



Technivision Knowledge Base  
**Search Engines**

A Brief Overview of How They Work  
In Everyday English!

January 1, 2009

Prepared by:  
**Kevin MacDonald**  
President

**1.0 INTRODUCTION.....3**  
Crawler-Based Search Engines .....3  
Human-Powered Directories.....3  
The Parts of a Crawler-Based Search Engine .....3

**2.0 GETTING YOUR SITE LISTED .....4**  
What Search Engines are Looking For .....4  
Keywords – Location and Frequency .....4  
Spice In The Recipe .....4  
Off The Page Factors .....4  
Information Captured by Search Engines – What is most important? .....5  
Relevant Content .....5  
Page Rank.....5  
How Important is Google? .....5  
How to Improve your Listings .....6

**3.0 WHAT TO EXPECT .....6**  
Give SEO Time!.....6  
Bottom Line .....6

## 1.0 Introduction

The term search engine is a generic term used to describe “crawler-based” search engines and “human-based” search directories. Both provide listings of web sites for searching users to help them locate the information they want.

### Crawler-Based Search Engines

Crawler-based search engines, such as Google, create their listings automatically. They “crawl” or “spider” the web, then people search through what they have found. If you change your web pages, crawler-based search engines eventually find these changes, and that can affect how you are listed. Page titles, body copy and other elements all play a role. All crawler-based search engines have the basic parts described above, but there are differences in how these parts are tuned. That is why the same search on different search engines often produces different results.

### Human-Powered Directories

A human-powered directory, such as the Open Directory (DMOZ.com) depends on humans for its listings. You submit a short description to the directory for your entire site, or editors write one for sites they review. Changing your web pages has no effect on your listing. Things that are useful for improving a listing with a search engine have nothing to do with improving a listing in a directory. The only exception is that a good site, with good content, might be more likely to get reviewed for free than a poor site.

This document will focus on the Crawler-Based search engines because most users on the internet search using these engines.

### The Parts of a Crawler-Based Search Engine

Crawler-based search engines have three major parts. First is the **spider**, also called the “crawler”. The spider visits a web page, reads it, and then follows links to other pages within the site. The spider then returns to the site on a regular basis, such as every month or two, to look for changes.

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

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

The search engine “**algorithm**” is the third part of a search engine. This is the program that sifts through the millions of pages recorded in the index to find matches to a search and rank them in order of “what it believes is most *relevant*” (We will discuss relevant content later in this document). 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 their own set of rules, in other words, their algorithm. Exactly how a particular search engine's algorithm works is a closely-kept trade secret!

## 2.0 Getting Your Site Listed

Web sites get found by Crawler-Based search engines one of two ways, either another site has linked to it or because the site was submitted manually to the search engine for inclusion into their index. Once the search engine becomes aware of the site, spiders are transmitted to gather information about what is on the site. This information is then stored on the index database.

The following describes the kind of information the spiders retrieve from a site.

### What Search Engines are Looking For

The effort known as Search Engine Optimization or SEO (as it is commonly referred to) is the work done to a web site that enables search engines to list the site properly, and provide the best possible listing position. What follows are a few of the key factors that go into the SEO work effort.

### Keywords – Location and Frequency

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

Think about the last time you were in the library. When you needed to find a book chances are you looked at the book titles listed in that old “card catalog” to determine what book you were most interested in. Then you would locate that book in a certain section of the library, based upon your search method. Search engines operate the same way. Pages with the search terms appearing in the HTML title tag are often assumed to be most relevant to the topic.

Search engines will also check to see if the search keywords/phrases appear near the top of a web page, such as in the headline or in the first few paragraphs of text. They presume that any page relevant to the topic will mention those words right from the beginning. These “keywords” are in the actual text copy of the page. This is a very important point and will be addressed later in this document.

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

### Spice In The Recipe

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

### Off The Page Factors

Off the page factors are those that webmasters cannot easily influence. Chief among these is *link popularity*. By analyzing how many other sites link to your site, a search engine can determine not only what a page is about, but can conclude whether that page is deemed to be “important”. Accordingly, link popularity could qualify the site as 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 “click-through” measurement. In short, this means that a search engine may monitor results selected for a particular search word or phrase, then eventually drop high-ranking pages that aren't attracting clicks, while promoting lower-ranking pages that do pull in visitors.

