doorway, gateway or bridge pages are orphaned. This means that they have no links pointing to them, but they have links pointing from them. The links pointing from them are usually aimed at your home page or a specific page on your website that the keyword set is targeting.

Although this technique still works today with some search engines, most, if not all of the major search engines and directories have published guidelines that state they do not want doorways, gateways or bridge pages in their indexes. Therefore, although we would not discount out of hand the use of gateway pages, there are more refined and strategical ways to improve the visibility of your website. And of course you improve your chance of rising up the rankings if the gateway pages you produce are not, or are not perceived to be, machine generated. Therefore, rotate them from time to time, keep them current and they, like the other procedures listed in this document can contribute to increasing the visibility of your website.

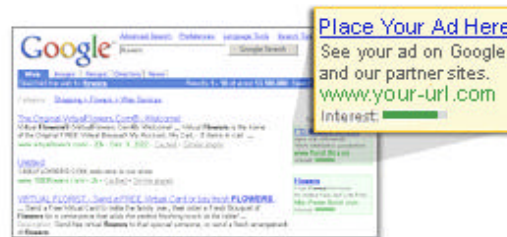
Paid For Listings: Google AdWords

Every major search engine with significant traffic accepts paid listings. This unique form of search engine advertising means that you can be guaranteed to appear in the top results for the terms you are interested in within a day or less. Given this, paid listings are an option that should be explored by site owners who wish to quickly build visibility. They may also be a long-term advertising option for some. We only explore Google AdWords here, but other Search Engines like Overture (GoTo), LookSmart and FindWhat have similar offerings.



Google AdWords

Google sells paid listings that appear above its editorial results called "Premium Sponsorships" and those that appear to the side of its regular results called "AdWords." Since it may take time for a new site to appear within Google, these advertising opportunities offer a fast way to get listed with the service. Also, as with Overture, they may be a continuing option you may wish to explore.



Google AdWords ads appear on Google and partner sites.

Google's self-service AdWords program charges a per click fee, similar to Overture. AdWords charges a \$5 activation fee. In either case, \$25 ought to last you at least one month, if you've carefully selected your terms. You can sign-up for Google's AdWords program via the URL below:

<https://adwords.google.com/select>

Google's Premium Sponsorships are for larger advertisers willing to pay for being listed on a CPM or cost-per-impression basis. They also have a \$10,000 minimum spend over a 3-month period. You can learn more here:

<http://www.google.com/ads/overview.html>

Google also distributes its ads to other partners, with some major sites listed on the Search Results Chart. Usually, it is only the AdWords ads that are distributed.

Appendix 1 - Using HTML Meta Tags

This Appendix is of a relatively technical nature and is intended for everybody with an interest in the HTML of your website, but especially its webmaster.

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. Err... 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.

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 will appear in the reverse bar of someone's browser when they view the web page.



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.

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. 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.

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 (all provide some support, but AltaVista, AllTheWeb and Teoma make most use of the tag).

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).

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.

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 (only Inktomi and Teoma support it to a limited extent).

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 realise 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, 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. As you can see - it's never as black and white as it might at first appear!

Far too many people new to search engine optimisation obsess with the meta keywords tag. Few crawlers support it. For those that do, it might 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, we strongly suggest that those new to search engine optimisation not even worry about the tag at all.

Even those who are experienced in search engine optimisation may decide it is no longer worth using the tags. Well, perhaps not - 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.

Still want to use the meta keywords tag? Inktomi says that you should include up to 25 words or phrases, with each word or phrase separated by commas. In the past, when the tag was supported by other search engines, they generally indexed up to 1,000 characters of text and commas were not required.

