

# Part I

## Understanding SEO

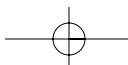
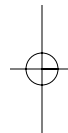
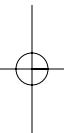
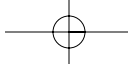
**S**earch engine optimization (SEO) is such a broad term. It can be quite overwhelming if you try to take the whole of it in a single bite. There are so many facets of search engine optimization, from how search engines work (and they all work a little differently) to how a web page is designed. There are enough elements to worry about that you could spend far more time than you can afford to invest in trying to achieve the SEO you have in mind. However, search engine optimization doesn't have to be such an onerous task that it can't be accomplished. Not if you understand what it is and how it works.

Part I explains the basics of search engine optimization. This part includes an explanation of what search engines are and how they work. There is also an explanation of the concept of an SEO plan. Together, these two elements will have you up to speed and ready to begin implementing the right SEO strategies to build the web site traffic that you need.

### IN THIS PART

**Chapter 1**  
**Search Engine Basics**

**Chapter 2**  
**Creating an SEO Plan**





# Chapter 1

## Search Engine Basics

**W**hat do you do when you need to find something on the Internet? In most cases, you pop over to one of the major search engines and type in the term or phrase that you're looking for and then click through the results, right? But of course search engines weren't always around.

In its infancy, the Internet wasn't what you think of when you use it now. In fact, it was nothing like the web of interconnected sites that's become one of the greatest business facilitators of our time. Instead, what was called the Internet was actually a collection of *FTP* (*File Transfer Protocol*) sites that users could access to download (or upload) files.

To find a specific file in that collection, users had to navigate through each file. Sure, there were shortcuts. If you knew the right people—that would be the people who knew the exact address of the file you were looking for—you could go straight to the file. That's assuming you knew exactly what you were looking for.

The whole process made finding files on the Internet a difficult, time-consuming exercise in patience. But that was before a student at McGill University in Montreal decided there had to be an easier way. In 1990, Alan Emtage created the first search tool used on the Internet. His creation, an index of files on the Internet, was called Archie.

If you're thinking Archie, the comic book character created in 1941, you're a little off track (at least for now). The name Archie was used because the file name Archives was too long. Later, Archie's pals from the comic book series (Veronica and Jughead) came onto the search scene, too, but we'll get to that shortly.

### IN THIS CHAPTER

**What is a search engine?**

**Anatomy of a search engine**

**Characteristics of search**

**Classifications of search engines**

**Putting search engines to work**

**Manipulating search engines**

**Part I****Understanding SEO**

Archie wasn't actually a *search engine* like those that you use today. But at the time, it was a program many Internet users were happy to have. The program basically downloaded directory listings for all of the files that were stored on *anonymous FTP* sites in a given network of computers. Those listings were then plugged into a searchable database of web sites.

The search capabilities of Archie weren't as fancy as the *natural language capabilities* you'll find in most common search engines today, but at the time it got the job done. Archie indexed computer files, making them easier to locate.

In 1991, however, another student named Mark McCahill, at the University of Minnesota, decided that if you could search for files on the Internet, then surely you could also search plain text for specific references in the files. Because no such application existed, he created Gopher, a program that indexed the plain-text documents that later became the first web sites on the public Internet.

With the creation of Gopher, there also needed to be programs that could find references within the indexes that Gopher created, and so Archie's pals finally rejoined him. Veronica (Very Easy Rodent-Oriented Net-wide Index to Computerized Archives) and Jughead (Jonzy's Universal Gopher Hierarchy Excavation and Display) were created to search the files that were stored in the Gopher Index System.

Both of these programs worked in essentially the same way, allowing users to search the indexed information by keyword.

From there, search as you know it began to mature. The first *real* search engine, in the form that we know search engines today, didn't come into being until 1993. It was developed by Matthew Gray, and it was called Wandex. Wandex was the first program to both index and search the index of pages on the Web. This technology was the first program to crawl the Web, and later became the basis for all search crawlers. And from there, search engines took on a life of their own. From 1993 to 1998, the major search engines that you're probably familiar with today were created:

- Excite — 1993
- Yahoo! — 1994
- Web Crawler — 1994
- Lycos — 1994
- Infoseek — 1995
- AltaVista — 1995
- Inktomi — 1996
- Ask Jeeves — 1997
- Google — 1997
- MSN Search — 1998

Today, search engines are sophisticated programs, many of which allow you to search all manner of files and documents using the same words and phrases you would use in everyday conversations. It's hard to believe that the concept of a search engine is just over 15 years old. Especially considering what you can use one to find these days!

