

A Guide to Search Engines



[Advanced Search](#)
[Language Tools](#)

Google Search

I'm Feeling Lucky

[Information, resources, and ways you can help](#) survivors of the Haiti earthquake.

[Advertising Programs](#) - [Business Solutions](#) - [About Google](#)

©2010 - [Privacy](#)

[Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

Search

YAHOO!®

© 2010 Yahoo! | [Page Tour](#) | [Privacy / Legal](#) | [About Our Ads](#) | [Submit Your Site](#)

bing™



Table of Contents

Understanding Search Engines	3
Classification of Search Engines	4
Crawler-Based Search Engines.....	7
Human-Powered Search Engines.....	15
Pay-for-Performance Search Engines	19
Who Feeds Who - Search Engine Relationships.....	23
How Search Engines Rank Pages	25
Suggested Reading.....	28

Understanding Search Engines

This is an introduction into search engines and the relationships between them. This information is valuable in helping SEO's understand how site rankings in one search engine / directory can influence rankings in other engines.

Throughout this paper, (and commonly used on the Internet as well) the term "*SE*" is used as an abbreviation for "*search engine(s)*".

Search engines are the most popular method for target customers to find you. As such, SEs are the most vital avenues for letting customers find you.

Currently, search engines around the world together receive around 400,000,000 searches per day. The searches are done with the help of keywords: as a rule, people type a short phrase consisting of two to five keywords to find what they are looking for. It may be information, products, or services.

(For more information on the share of world searches received by major search engines per day, see <http://searchenginewatch.com/reports/article.php/2156431>).

In response to this query, a search engine will pick from its huge database of Web pages those results it considers relevant for the Web surfer's terms, and display the list of these results to the surfer. The list may be very long and include several million results (remember that nowadays the number of pages on the Web reaches 2.1 trillion, i.e. 2,100,000,000,000); so the results are displayed in order of their relevancy and broken into many pages (most commonly 10 results per page). Most Web surfers rarely go further than the third page of results, unless they are considerably interested in a wide range of materials (e.g. for a scientific research). One reason for this is that they commonly find what they look for on those first pages without the needing to dive in any deeper.

That's why a position among the first 30 results (or "*top-30 listing*") is a coveted goal. There used to be a great variety of search engines, but now after major reshuffles and partnerships there are just several giant search monopolies that are most popular among Web surfers and which need to be targeted by optimizers.

There are – and the search engines are aware of this – more popular searches and less popular searches. For instance, a search on the word "*myrmecology*" is conducted on the Web much more rarely than a search for "*Web hosting*". Search engines make money by offering special high positions (most often called "*sponsored results*") for popular terms, ensuring that a site will appear to Web surfers when they search for this term, and that it will have the best visibility. The more popular the term, the more you will have to pay for such a listing.

Classification of Search Engines

The term "*search engine*" (SE) is often misused to describe both directories and pure search engines. In fact, they are not the same; the difference lies in how result listings are generated.

There are four major search engine types you should know about. They are:

- crawler-based (traditional, common) search engines;
- directories (mostly human-edited catalogs);
- hybrid engines (META engines and those using other engines' results);
- pay-per-performance and paid inclusion engines.

Crawler-based SEs, also referred to as spiders or Web crawlers, use special software to automatically and regularly visit websites to create and supplement their giant Web page repositories.

This software is referred to as a "bot", "robot", "spider", or "crawler". All these terms denote the same concept. These programs run on the search engines. They browse pages that already exist in their repositories, and find your site by following links from those pages. Alternatively, after you have submitted pages to a search engine, these pages are queued for scanning by a spider; it finds your page by looking through the lists of pages pending review in this queue.

After a spider has found a page to scan, it retrieves this page via HTTP (like any ordinary Web surfer who types an URL into a browser's address field and presses "enter"). Just like any human visitor, the crawling software leaves a record on your server about its visit. Therefore, it's possible to know from your server log when a search engine has dropped in on your online estate.

Your Web server returns the HTML source code of your page to the spider. The spider then reads it (this process is referred to as "*crawling*" or "*spidering*") and this is where the difference begins between a human visitor and crawling software.

While a human visitor can appreciate the quality graphics and impressive Flash animation you've loaded onto your page, a spider won't. A human visitor does not normally read the META tags, a spider can. Only seasoned users might be curious enough to read the code of the page when seeking additional information about the Web page. A human visitor will first notice the largest and most attractive text on the page. A spider, on the other hand, will give more value to text that's closest to the beginning and end of the page, and the text wrapped in links.

Perhaps you've spent a fortune creating a killer website designed to immediately captivate your visitors and gain their admiration. You've even embedded lots of quality Flash animation and JavaScript tricks. Yet, a search engine spider is a robot which only sees that there are some images on the page and some code embedded into the "<script>" tag that it is instructed to skip. These design elements are additional obstacles on its way to your content. What's the result? The spider ranks your page low, no one finds it on the search engine, and no one is able to appreciate the design.

SEO (search engine optimization) is the solution for making your page more search engine friendly. The optimization is mostly oriented towards crawler-based engines, which are the most popular on the Internet. We're not telling you to avoid design innovations; instead, we will teach you how to properly combine them with your optimization needs.

Let's return to the way a spider works. After it reads your pages, it will compress them in a way that is convenient to store in a giant repository of Web pages called a search engine index. The data are stored in the search engine index the way that makes it possible to quickly determine whether this page is relevant to a particular query and to pull it out for inclusion in the result page shown in response to the query. The process of placing your page in the index is referred to as "*indexing*". After your page has been indexed, it will appear on search engine results pages for the words and phrases most common on the indexed Web page. Its position in the list, however, may vary.

Later, when someone searches the engine for particular terms, your page will be pulled out of the index and included in the search results. The search engine now applies a sophisticated technique to determine how relevant your page is to these terms. It considers many on-page and off-page factors and the page is given a certain position, or rank, within other results found for the surfer's query.

This process is called "*ranking*". Google (www.google.com) is a perfect example of a crawler-based SE.