Meta Robots Tag

One other meta tag worth mentioning is the robots tag. This lets you specify that a particular page, for whatever reason, 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:

```
<HEAD>
<TITLE>Don't Index this page.</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. Every Softflare hosting account includes a robots.txt in the web root by default. If you use do a robots.txt file to block indexing, there is no need to also use meta robots tags. This is the syntax for the robots.txt file:

```
User-agent: *
Disallow: /cgi-bin/
Disallow: /tmp/
Disallow: /~joe/
```

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 some web pages, you might find "author," "channel" and "date" meta tags. These mean nothing to web-wide crawlers such as Google.

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 specialised search engines.

Appendix 2 - Submitting to Google



Arguably the most important crawler-based search engine is Google, as several major search engines such as Yahoo and AOL Search have their main results "powered" by Google. In addition, many people also go directly to Google to search.

The absolutely best way to get listed with Google is to build links to your web site. Indeed, this is the best way to get listed for free with all the major crawlers listed on this page. Crawlers follow links, so if you have good links pointing at your web site, the crawlers are more likely to find and include your pages.

If you submitted your site to the major directories and got listed with one or more of them, then Google and other crawlers will almost certainly pick up the URL that was listed. This means you may not need to do additional work to get listed with crawlers. Nevertheless, you may find it helpful to do some link building beyond the directories.

Aside from link building, Google provides an **Add URL** page that lets you submit a URL directly to its crawler. There's no guarantee that Google will actually include a URL submitted to it this way, however. Despite this, it makes sense to submit your home page and perhaps one or two other URLs from "inside" your web site via the **Add URL** page.

<http://www.google.com/addurl.html>

You really don't need to submit more than this. The only reason for submitting some of your inside pages is in case there is a problem reaching your home page. This gives Google an alternate route into your site. From whatever page it visits, it will look for links to other pages that you have and perhaps include those. This is true for other crawlers, as well.

If you have a brand new web site, it will probably take about a month before Google lists your web pages. Because of this, you might consider making use of its paid placement program.

Appendix 3 - Yahoo!



Launched in 1994, Yahoo is the web's oldest "directory," a place where human editors organise web sites into categories. However, in October 2002, Yahoo made a giant shift to using Google's crawler-based listings for its main results.

If Yahoo is now powered by Google, then why bother using it? For one thing, you might find that the way Yahoo "enhances" Google's listings with information from its own directory may make search results more readable. In addition, Yahoo's search results pages still show **Categories** links. When offered, these will take you to a list of web sites that have been reviewed and approved by a human editor.

It's also possible to do a pure search of just the human-compiled **Yahoo Directory**, which is how the old or "classic" Yahoo used to work. To do this, search from the Yahoo Directory home page, as opposed to the regular Yahoo.com home page. Then you'll get both directory category links ("Related Directory Categories") and "Directory Results," which are the top web site matches drawn from all categories of the Yahoo Directory.

Sites pay a fee to be included in the Yahoo Directory's commercial listings, though they must meet editor approval before being accepted. Non-commercial content is accepted for free.

Consider Yahoo any time you think you might be well served by having a list of human-reviewed web sites. It's also a good choice for popular queries, since the category listings it provides may help you narrow in and refine your query. Doing a pure Yahoo Directory search also provides an unique human view of the web.

The company completed its purchase of Inktomi in March 2003, a crawler-based rival to Google.

Getting Listed

Yahoo has two submission options: "Standard," which is free, and "Yahoo Express," which involves a submission fee. Look for **Suggest your site** on the main Yahoo home page (yes, they don't make it easy, do they?).

Anyone can use Standard submission to submit for free to a non-commercial category. You'll know the category is non-commercial if the "reverse bar" containing the category name at the top of the category page is blue. You'll also know because if you try to submit to a non-commercial category, the Standard submission option will be offered in addition to the Yahoo Express paid option, discussed further below.

Why might you choose to pay when the free search engine submission option is available? Simply for a fast turnaround time. If you use the free submit choice, there's no guarantee that your submission will be reviewed quickly or at all.

Your submission to a non-commercial category is more likely to be accepted if your content is not overtly commercial. For example, submitting the home page of a site that sells running shoes is likely to be seen as commercial and not accepted. However, if you have a page within that web site that discusses in depth how to select the right type of shoes for different running races,