## What Is a Search Engine?

Okay, so you know the basic concept of a search engine. Type a word or phrase into a search box and click a button. Wait a few seconds, and references to thousands (or hundreds of thousands) of pages will appear. Then all you have to do is click through those pages to find what you want. But what exactly is a search engine, beyond this general concept of “seek and ye shall find”?

It's a little complicated. On the back end, a search engine is a piece of software that uses applications to collect information about web pages. The information collected is usually key words or phrases that are possible indicators of what is contained on the web page as a whole, the URL of the page, the code that makes up the page, and links into and out of the page. That information is then indexed and stored in a database.

On the front end, the software has a user interface where users enter a search term — a word or phrase — in an attempt to find specific information. When the user clicks a search button, an algorithm then examines the information stored in the back-end database and retrieves links to web pages that appear to match the search term the user entered.

### CROSS-REF

You can find more information about web crawlers, spiders, and robots in Chapter 14.

The process of collecting information about web pages is performed by an agent called a crawler, spider, or robot. The crawler literally looks at every URL on the Web, and collects key words and phrases on each page, which are then included in the database that powers a search engine. Considering that the number of sites on the Web went over 100 million some time ago and is increasing by more than 1.5 million sites each month, that's like your brain cataloging every single word you read, so that when you need to know something, you think of that word and every reference to it comes to mind.

In a word . . . overwhelming.

## Anatomy of a Search Engine

By now you probably have a fuzzy picture of how a search engine works. But there's much more to it than just the basic overview you've seen so far. In fact, search engines have several parts. Unfortunately, it's rare that you find an explanation for just how a search engine is made — and that information is vitally important to succeeding with *search engine optimization* (SEO).

**Part I****Understanding SEO**

## Query interface

The query interface is what most people are familiar with, and it's probably what comes to mind when you hear the term "search engine." The query interface is the page that users see when they navigate to a search engine to enter a search term.

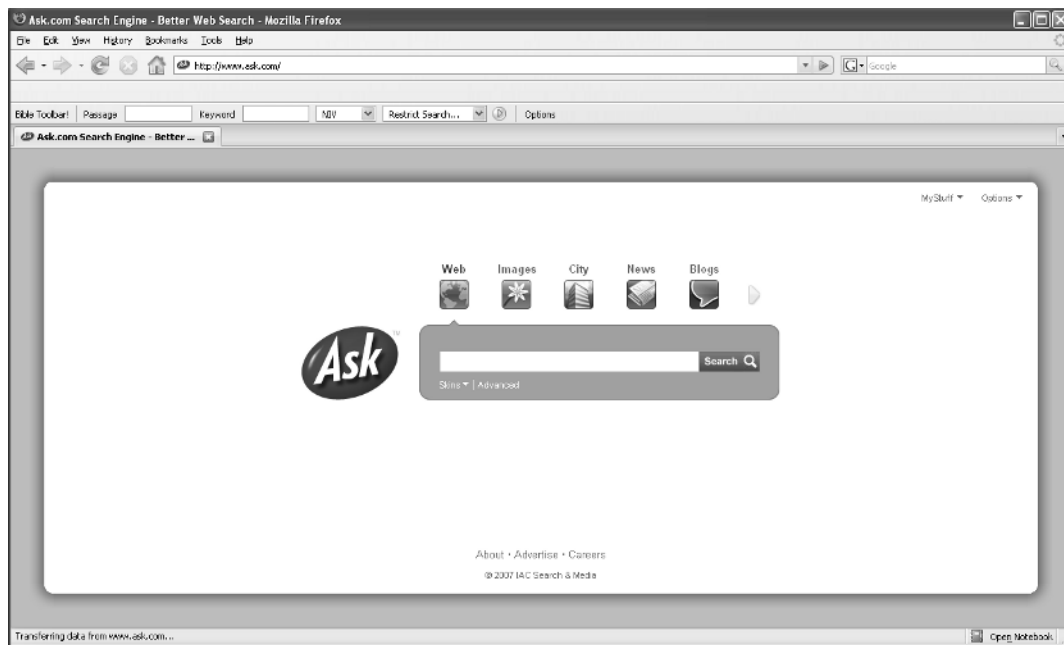
There was a time when the search engine interface looked very much like the Ask.com page shown in Figure 1-1. The interface was a simple page with a search box and a button to activate the search.