Human-edited directories are different. The pages that are stored in their repository are added solely through manual submission. The directories, for the most part, require manual submission and use certain mechanisms (particularly, CAPTCHA images) to prevent pages from being submitted automatically. After completing the submission procedure, your URL will be queued for review by an editor, who is, luckily, a human.

When directory editors visit and read your site, the only decision they make is to accept or reject the page. Most directories do not have their own ranking mechanism - they use various obvious factors to sort URLs, such as alphabetic sequence or Google PageRank™ (explained later in this course). It is very important to submit a relevant and precise description to the directory editor, as well as take other parts of this manual submission seriously.

Spider-based engines often use directories as a source of new pages to crawl. As a result, it's self-evident in SEO that you should treat directory submission and directory listings as seriously and responsibly as possible.

While a crawler-based engine would visit your site regularly after it has first indexed it, and detect any change you make to your pages, it's not the same with directories. In a directory, result listings are influenced by humans. Either you enter a short description of your website, or the editors will. When searching, only these descriptions are scanned for matches, so website changes do not affect the result listing at all.

As directories are usually created by experienced editors, they generally produce better (at least better filtered) results. The best-known and most important directories are Yahoo (www.yahoo.com) and DMOZ (www.dmoz.org).

Hybrid engines: Some engines also have an integrated directory linking to them. They contain websites which have already been discussed or evaluated. When sending a search query to a hybrid engine, the sites already evaluated are usually not scanned for matches; the user has to explicitly select them. Whether a site is added to an engine's directory generally depends on a mixture of luck and content quality. Sometimes you may "apply" for a discussion of your website, but there's no guarantee that it will be done.

Yahoo (www.yahoo.com) and Google (www.google.com), although mentioned here as examples of a directory and crawler respectively, are in fact hybrid engines, as are nowadays most major search machines. As a rule, a hybrid search engine will favor one type of listing over another. For example, Yahoo is more likely to present human powered listings, while Google prefers its crawled listings.

Meta Search Engines: Another approach to searching the vast Internet is the use of a multi-engine search, or meta-search engine that combines results from a number of search engines at the same time and lays them out in a formatted result page. A common or natural language request is translated to multiple search engines, each directed to find the information the searcher requested. The search engine's responses thus obtained are gathered into a single result list. This search type allows the user to cover a great deal of material in a very efficient way, retaining some tolerance for imprecise search questions or keywords.

Examples of multi-engines are MetaCrawler (<http://www.metacrawler.com>) and DogPile (<http://www.dogpile.com>). MetaCrawler refers your search to seven of the most popular search engines (including AltaVista and Lycos), then compiles and ranks the results for you.

Pay-for-performance and paid inclusion engines: As is clear from the title, with these engines you have no way other than to pay a recurring or one-time fee to keep your site either listed, re-spidered, or top-ranked for keywords of your choice. There are very few search engines that solely focus on paid listings. However, most major search engines offer a paid listing option as a part of their indexing and ranking system.

Unlike paid inclusion where you just pay to be included in search results, in an advertising program listings are guaranteed to appear in response to particular search terms, and the higher your bid, the higher your position will be for these terms. Paid placement listings can be purchased from a portal or a search network. Search networks are often set up in an auction environment where keywords and phrases are associated with a cost-per-click (CPC) fee. Such a scheme is referred to as Pay-Per-Click (PPC).

Yahoo and Google are the largest paid listing providers, and Live Search (formerly MSN) also sells paid placement listings.

Crawler-Based Search Engines

In the previous section we discussed how crawler-based engines work. Typically, special crawler software visits your site and reads the source code of your pages. This process is called "crawling" or "spidering". Your page is then compressed and put into the search engine's repository which is called an "index". This stage is referred to as "indexing". Finally, when someone submits a query to the search engine, it pulls your page out of the index and gives it a rank among the other results it has found for this query. This is called "ranking".

Usually for indexing, crawler-based engines consider many more factors than those they can find on your pages. Thus, before putting your page into an index, a crawler will look at how many other pages in the index are linking to yours, the text used in links that point to you, what the PageRank is of linking pages, whether the page is present in directories under related categories, etc. These "*off-page*" factors are a significant consideration when a page is evaluated by a crawler-based engine. While theoretically, you can artificially increase your page relevance for certain keywords by adjusting the corresponding areas of your HTML code, you have much less control over other pages in the Internet that are linking to you. Thus, off-page relevance prevails in the eyes of a crawler.

In this section, we look at the main spider-based search engines, and learn how we can get each of them to index our site and rank it highly. Although this step does not closely deal with the optimization process itself, we provide information on how each search engine looks at your pages so that you can come back to this section for later reference.

Google (<http://www.google.com>)

Google is the number one search engine among such giants of the SEs' market as Yahoo! and Bing/Live Search/MSN. Its search share is over 60%. Google indexes billions of Web pages, so that users can search for the information they desire. Plus it creates services and tools including Web applications, advertising networks and solutions for businesses to hold on the leading position successfully.

You can submit your site to Google here: <http://www.google.com/addurl/> and your site will probably be indexed in around 1-2 months.

Alternatively, you can sign in to your Google account, go to Google Webmaster Tools and submit a sitemap of your site.

Please keep in mind that Google may ignore your submission request for a significant period of time. Even if it happens to crawl your site, it may not actually index it if there are no links pointing to it. However, if Google finds your site by following the links from other pages that have already been indexed and are regularly re-spidered, chances are you will be included without any submission. These chances are much higher if Google finds your site by reading a directory listing, such as DMOZ (www.dmoz.org).

So submit your site and it may help, but links are the best way to get indexed.

In the past, Google typically performed monthly updates called the "Google Dance" among the experts. At the beginning of the month, a deep crawl of the web took place, then after a couple of weeks the PageRank for the retrieved pages was calculated, and at the end of the month the index database was finally updated. Nowadays, Google has switched to an incremental daily update model (sometimes referred to as *everflux*) so the concept of Google dance is quickly becoming historical.