then that page might be deemed helpful, non-commercial information and accepted.

As for commercial categories, Yahoo requires that sites pay a Yahoo Express submission fee of, at the time of writing, \$299 per year. This fee doesn't guarantee that you will be listed, only that you'll get a yes or no answer about being accepted within seven business days. However, the vast majority of most decent sites are accepted.

If accepted, you'll be reevaluated after a year and charged the submission fee again, if you want to stay in Yahoo's commercial area. You should review the traffic you received from Yahoo over the past year*, to decide if it is worth paying the fee again. If not, you can decline to be listed, and you will not be charged.

* for information on how to view your server log files please visit this page:

<http://support.softflare.com/stats.html>

Appendix 4 - Verifying your Listing

Some crawler-based search engines make it easy to confirm whether 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.

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://<www.yourdomain.com>/thispage.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://<www.yourdomain.com>/services/

would list all the pages from within the /services/ area of your website. 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:<www.yourdomain.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.softflare.com

found 17 results

host:softflare.com

found 46!

In fact, this is the more common way that people come to Softflare's website. Consequently, AltaVista has actually indexed more pages from the site without the www prefix.

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:<www.yourdomain.com>/services/thispage.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:<www.yourdomain.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:<www.yourdomain.com>/services/thispage.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.

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:<www.yourdomain.com> softflare

or

softflare site:<www.yourdomain.com>

would bring up all (or nearly all) of the pages Google lists from Softflare's website, because all the pages should have the word "Softflare" on them.

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

Appendix 5 - Resources

The information found in this document is compiled from a variety of freely available but independent, expert sources on the Internet including but not limited to:

<http://www.unmelted.com>

<http://www.search-engine-secrets.net>

<http://www.searchengineshowdown.com>

<http://www.searchenginewatch.com>

<http://gorilla.johnrecommends.com>

<http://www.searchsecrets.com>

<http://www.optimisation-nation.com>

Appendix 6 - Softflare Products

As well as designing eye-catching websites with or without built-in bespoke software solutions, Softflare Limited have developed a portfolio of browser-based products to enhance communication with your customers and increase your productivity. For more information about these or any other products/services please visit our website (www.softflare.com) or give a member of our Customer Support Team a call (+44 (0)870 770 8000):

FaxFlare™

A fax server that allows you to send faxes from your desktop and delivers incoming faxes to your e-mail address! Totally configurable within a browser.

FireFlare™

Softflare's Unix server-based corporate firewall that allows your users to get out, but stops undesirables getting in!

FotoFlare™

Get a Product/Photo Gallery on your website that you can manage straight from your browser without any technical knowledge at all!

ListFlare™

Keep in touch with all your customers using a simple yet powerful Mailing List Manager - all handled via your web browser!

NetMonitor™

The Network Monitor can monitor routers, computers, modem banks - essentially any item that has an IP address. It will monitor CPU usage, disk space, connectivity, system messages and services such as DNS, FTP, HTTP, IMAP, POP3 and SMTP, to name but a few. There are even scripts to monitor your favorite stock price or your local weather!

ShopFlare™

ShopFlare is an out-of-box e-commerce solution that will allow you, the shopkeeper, to start selling your products to an audience of 100's of millions within minutes! ShopFlare gives you all the tools you need - all you need is a few minutes a day to check your orders and the enthusiasm to exploit the potential of the Internet!

SMSFlare™

SMSFlare is an SMS gateway - a bridge between the text-messaging you need to do and the familiar computer you can send it from. Never use that tiny mobile keypad again!

StorFlare™

Why risk losing your e-mail and critical data with a potential server hard disk failure when StorFlare can synchronise with another remote server and keep all your data backed up?

VirusFlare™

Have your mail scanned for viruses before it gets to your desktop using antivirus software that is updated every 10 minutes!

VPNFlare™

A private network that uses some public segments is called a VPN (Virtual Private Network) and is significantly cheaper and more flexible than a dedicated private network. Each of the secure networks need only be internet-enabled, and adding new connections is simple and inexpensive. VPNFlare - flexible yet secure, encrypted, remote access to business applications.

QMail™

We have been using and recommending Qmail™ as an alternative to Sendmail as a MTA (Mail Transfer Agent) for several years. We offer installation and ongoing support for Qmail™ by phone, email, and onsite. Softflare is an accredited international mirror site for QMail™.