Today, many search engines on the Web have added much more personalized content in an attempt to capitalize on the real estate available to them. For example, Yahoo! Search, shown in Figure 1-2, allows users to personalize their pages with a free e-mail account, weather information, news, sports, and many other elements designed to make users want to return to that site to conduct their web searches.

One other option users have for customizing the interfaces of their search engines is a capability like the one Google offers. The Google search engine has a customizable interface to which users can add different *gadgets*. These gadgets allow users to add features to their customized Google search home that meet their own personal needs or tastes.

**FIGURE 1-1**

The Ask.com search page shows how most search engine interfaces used to look.



**FIGURE 1-2**

Yahoo! Search allows users to make their search page more personal.



When it comes to search engine optimization, Google's user interface offers the most ability for you to reach your target audience, because it does more than just optimize your site for search; if there is a useful tool or feature available on your site, you can allow users to have access to this tool or feature through the *Application Programming Interface (API)* made available by Google. This allows you to have your name in front of users on a daily basis.

**CROSS-REF** You can find more information about Google APIs in Appendix A in the section "Optimization for Google."

For example, a company called PDF24.org has a Google gadget that allows users to turn their documents into PDF files, right from their Google home page once the gadget has been added. If the point of search engine optimization is ultimately to get your name in front of as many people as possible, as often as possible, then making a gadget available for addition to Google's personalized home page can only further that goal.

## Crawlers, spiders, and robots

The query interface is the only part of a search engine that the user ever sees. Every other part of the search engine is behind the scenes, out of view of the people who use it every day. That doesn't mean it's not important, however. In fact, what's in the back end is the most important part of the search engine.

**CROSS-REF** There's more in-depth information about crawlers, spiders, and robots in Chapter 14.

## Part I

## Understanding SEO

If you've spent any time on the Internet, you may have heard a little about spiders, crawlers, and robots. These little creatures are programs that literally crawl around the Web, cataloging data so that it can be searched. In the most basic sense all three programs—crawlers, spiders, and robots—are essentially the same. They all “collect” information about each and every web URL.

This information is then cataloged according to the URL on which they're located and are stored in a database. Then, when a user uses a search engine to locate something on the Web, the references in the database are searched and the search results are returned.

### Databases

Every search engine contains or is connected to a system of databases, where data about each URL on the Web (collected by crawlers, spiders, or robots) is stored. These databases are massive storage areas that contain multiple data points about each URL.

The data might be arranged in any number of different ways, and will be ranked according to a method of ranking and retrieval that is usually proprietary to the company that owns the search engine.

### Search algorithms

All of the parts of the search engine are important, but the *search algorithm* is the cog that makes everything work. It might be more accurate to say that the search algorithm is the foundation on which everything else is built. How a search engine works is based on the search algorithm, or the way that data is discovered by the user.

In very general terms, a search algorithm is a problem-solving procedure that takes a problem, evaluates a number of possible answers, and then returns the solution to that problem. A search algorithm for a search engine takes the problem (the word or phrase being searched for), sifts through a database that contains cataloged keywords and the URLs those words are related to, and then returns pages that contain the word or phrase that was searched for, either in the body of the page or in a URL that points to the page.

This neat little trick is accomplished differently according to the algorithm that's being used. There are several classifications of search algorithms, and each search engine uses algorithms that are slightly different. That's why a search for one word or phrase will yield different results from different search engines. Some of the most common types of search algorithms include the following:

- **List search:** A list search algorithm searches through specified data looking for a single key. The data is searched in a very linear, list-style method. The result of a list search is usually a single element, which means that searching through billions of web sites could be very time-consuming, but would yield a smaller search result.
- **Tree search:** Envision a tree in your mind. Now, examine that tree either from the roots out or from the leaves in. This is how a tree search algorithm works. The algorithm searches a data set from the broadest to the most narrow, or from the most narrow to the broadest. Data sets are like trees; a single piece of data can branch to many other pieces of data, and



this is very much how the Web is set up. Tree searches, then, are more useful when conducting searches on the Web, although they are not the only searches that can be successful.