The "*Dance*" took place from time to time but only when they need to make major changes to their algorithm. For example, their dance in November 2003 (known as Google Florida Update) was actually their first after about six months. In January 2004, Google started another dance (Austin Update) where pages that had disappeared during the "*Florida*" showed up again, and many pages that hadn't disappeared the first time were gone.

In February 2004 Google updated once more and things settled down. Most people who had lost pages saw them return and although the results were rather different than those shown before Florida, at least pages didn't seem to disappear for no reason.

Google claims to have 1 trillion (as in 1,000,000,000,000) unique URLs in its index. The engine constantly adds new pages to the index database - usually it takes around two days to list a new page after the Googlebot (Google's spider) has crawled it. The Google team works industriously towards algorithm perfection to keep their leading position amongst search engines.

These days, Google maintains a database which is continuously updated. Matt Cutts (head of Google's Webspam team) reported in his personal blog that: *«Google switched to an index that was incrementally updated every day (or faster). Instead of a monolithic monthly event, the Google would refresh some of its index pretty much every day, which generated much smaller day-to-day changes that some people called everflux.»*

Google has lots of so-called "*regional*" branches, such as "Google Australia", "Google Canada," etc. These branches are modifications of their index database stored on servers located in the corresponding regions. They are meant to further adjust search results to searcher needs: when you're searching, Google detects your IP address (and thus approximate location) and feeds the results from the most appropriate index database. Submission to the "*Main Google*" will list your site in all its regional branches – after Google indexes you, of course.

Google has a number of crawlers to do the spidering. They all have the name "GoogleBot" but they come from a number of different IP addresses. You can see if Google has visited your site by looking through your server logs: just find the IP address something like 82.110.xxx.xx and most probably you will see the user-agent defined as GoogleBot "Googlebot/2.1+(+http://www.google.com/bot.html)". Verify that IP address using the reverse DNS lookup to find out the registered name of the machine. Generally all Google crawlers host names will end with 'googlebot.com'

Google is by far the most important search engine. Apart from their own site receiving 350 million searches per day, they also provide the search results for AOL Search, Netscape Search,

Ask.com, Iwon, ICQ Search and MySpace Search. For this reason, most optimizers first focus on Google. Generally, this makes sense.

How to optimize for Google

Most important for Google are three factors: PageRank, link anchor text and semantics. PageRank is an absolute value which is regularly calculated by Google for each page it has in its index. Later in this course we will give you a detailed description, but for now it's just important to know that the number of links you've got from other sites outside your domain matters greatly, as well as the link quality. The latter means that in order to give you some weight, the sites linking to yours must themselves have high PageRank, be content-rich and regularly updated.

MiniRank/Local Rank is a modification of the PageRank based on the link structure of your single site only. Since search engines rank pages, not sites, certain pages of your site will rank higher for given keywords than others. Local Rank has a significant influence on the general PageRank.

Anchor text is the text of the links that point to your pages. For instance, if someone links to you with the words "see this great website", this is a useless link. However, let's say you sell car tires and a link from another site to yours says "car tires from leading brands", such a link will boost your rank when someone searches for car tires on Google.

Semantics is a new factor that appears to have made the biggest difference to the results. This term refers to the meaning of words and their relationships. Google bought a company called Applied Semantics back in 2003 and has been using the technology for their AdSense contextual advertising program. According to the principles of applied semantics, the crawler attempts to define which words mean the same thing and which ones are always used together.

For example, if there are a certain number of pages in Google's index saying that an executive desk is a piece of office furniture, Google associates the two phrases. After this, a page about executive desks using the keywords "office furniture" won't show up in a search for the keywords "executive desk". On the other hand, a page that mentions "executive desk" will rank better if it mentions "office furniture".

Now, there are two other terms related to Google's way of ranking pages: Hilltop and Sandbox. Hilltop is an algorithm that was created in 1999. Basically, it looks at the relationship between "*Expert*" and "*Authority*" pages. An "*Expert*" is a page that links to lots of other relevant documents. An "*Authority*" is a page that has links pointing to it from the "*Expert*" pages.

In theory, Google would find "*Expert*" pages and then the pages that they link to would rank well. Pages on sites like Yahoo, DMOZ, college sites and library sites would be considered experts.

Sandbox refers to Google's algorithm which detects how old your page is and how long ago it has been updated. Usually pages with stale content tend to gradually slip down the result list, while new pages just crawled initially have higher positions than they would if based on

PageRank only. However, some time after gaining boosted positions, new website disappear from the top places in search results, since Google wants to verify whether your website is really continued and was not created with the sole purpose to benefit from artificially high rankings over the short term. The period when a website is unable to make it to the top of search results is referred to as "being in the sandbox". This can last from 6 months to one year, then the positions usually restore gradually. However, not all brand new site owners observe the sandbox effect on their own sites, which has led to discussions on whether the sandbox filter really exists.

On-page factors considered by Google

Now that we've examined off-page factors that have primary importance for Google, let's take a look at on-page factors that should be given attention before submitting to Google. Google does not consider the META keyword tag for counting relevancy. While your META description tag contents can be used by Google as the description of your site in the search results, the META description does not have any influence for relevancy count.

Nowadays META tags don't influence website position in search results absolutely. They can be of use as additional information source about the Web page for surfers only.

When targeting optimization for Google, be sure to use your keywords in the following:

- Your domain name - important!
- First words of the TITLE tag; HTML heading tags H1 to H6
- ALT text as long as you also describe the image;
- Quality content on your index page. Try to make the length of your home page at least 300 words, however, don't hide anything from visitors' eyes (VERY IMPORTANT!).
- Link text for outgoing links.
- Drop-down form boxes created with the help of the SELECT tag.
- Finally, try to have some keywords in BOLD.

Additionally, try to center your pages on one central theme. Use synonyms of your important keyword phrases. Keep everything on the page on that ONE main topic, and provide good, solid content.

Pages that are optimized for Google will score best when there are at least a few links to outside sites that are related to your topic because this establishes your page's reputation as an authority. Google also measures how many websites outside your domain have links pointing to your site and factor in an "importance rating" for each of those referring sites. The more popular a site appears to a search engine, the higher up in the search listings they will place it.