## Information Captured by Search Engines – What is most important?

Data Recorded by Search Engines – Listed in relative importance

	Data Description	YES	NO
1	Page Titles	ALL Search Engines	N/A
2	Text Used on Webpage	ALL Search Engines	N/A
3	Meta Description	ALL Search Engines	N/A
4	Page URL file name	Google	AltaVista, MSN, Yahoo, Ask
5	ALT text ( <i>image description code</i> )	AltaVista, Google, Ask	MSN, Yahoo
6	Meta Keywords	Yahoo, Ask	MSN, AltaVista, Google

This chart clearly demonstrates that the most important factors in getting good search engine rankings are page titles, the text on a web page, and the metadata (code) description of that page. The more consistent and relative they are to one another, the more favorable a page will appear.

*Note that Google, the most popular search engine in the world, does not collect Keyword data! The use of keyword metadata is actually less important than is commonly thought.*

### Relevant Content

It is extremely important to be aware that the increased relevance between page titles, content, and metadata, it is likely a site will be promoted to a more advantageous listing. The more content that exists on a page about a relevant topic the better! Therefore, one can deduce that relevant content is now what drives search engine rankings.

### Page Rank

However, there is one other aspect that also impacts a sites listing, it is called Page Rank. The only search engine that considers Page Rank is Google. They created Page Rank as a way to weigh the importance of one site over another. Google displays search results initially by the page rank. Google's Page Rank is determined when links are established from one site (in promotion of, or in relation) to another, by attaching a site's URL. A greater weight (or value) is given to authoritative sites. This is considered "Link Analysis". Every major search engine uses link analysis as part of its ranking algorithm. Link analysis gives search engines a more useful means of determining which pages are beneficial for particular topics.

By increasing inbound links, you can help improve how well your pages rank on some search engines. The key to understanding page rank is that link analysis is not about "popularity". In other words, it's not an issue of attaching a multitude of links from just anywhere. Instead, you want links from good web pages / sites that pertain to the topics in your site content. This is what ultimately determines your site rank.

### How Important is Google?

In December of 2008, Americans conducted 12.7 billion searches on the main search engines. This is up 3 percent from November 2008 and continues to grow. Google sites handled 8 billion core searches, followed by Yahoo! sites with 2.6 billion and Microsoft sites with 1.1 billion.

Five primary search engines used in 2008 and their percent of use by internet searchers.

1. Google 65%
2. Yahoo. 20%
3. Microsoft (MSN) 8%
4. Ask 4%
5. Others 3%

### How to Improve your Listings

Focus most of your SEO efforts to on the top three Data Elements - page titles, text used in web content and Meta descriptions - captured by all search engines, and increase their "relevance" to one another. Interestingly, if your site has plenty of relevant content to a specific topic, it is likely that authoritative sites will link to yours. This will improve your Page Rank and will ultimately elevate your listing on Google. As this process continues to grow, your site listing will continue to advance. The more content you add to your site about a relevant topic, the more interest it gains, more sites link to it, your Page Rank improves and **Bam!** Your site is now Finally on the First Page! It sounds simple, right? The question is, how long will it take?

It is also important to realize, that while you are just learning how this process works, others have been scrutinizing the search engine methods for a long time (years), and using this knowledge to enhance their site listings. Realistically, it is fair to assume there will be plenty of competition for the best listing positions on search engines. This unfortunately, makes it even more difficult to obtain a good listing for your site, and it will take even longer. Then, once you have finally arrived in the top listings, you can bet those who once held that position are now going to fight to get it back! The SEO process never ends!

## 3.0 What to Expect

### Give SEO Time!

#### How Long Does it Take for SEO efforts to Show Results?

The answer is, there is no answer!

We don't know for certain when results will kick in. We don't own the search engines. SEO service companies are like public relations companies. We can do the work to help your presence, but we don't control the rankings. For a new Web site trying to compete in one of the most competitive areas of SEO, the answer for "how long" might be two years, or even more.

Now, for sites trying to compete against local keywords, with a very niche product or service offering, there's a great chance to see results quickly. Even "quickly" depends upon factors such as domain age, links, a properly built site, quality of content and most importantly, relevancy. Graphics-heavy or Flash only web sites, particularly those built in frames with virtually no text content, have little hope of ranking for anything, ever!

What about companies that have had a Web site for many years, but haven't given any consideration to SEO? These sites probably already have some good content and natural links to their Web site. All they need is some help structuring things correctly, using good SEO practices and gaining more valuable links. Even then, without knowing the competitive nature of the other Web, it's difficult to project.

### Bottom Line

SEO is an investment that will show you returns if you're committed to the process for however long it takes.

There are only 10 "top" positions in the search engines. Chances are, many of those in the top 10 have been doing SEO for a while. To compete with them, you'll have to do it for a while, too.