- **SQL search:** One of the difficulties with a tree search is that it's conducted in a hierarchical manner, meaning it's conducted from one point to another, according to the ranking of the data being searched. A SQL (pronounced See-Quel) search allows data to be searched in a non-hierarchical manner, which means that data can be searched from any *subset* of data.
- **Informed search:** An informed search algorithm looks for a specific answer to a specific problem in a tree-like data set. The informed search, despite its name, is not always the best choice for web searches because of the general nature of the answers being sought. Instead, informed search is better used for specific queries in specific data sets.
- **Adversarial search:** An adversarial search algorithm looks for all possible solutions to a problem, much like finding all the possible solutions in a game. This algorithm is difficult to use with web searches, because the number of possible solutions to a word or phrase search is nearly infinite on the Web.
- **Constraint satisfaction search:** When you think of searching the Web for a word or phrase, the constraint satisfaction search algorithm is most likely to satisfy your desire to find something. In this type of search algorithm, the solution is discovered by meeting a set of constraints, and the data set can be searched in a variety of different ways that do not have to be linear. Constraint satisfaction searches can be very useful for searching the Web.

These are only a few of the various types of search algorithms that are used when creating search engines. And very often, more than one type of search algorithm is used, or as happens in most cases, some proprietary search algorithm is created. The key to maximizing your search engine results is to understand a little about how each search engine you're targeting works. Only when you understand this can you know how to maximize your exposure to meet the search requirements for that search engine.

## Retrieval and ranking

For a web search engine, the retrieval of data is a combination activity of the crawler (or spider or robot), the database, and the search algorithm. Those three elements work in concert to retrieve the word or phrase that a user enters into the search engine's user interface. And as noted earlier, how that works can be a proprietary combination of technologies, theories, and coding whizbangery.

The really tricky part comes in the results ranking. Ranking is also what you'll spend the most time and effort trying to affect. Your ranking in a search engine determines how often people see your page, which affects everything from revenue to your advertising budget. Unfortunately, how a search engine ranks your page or pages is a tough science to pin down.

The most that you can hope for, in most cases, is to make an educated guess as to how a search engine ranks its results, and then try to tailor your page to meet those results. But keep in mind that, although retrieval and ranking are listed as separate subjects here, they're actually part of the search algorithm. The separation is to help you better understand how search engines work.

## Part I

## Understanding SEO

Ranking plays such a large part in search engine optimization that you'll see it frequently in this book. You'll look at ranking from every possible facet before you reach the last page. But for now, let's look at just what affects ranking. Keep in mind, however, that different search engines use different ranking criteria, so the importance each of these elements plays will vary.

- **Location:** Location doesn't refer here to the location (as in the URL) of a web page. Instead, it refers to the location of key words and phrases on a web page. So, for example, if a user searches for "puppies," some search engines will rank the results according to where on the page the word "puppies" appears. Obviously, the higher the word appears on the page, the higher the rank might be. So a web site that contains the word "puppies" in the *title tag* will likely appear higher than a web site that is about puppies but does not contain the word in the title tag. What this means is that a web site that's not designed with SEO in mind will likely not rank where you would expect it to rank. The site [www.puppies.com](http://www.puppies.com) is a good example of this. In a Google search, it appears ranked fifth rather than first, potentially because it does not contain the key word in the title tag.
- **Frequency:** The frequency with which the search term appears on the page may also affect how a page is ranked in search results. So, for example, on a page about puppies, one that uses the word five times might be ranked higher than one that uses the word only two or three times. When word frequency became a factor, some web site designers began using hidden words hundreds of times on pages, trying to artificially boost their page rankings. Most search engines now recognize this as *keyword spamming* and ignore or even refuse to list pages that use this technique.
- **Links:** One of the more recent ranking factors is the type and number of links on a web page. Links that come into the site, links that lead out of the site, and links within the site are all taken into consideration. It would follow, then, that the more links you have on your page or leading to your page the higher your rank would be, right? Again, it doesn't necessarily work that way. More accurately, the number of relevant links coming into your page, versus the number of relevant links within the page, versus the number of relevant links leading off the page will have a bearing on the rank that your page gets in the search results.
- **Click-throughs:** One last element that might determine how your site ranks against others in a search is the number of *click-throughs* your site has versus click-throughs for other pages that are shown in page rankings. Because the search engine cannot monitor site traffic for every site on the Web, some monitor the number of clicks each search result receives. The rankings may then be repositioned in a future search, based on this interaction with the users.