According to Craig Silverstein with Google, "External links that you grant from a particular page on your website can become diluted. In other words, if you place 10,000 links to other Web pages from a particular page of your website, each link is less powerful than if you were to link to only five other Web pages. Or, the contribution value to another website of each individual link is weakened the more you grant."

Yahoo! (<http://www.yahoo.com>)

You can submit your URL to Yahoo! Search for free here:

<http://submit.search.yahoo.com/free/request> (note: you need to register at their portal first) and it will be indexed in about 1-2 months.

Yahoo! is still the most popular site on the Web, according to its traffic rank reported by Alexa (www.alexa.com). Nevertheless, in terms of the number of searches performed Google carries the day.

Yahoo! provides results in a number of ways. First, it has one of the most complete directories on the Web. There's also Yahoo! Search, which lists results in a way similar to other crawler-based engines. Here, in this section on crawler-based engines, we deal with the second service.

Sponsored results are found at the top, side, and bottom of the search results pages fed by Yahoo! Yahoo! now owns its Yahoo! Search Marketing pay engine and provides search results to AltaVista, AllTheWeb and Lycos.

The search results at Yahoo! changed in February 2004. For the previous couple of years, Google was their search-results supplier. Nowadays, Yahoo! is using its own database. Yahoo! bought engines that had earlier pioneered the search world, Yahoo! Search officially provides all search engines acquired through these acquisitions with its own results. Therefore, when you optimize your Web pages for Yahoo!, there's a good chance of appearing in the top results of other popular search engines, such as AllTheWeb and AltaVista.

To find out if Yahoo's spider has visited your site, search the following information in your server logs. Their crawler is now called Yahoo Slurp (formerly, it was just Slurp).

For each request from 'Yahoo! Slurp' user-agent, you can start with the IP address (i.e. 74.6.67.218). Then check if it really is coming from Yahoo! Search using the reverse DNS lookup. The name of all Yahoo! Search crawlers will end with 'crawl.yahoo.net'.

To rank well in Yahoo, you need to do the same things that help your rankings in Google. Off-page factors (link popularity, anchor text, etc.) are very important. Some experts consider it easier to rank well on Yahoo because your own internal links are more important and there also appears to be no requirement for link relevancy. Whereas Google claims that the PageRank of the relevant linking sites is worth more than the PageRank of irrelevant sites, links don't need to be relevant to do well on Yahoo.

Like all other search engines, they'll list you for free if you get links to your site. Much like Google, their crawler is very active and updates listings on a daily basis. However, it can take a few weeks for Yahoo to list new pages after they have found and crawled through referring links. The pages that have already been included in the listing are updated much more often, usually every several days.

In March 2004, Yahoo launched its paid inclusion program called Site Match. You can find out details about Yahoo's paid inclusion program here:

<http://searchmarketing.yahoo.com/srch/index.php> and the pricing here -

http://searchmarketing.yahoo.com/srchsb/sse_pr.php?mkt=us Site Match guarantees your site will appear in non-sponsored search results on Yahoo!, and other portals such as AltaVista and AllTheWeb.

Let's review some quick insights into the factors on your pages that will help you rank higher in Yahoo: use keywords in your domain name and use keywords in the TITLE tag. Your title must include your most important keyword phrase once toward the beginning of the tag. Don't split the important keyword phrase in your title!

META Keywords and META Description tags.

The Yahoo! family of search engines DOES NOT consider META tags when estimating relevancy. Use keywords in heading tags H1 through H6; Use keywords in link anchor text and ALT attributes of your images; the body main content and page names (URLs) need to have keywords in them too; the recommended keyword weight in the BODY tag for Yahoo is 5%, maximum 8%. A catalog page with lots of links, for instance a site map, will help a lot for your indexing and ranking by Yahoo!.

You can submit just the main page to Yahoo!, and let its spider find, crawl, and index the rest of your pages. If it doesn't find an important page, however, make sure you submit it manually.

Like with all of the other engines, solid and legitimate link popularity is considered important by Yahoo's spider as a ranking factor.

Yahoo! frowns upon having satellite sites that revolve around the theme of a main site. For example, if you sell office furniture and set up a main company site and then plant several satellite sites for each kind of furniture, it may seem suspicious to Yahoo!. Therefore, make sure that each site is a stand-alone site and serves a unique purpose, and that it's valuable to both the search engines and your users.

As with any other search engine it is vitally important for Yahoo! that you create valuable content for your search visitors.

Bing/Live Search/MSN (<http://www.bing.com/>)

You can submit your site to Bing/Live Search (formerly known as MSN Search) for free at <http://www.bing.com/docs/submit.aspx>; however they are sure to find it without your submission if you have links from sites already listed there.

MSN stands for Microsoft Network and was initially meant to be Microsoft's solution for Web search, among other goals. Nevertheless, it was powered by Inktomi's results and did not have its own crawler.

Since February 2005, MSN switched from another engine result base and introduced its own Web crawler. And their next significant step was on September 11, 2006 when **Live Search** release replaced MSN Search.

Another level of evolution was the **Bing** search engine launched in June 2009. Bing pulls content from indexed websites and displays the navigation path and variations on the search query in the left-hand side of the screen. Clicking on the new keyword or key phrase helps the searcher find information quickly.

For this moment Bing is being actively promoted. The first step of promotion was redirecting users from Live Search and MSN sites to Bing.com. Hopefully Bing will go on evolving and grow into an independent search engine and a stable source of visitor traffic.

Before Bing was launched, Live Search was one of the most popular of world-class search engines, with around 9% of all search traffic. It definitely makes sense to target placement at the top of Bing/Live Search/MSN's listings as the amount of traffic you will receive as a result is considerable. However, with Bing it's especially important to avoid spam methods since they claim to use a sophisticated series of technologies to fight against even potential spammers.

The information for finding the Bing spider in your server visit logs is as follows: The Spider's name is MSNBot, with the IP addresses something like 65.55.xx.xx and 65.64.xx.xx, host msnbot.msn.com, and user agent "msnbot/1.1 (+http://search.msn.com/msnbot.htm)".

At this stage, the general rules of optimizing for Bing would be similar to the optimization rules for other search engines. Get your site listed in the directories, obtain solid and quality link popularity, and balance your keyword theme. You may also consider purchasing an ad from Bing/Live Search to get listed in the sponsored results. Bing equally treats both off-page and on-page factors when ranking pages.

FAST Search / AllTheWeb and AltaVista (<http://www.alltheweb.com>)

FAST (now called Fast Search and Transfer) is a Norwegian company that specializes in search and filter technology solutions. Some time ago it built the AllTheWeb search engine. In 2003, AltaVista and AllTheWeb were bought out by the Overture engine, the second being bought from FAST Search. Overture, in its turn, was purchased by Yahoo!. It is through this route that AllTheWeb has joined the Yahoo! family of search engines. As a result, AllTheWeb's spider is no longer crawling the Web, and you can no longer submit to the engine. Its XML feed and paid inclusion programs have been changed over to Yahoo! Search Marketing programs. Submitting to Yahoo! Search (and getting indexed there) will get your pages in AltaVista, Yahoo! Web Results, AllTheWeb, HotBot and other engines.

AltaVista (<http://www.altavista.com>) was once a big player in the search industry. Until about seven years ago they could claim to be the most used search engine. Since that time, this engine has lost its independence and much of its popularity. Nowadays, it's sending very little traffic to websites. As with AllTheWeb, it was purchased by Yahoo together with Overture.

Ask.com (<http://www.ask.com>)

Ask.com (formerly Ask Jeeves) is the last crawler-based search engine we are going to speak about. Ask.com supplies its search results to the Iwon search engine and is a member of the Google family of search engines.

It was designed as **Ask Jeeves**, where "Jeeves" is the name of the "gentleman's personal gentleman", or valet (illustrated by Marcos Sorensen), fetching answers to any question asked. On September 23, 2005 the company announced plans to phase out Jeeves and on February 27, 2006 the character was disassociated with Ask.com (according to the Wikipedia).

Ask.com has acquired DirectHit and Teoma which were big players in the search industry too. DirectHit was a search engine that provided results based on click popularity. Therefore, sites that received the most clicks for a particular keyword were listed at the top. Teoma was also unique due to its link popularity algorithm. Teoma claimed that it produces well ordered and relevant search results by initially eliminating all the irrelevant sites and then considering the popularity of only those that relate to the search subject in the first place.

Optimizing for Google will guarantee your appearance on the Ask.com results along with the other large engines of Google's family (AOL, Netscape and Iwon).

Human-Powered Search Engines

The term "human powered" mainly refers to directories, i.e. online catalogs which categorize websites into thematic sections. Yahoo!, along with the regular Web search, offers one of the most complete catalogs on the Web.

When you submit a site to a directory, it is queued for editorial review. Usually, when you submit, you are allowed to choose the category your site will be placed under, and the desired description and title for your site, which will show up in a related category. However, the actual presence of your site is subject to the editor's decision when he or she browses your site.

Directories DO NOT accept automated submissions and have special methods to protect themselves from auto-submission software. You should always submit manually to a directory.

What the directories are useless for

With the directories, there's no concept of "*ranking*", because once you've been included into the index, your site is ranked within the appropriate category by an independent factor, for instance, alphabetically or according to the Google PageRank. Thus, the concept of "*optimization*" has no bearing when applied to directories like DMOZ or Yahoo.

In terms of traffic, these directories are also of little value. The percentage of Web surfers that would prefer browsing categories in a directory over a regular search is quite small; few people tend to dive into the depth of the branchy vertical hierarchy of categories. Thus, on its own your category listing won't bring much traffic.

What the directories are useful for

Since the directories are compiled by humans and are for humans (unlike search engine listings which are compiled by robots for humans), the relevancy of directory results is very high. Search engines know that – almost all search engine spiders start their regular crawl at a directory like DMOZ or Yahoo. If you are listed there, the search engines will find you during their next crawl, even without your submission.

Additionally, directories themselves have high PageRank values (as a rule). When you are listed in a directory, the directory is linking to your site. This link is considered high quality and is very good for your overall link popularity, which is an influential factor when it comes to ranking. For search engines that are able to analyze the context of your link (such as Google), your directory listing will be important because your link is supplemented by a relevant description.

This paper describes two major search services that mainly function as directories and which are extremely valuable for ranking improvement: Yahoo! and DMOZ.

Yahoo! Directory (<http://dir.yahoo.com>)

To submit pages to the Yahoo! Directory, it will cost \$299 for an annual review (for adult-related sites - \$600). If you pay, it's possible to be listed within a few days. If you go the non-commercial route, it can take months and quite often they won't list you at all. Registration at the Yahoo! portal is required for both cases.

Yahoo is now making use of click through measurements as part of its relevancy ranking system. Searches performed at Yahoo!, instead of leading directly to websites, point to yahoo's internal redirecting script so that Yahoo can measure what people are searching for. Yahoo also tracks and measures pages that are not directly listed in Yahoo but are linked from pages that are listed.

There are some tips that can help you get indexed in the directory faster if you are submitting as a standard (non-commercial) site. First, put a copyright notice and date on your page. Avoid using "trademarks" for your company or product name until after you've submitted and gotten indexed by Yahoo!. It appears that Yahoo! is concerned about posting anything that is trademarked as a company name in the listings. Yahoo! Directory will list additional pages of a site (beyond the one submitted) and/or sub domains if the editor finds the content unique and if there aren't many listings for the same topic in the category.

It appears that Yahoo will reject your domain if it has more than 54 characters. Keeping Yahoo! compliance in mind, stay under 55 characters (not including the .com or other suffix) when choosing your domain name.

Yahoo! regularly performs checks for dead links. If your site is down during one of these checks, you may find it deleted from the Yahoo! Directory.

To sum up all of the above, the following is recommended for the Yahoo! Directory.

First, try to get listed in their directory for free (register at the Yahoo! Portal, go to <http://dir.yahoo.com> and submit your URL under the related category).