Page ranking is a very precise science. And it differs from search engine to search engine. To create the best possible SEO for your site, it's necessary to understand how these page rankings are made for the search engines you plan to target. Those factors can then be taken into consideration and used to your advantage when it's time to create, change, or update the web site that you want to optimize.

## Characteristics of Search

Understanding how a search engine works helps you to understand how your pages are ranked in the search engine, but how your pages are found is another story entirely. That's where the human element comes in. Search means different things to different people. For example, one of my colleagues searches the Internet using the same words and phrases he would use to tell someone about a topic or even the exact question that he's trying to get answered. It's called *natural language*. Another, however, was trained in search using *Boolean search techniques*. She uses a very different *syntax* when she's creating a search term. Each of them returns different search results, even when each is using the same search engines.

The characteristics of search refer to how users search the Internet. This can be everything from the *heuristics* they use when creating a search term to the selection the user makes (and the way those selections are made) once the search results are returned. One interesting fact is that more than half of American adults search the Internet every time they go online. And in fact, more people search the Internet than use the yellow pages when they're looking for phone numbers or the locations of local businesses.

This wealth of search engine users is fertile ground for SEO targeting. And the better you understand how and why users use search engines, and exactly how search engines work, the easier it will be to achieve the SEO you're pursuing.

## Classifications of Search Engines

With a decent understanding of how search engines work and how people use those search engines, you can now concentrate on some more detailed information about these engines. For example, you know that all search engines aren't created equal, right? But did you know that there are different types, or classifications, of search engines? There are.

Search engines can be broken down into three different types (in the broadest of terms): primary, secondary, and targeted.

### Primary search engines

A *primary search engine* is the type you think of most often when search engines come to mind. Some index most or all sites on the Web. For example, Yahoo! Google, and MSN are primary (also called major) search engines.

Primary search engines will generate the majority of the traffic to your web site, and as such will be the primary focus of your SEO efforts. Each primary search engine differs slightly from the others.

**Part I****Understanding SEO**

For example, Lycos has been around much longer than Google, yet Google is the most popular search engine on the Web. Why is that? Most likely because people find that, when searching the Web, Google provides better search results.

The difference in those search results is all in the search algorithm used to create the search engine.

Most primary search engines are also more than just search. Additional features such as e-mail, mapping, news, and different types of entertainment applications are also available from most of the primary search engine companies. These elements were added long after the search was established, as a way to draw more and more people to the search engine. Although those features don't change the way people search, they might affect which search engine people choose.

**Overview of Google**

Each of the major search engines differs in some small way. Google is the king of search engines, in part because of the accuracy with which it can pull the results from a search query. Sure, Google offers all kinds of extras like e-mail, a personalized home page, and even productivity applications, but those value-added services are not what made Google popular.

What turned Google into a household word is the accuracy with which the search engine can return search results. This accuracy was developed when the Google designers combined keyword searches with link popularity. The combination of the keywords and the popularity of links to those pages yields a higher accuracy rank than just keywords alone.

However, it's important to understand that link popularity and keywords are just two of hundreds of different criteria that search engines can use in ranking the relevancy of web pages.

**Overview of Yahoo!**

Most people assume that Yahoo! is a search engine, and it is. But it's also a *web directory*, which basically means that it's a list of the different web pages available on the Internet, divided by category and subcategory. In fact, what few people know is that Yahoo! started as the favorites list of the two young men who founded it. Through the acquisition of companies like Inktomi, All the Web, AltaVista, and Overture, Yahoo! gradually gained market share as a search engine.

Yahoo!, which at one time used Google to search its directory of links, now ranks pages through a combination of the technologies that it acquired over time. However, Yahoo!'s link-ranking capability is not as accurate as Google's. In addition, Yahoo! also has a paid inclusion program, which some think tends to skew search results in favor of the highest payer.

**Overview of MSN**

MSN's search capabilities aren't quite as mature as those of Google or Yahoo! As a result of this immaturity, MSN has not yet developed the in-depth link analysis capabilities of these other primary search engines. Instead, MSN relies heavily on web-site content for ranking purposes. However, this may have a beneficial effect for new web sites that are trying to get listed in search engines.