When submitting for free, there is no guarantee your site will be reviewed and listed anytime soon. However there are some guidelines that might help. Try to purchase a domain name with your most important keywords in it, preferably separated by hyphens. If these keywords are your company name, it's the best variant as you can submit the same words you use in the domain name as the name of your company. Create unique and original content for your website. Ensure it has no broken links or "*Under construction*" pages. If possible, make your design look as professional as possible because – remember – the site will be reviewed by a human editor. Generally, focus on proving to the directory that your site is all about your keyword phrase.

It will help if your site is more than just one page. It should have at least 6 to 7 pages. Put contact information at the bottom of the main page. When submitting a description, make sure it uses your key phrase. When choosing a subcategory, provide a second subcategory as well. Try to choose a subcategory that is close to the upper-level categories. If your keyword phrase is in the name of a subcategory, that's an added bonus too.

Once you have submitted, wait for a couple of weeks. If your site still isn't in the index, you may resubmit (there's no penalty for resubmission and sometimes it helps), submit to another category or choose paid submission if you don't have much time.

DMOZ / ODP (www.dmoz.org)

DMOZ stands for "*Directory.MOZilla*" and is also known ODP which stands for Open Directory Project. It is a multilingual open content directory of World Wide Web links owned by Time Warner that is constructed and maintained by a community of volunteer editors.

After Google made the decision to remove links to DMOZ results from its result pages, DMOZ lost some of its importance. Listing there, however, can still be extremely profitable for your rankings and general Web visibility.

To submit, go to their main page, then find the most appropriate category for your site and click the "*Suggest URL*". Unlike Yahoo! Directory, here you shouldn't submit to a top-level category. Instead, drill down until you find the best subcategory for your site.

If you submit to an inappropriate category, the editor of that category has to transfer your submission to another category. In order to do that, he/she has to visit the main page and click through to find the "*perfect*" category. This is time consuming, and a lot of editors won't do it. They'll simply reject your submission. So, you should be very careful about choosing the best category for your site. When searching for an appropriate subcategory visit the different engines that utilize ODP results (Google and Google family: AOL Search and Netscape Search; then, AltaVista and finally Lycos). Search for your keyword phrase, and see which categories come up. Give your preference to results from the engine you'd most like to target.

Create a captivating title and description which contain your keywords. Make sure it is professional and sincere. Then complete the submission form.

Submit your main URL in the best subcategory. If you have an interior page that stands on its own and has a lot of relevant information, you can try submitting it into a second subcategory.

A "*last updated on*" note on your site can be informative to the editors, especially if it's been updated very recently. Like the Yahoo! Directory, ensure that your site is error-free and there are no missing images or broken links. Don't submit directly to the ODP if you've submitted your site to the ODP links on any of the other engines.

A smart technique would be to attempt submitting to the regional categories at the ODP as well as career categories. You may be able to get listed in several categories in this manner.

The directory is compiled by volunteers and this may be one of the reasons why sometimes it's quite difficult to get listed. So, once you have submitted your site, wait at least a month. Then, if you aren't listed, go to <http://www.resource-zone.com> and post a question about your listing in the Site Submission Status Forum. If you want to detect whether you've been listed, don't do it

through the search as the index isn't synchronized with the database very frequently. Instead, go to the correct category and look for your listing.

Another way to find out the results of your submission is to contact the category editor personally. When doing this be as polite as possible – these editors are volunteers and they do not like being told what they should and shouldn't do. They will help as long as you show that you appreciate the job that they do.

Unlike Yahoo!, you shouldn't re-submit to DMOZ because it pushes your site to the bottom of the queue and would just extend the waiting period. As with Yahoo!, good content matters a great deal. Generally DMOZ won't list sites that are only set up to earn affiliate income or sell something. Try to make your pages as informative about your subject as possible.

Pay-for-Performance Search Engines

As opposed to organic search results (free by nature), the majority of search engines now offer Pay for Performance (PFP) options. Pay for Performance lets you promote your site by paying for SE exposure, rather than by relying on solely organic listings determined by your SEO efforts.

The picture below demonstrates the difference between organic and paid search results in Google.



There are three main types of Pay for Performance options:

- Pay-per-click - the best examples are Google AdWords and Yahoo! Search Marketing Solutions (formerly Overture). With pay-per-click (PPC) advertising, advertisers place bids for different search keywords. When users perform searches for these keywords, advertiser short, textual ads are shown together with organic results. If a user clicks on one of these ads, the advertiser is charged the per-click sum agreed to earlier.
- Paid inclusion (or paid submission) is a fee-based inclusion into the database or directory of a search engine. The more prominent paid inclusion programs are Yahoo! Search Submit and Yahoo! Directory Submit.

- Paid sponsorship - with this model, an advertiser pays a flat fee to a search engine. In return, the search engine shows the advertiser's ads together with search results for pre-selected keywords. ExactSeek, for example, features this pay-for-performance model.

Pay-per-click

PPC advertising is by far the most widespread form of pay-for-performance search marketing. As it's the most effective way for a search engine to make money, PPC is offered by almost every SE on Earth. However, the two most prominent providers of PPC advertising are Google and Yahoo.

Google AdWords

Google AdWords is the world leader in pay-per-click advertising. Currently it has more than 150,000 advertisers. The ads show up not only with Google search results, but also with Google partners such as AOL search, About.com and thousands of other websites that publish AdWords ads. Google has an interesting ad ranking system. It ranks ads not by the bid (the amount their owners are ready to pay for one click), but by the combination of the bid and the click-through ratio. This way, Google maximizes its revenue stream (since Revenue to Google = Bid x CTR x Views) and gives small advertisers an opportunity to effectively compete with big companies. A small advertiser cannot compete on the cost-per-click basis, but can successfully overcome any big company in terms of the click-through ratio.

AdWords ads can only contain 95 characters: 25 for the headline, then two 35-characterlong description lines, and a visible URL field.

AdWords gives advertisers several options to target keywords: broad matching, exact matching, phrase matching, and negative keywords. The matching options define how close the search string entered by a user should be to a keyword selected by an advertiser. If the advertiser has chosen [*tennis ball*] as their keyword (square brackets mean exact match), their ad will be shown only if a user enters *tennis ball* into the search box. If the advertiser has chosen "tennis ball" (quotes mean phrase match), the ad shows up if a user searches for *red tennis ball* or *yellow tennis ball* or simply *tennis ball*. Finally, if the advertiser has chosen *tennis ball* with no brackets or quotes around it (for a broad match), the ad will show up even if a user enters *wilson rackets*. With negative keywords, advertisers can prevent their ad from showing up if a user enters this keyword. For example, a retailer would usually add *-free*, *-replica* to the keywords list to avoid targeting "free stuff" hunters.

One recent development with AdWords was the release of AdWords API (application program interface) that will allow third-party developers to create applications that will work directly with AdWords accounts facilitating and automating many bid and ad management tasks.

You can learn more and sign up for Google AdWords at <http://adwords.google.com>

Yahoo

Yahoo! Search Marketing Solutions <http://searchmarketing.yahoo.com> (formerly Overture) is the second top player after Google in the field of PPC advertising. Overture (originally called GoTo) was the first company to offer PPC advertising and was later acquired by Yahoo. Yahoo ranks ad listings exclusively based on the bid amount; furthermore, if you get a top-3 listing with Yahoo, your ad will be prominently placed with MSN, Altavista, CNN, and other search and news portals.

Yahoo Search Marketing offers Sponsored Search and Content Match. Sponsored search ads show up among search results in Yahoo and its partners; Content Match ads appear on the content partners of Yahoo (this is not search, but conventional websites that are capitalizing on their content by showing third-party ads from Yahoo or Google).

Other PPC Engines

A PPC model is an excellent source of income for any search engine. Therefore, almost every search engine offers PPC advertising opportunities. Among the more prominent PPC Engines are:

- LookSmart
- MIVA
- ePilot
- Mamma, etc.

Paid inclusion

With paid inclusion and submission, site owners pay search engines to get their sites reviewed and included either into the general search index, or into a directory. Yahoo! is the main player in the market of paid inclusion. It's paid for inclusion programs are Search Submit, and Directory Submit.

Yahoo! Search Submit

Search Submit adds your site into Yahoo's index and ensures that it will be re-crawled every 48 hours (as opposed to around two weeks for sites that were automatically added by crawler to the index). Search Submit costs include a \$49 initial fee and cost-per-click fees (starting from \$0.15) for each click you receive.

Yahoo! Directory Submit

With Directory Submit, there is a \$299 fee for your site to be reviewed and (if accepted) added to Yahoo! Directory. After this, you will be subject to an annual recurring fee of \$299. For more information about Yahoo! Pay for Performance products, please visit <http://searchmarketing.yahoo.com>

Paid sponsorship

The category of Pay for Performance search mainly includes flat fee search advertising. ExactSeek (<http://exactseek.com>) is a nice example of such a model.

ExactSeek

With ExactSeek, you purchase a so-called "*Featured Listing*" (actually, your ad) on a selected keyword. Your listing then shows next to the search results for the selected keyword. You do not pay for every click - you pay for your presence in search results. There is no ad ranking system, if there are several listings for one keyword, they will show in random order.

Who Feeds Who - Search Engine Relationships

Search engines are large corporations with complex ownership and partnerships between them. While some are more technology oriented and mainly outsource their database and crawling software, others are commercially-specific and will speculate on their large user audience to sell you listings and advertising opportunities.

This brief section is intended to provide some insight into who owns who in the contemporary search world and which search engines provide results for which. By knowing this, you will better understand which of them are the best to target when optimizing your pages.

A while ago, there were many different engines, almost equally popular among visitors that were constantly competing in popularity, number of searches, efficiency of search technology, number of indexed pages, etc. Nowadays, after a period of incorporation, all the minor companies are owned by the larger entities.

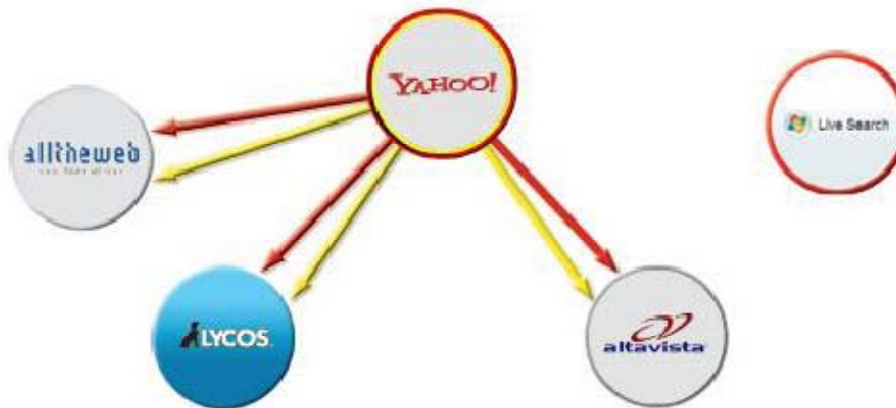
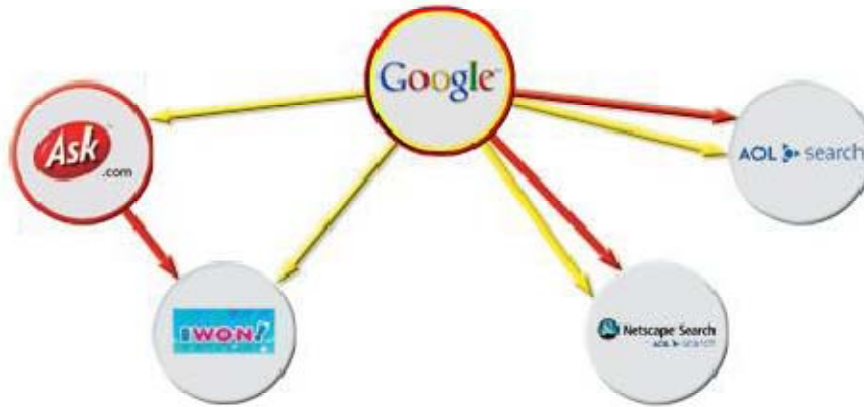
Who Owns Who

- Yahoo! owns AllTheWeb, Lycos and AltaVista. Yahoo! also owns and uses Inktomi's technology and database for all the partnering engines under its wing.
- Google partners with AOL, it is also a major result provider for many large engines because of its own cutting-edge technology (Netscape, Ask.com, and Iwon).
- Live Search does not own anybody and uses its own database (since February, 2005).



Who Feeds Who

The most outstanding work in the field of relationships between search result providers and consumers, and known to all SEO experts, is Bruce Clay's SE relationship chart. Here we provide a static representation. The dynamic (and much more convenient) version is available at <http://www.bruceclay.com/searchenginerelationshipchart.htm>

Here's the search engine relationship chart:



LEGEND

SUPPLIES  RECEIVES PRIMARY SEARCH RESULTS
SUPPLIES  RECEIVES PAID RESULTS

CLICK ON A LOGO FOR
SEARCH ENGINE INFORMATION

CLICK HERE TO SELECT A
DIFFERENT CHART

How Search Engines Rank Pages

Every smart Search Engine Optimizer starts his or her career by looking at Web pages with the eye of a search engine spider. Once the optimizer is able to do that, the path is half way complete to full mastery.

The first thing to remember is that the search engines rank "pages", not "sites". What this means is that you will not achieve a high ranking for your site by attempting to optimize your main page for ten different keyword phrases. However, different pages of your site WILL appear up the list for different key phrases if you optimize each page for just one of them. If you can't use your keyword in the domain name, no problem – use it in the URL of some page within your site, e.g. in the file name of the page. This page will rise in relevance for the given keyword. All search engines show you URLs of specific PAGES when you search – not just the root domain names like www.webceo.com but the paths like www.webceo.com/voxpathuli/index.htm.

Second, understand that the search engines do not see the graphics and JavaScript dynamics your page uses to captivate visitors. You can use a graphic image of written text that says you sell beautiful Christmas gifts. But it does not tell the search engine that your website is related to Christmas Gifts – unless you use an ALT attribute where you write about it.

Here's an example to illustrate.

What the visitor sees:



**Beautiful
Christmas
Gifts!!!**

What the search engine will read in this place:

```

```

As you see there's nothing in the code which could tell the search robots that the content relates to "Christmas", "Gifts", or "Beautiful". The situation will change if we rewrite the code like this:

```

```

As you can see we've added the ALT attribute with the value that corresponds to what the image tells your visitors. Initially, the "alt" attribute was meant to provide alternative text for an image that for some reason could not be shown by the visitor's browser. Nowadays it has acquired one

more function – to bring the same message to the search engines that the image itself brings to human Web surfers.

The same concerns the usage of JavaScript. Look at these two examples:

1. Visit our page about discounted Samsung Monitors!
2. `<script language="JavaScript" type="text/javascript"><!--document.write("Visit our page about " + goods[Math.round(0.5 +(3.99999 * Math.random()))-1]); --> </script>`

The first example is what visitors see, the second is the source code script that produces the output. Assume the search engine spider is intelligent enough to read the script (however, actually not all the spiders do); is there anything in the code that can tell it about the Samsung Monitor? Hardly.

As a rule, search engine spiders have a limit on loading page content. For instance, the Googlebot will not read more than 100 KB of your page, even though it is instructed to look whether there are keywords at the end of your page. So if you use keywords somewhere beyond this limit, this is invisible to spiders. Therefore, you may want to acquire the good habit of not overloading the HEAD section of your page with scripts and styles. Better link them from outside files, because otherwise they just push away your important textual content.

There are many more examples of relevancy indicators a spider considers when visiting your page, such as the proximity of important words to the beginning of the page. Here, as well, the spider does not necessarily see the same things a human visitor would see. For instance, a left-hand menu pane on your Web page. People visiting your site will generally not first pay attention to this, focusing instead on the main section. The spider, however, will read your menu before passing to the main content – simply because it is closer to the beginning of the code.

Remember: during the first visit, the spider does not yet know which words your page relates to! Keep in mind this simple truth. By reading your HTML code, the spider (which is just a computer program) must be able guess the exact words that make up the theme of your site.

Then, the spider will compress your page and create the index associated with it. To keep things simple, you can think of this index as an enumeration of all words found on your page, with several important parameters associated with each word: their proximity, frequency, etc.

Certainly, no one really knows what the real indices look like, but the principals are as they have been outlined here. The words that are high in the list according to the main criteria will be considered your keywords by the spider. In reality, the parameters are quite numerous and include off-page factors as well, because the spider is able to detect the words every other page out there uses when linking to your page, and thus calculate your relevance to those terms also.

When a Web surfer queries the search engine, it pulls out all pages in its database that contain the user's query. And here the ranking begins: each page has a number of "onpage" indicators associated with it, as well as certain page-independent indicators (like PageRank). A combination of these indicators determines how well the page ranks.

It's important to keep this in mind: after you have made your page attractive for visitors, ask yourself whether you have also made it readable for the search engine spiders.

Suggested Reading

Google Guide [How Google Works](#)

Live Search [About Website Ranking](#)

Danny Sullivan [Search Engine Technology](#)

Robert K. McCourty [Search Engine Observations: Return of The Metacrawler](#)

Bruce Clay, Inc. [The Latest Search Engine Relationship Chart](#)

Danny Sullivan [How Search Engines Rank Web Pages](#)

Chris Sherman [Metacrawlers and Metasearch Engines](#)

Webopedia.com [How Web Search Engines Work](#)

Dave Davies [Anatomy of an Internet Search Engine](#)

Rob Sullivan [Anatomy of a Search Engine Crawler](#)

Dave Davies [SEO For The Big Three](#)

Derek Taylor [META SEARCHING: THE GOOD, THE BAD, & THE UGLY](#)

Derek Fulford [Website Analysis: Will Search Engines Draw a Blank or Give You a Rank?](#)