The link-ranking capabilities of Google and Yahoo! can preclude new web sites from being listed for a period of time after they have been created. This is because (especially where Google is concerned)

the quality of the link may be considered during ranking. New links are often ignored until they have been in place for a time.

Because MSN relies heavily on page content, a web site that is tagged properly and contains a good ratio of keywords will be more likely to be listed — and listed sooner — by the MSN search engine. So, though it's not the most popular of search engines, it is one of the primaries, and being listed there sooner rather than later will help increase your site traffic.

## Secondary search engines

*Secondary search engines* are targeted at smaller, more specific audiences, although the search engine's content itself is still general. They don't generate as much traffic as the primary search engines, but they're useful for regional and more narrowly focused searches. Examples of secondary search engines include Lycos, LookSmart, Miva, Ask.com, and Espotting.

Secondary search engines, just like the primary ones, will vary in the way they rank search results. Some will rely more heavily upon keywords, whereas others will rely on reciprocal links. Still others might rely on criteria such as *meta tags* or some proprietary criteria.

Secondary search engines should be included in any SEO plan. Though these search engines might not generate as much traffic as the primary search engines, they will still generate valuable traffic that should not be overlooked. Many users of secondary search engines are users because they have some loyalty to that specific search engine. For example, many past AOL users who have moved on to broadband Internet service providers still use the AOL search engine whenever possible, because it's comfortable for them.

## Targeted search engines

*Targeted search engines* — sometimes called topical search engines — are the most specific of them all. These search engines are very narrowly focused, usually to a general topic, like medicine or branches of science, travel, sports, or some other topic. Examples of targeted search engines include CitySearch, Yahoo! Travel, and MusicSearch, and like other types of search engines, ranking criteria will vary from one to another.

When considering targeted search engines for SEO purposes, keep in mind that many of these search engines are much more narrowly focused than primary or secondary search engines. Look for the targeted search engines that are relevant to your specific topic (like pets, sports, locations, and so on).

## Putting Search Engines to Work for You

All this information about search engines has one purpose — to show you how they work, so that you can put them to work for you. Throughout this book, you'll find various strategies for optimizing your web site so it appears high in search engine rankings when relevant searches are performed. But this requires that you know how to put search engines to work for you.



**Part I****Understanding SEO**

Search engine optimization is essentially the science of designing your web site to maximize your search engine rankings. This means that all of the elements of your web site are created with the goal of obtaining high search engine rankings. Those elements include:

- Entry and exit pages
- Page titles
- Site content
- Graphics
- Web site structure

In addition to these elements, however, you also have to consider things like keywords, links, HTML, and meta-tagging. Even after you have all the elements of your page optimized for search-engine friendliness, there are other things to consider. For example, you can have all the right design elements included in your web pages and still have a relatively low search engine ranking. Factors such as advertising campaigns and update frequencies also affect your SEO efforts.

All of this means that you should understand that the concept of search engine optimization is not based on any single element. Instead, search engine optimization is based on a vast number of elements and strategies. And it's an ongoing process that doesn't end once your web site is live.

SEO is a living, breathing concept of maximizing the traffic that your web site generates, and because it is, that means that it's a constantly moving target. If you've ever played a game of Whack-a-Mole, you can understand how difficult search engine optimization is to nail. In the game, a little mole pops up out of a hole. Your job is to whack the mole on the top of the head before it disappears back down the hole and appears in another.

Search engine optimization is much the same concept. Search engines are constantly changing, so the methods and strategies used to achieve high search engine rankings must also change. As soon as that little mole pops up in one hole, he disappears and then reappears in another. It's a frustrating game, but given enough time and concentration, you can become very good at it.

## **Manipulating Search Engines**

There's one more topic to touch on before this chapter is finished. SEO is about manipulating search engines—to an extent. Beyond that, the manipulation becomes something more sinister and you run the risk of having your web site removed from the search engine rankings completely. It's true. It happens.

So what exactly can and can't you do? There's a list. Here is part of it.

You can:

- Create a web site that contains meta tags, content, graphics, and keywords that help improve your site ranking.
- Use keywords liberally on your site, so long as they are used in the correct context of your site topic and content.
- Include reciprocal links to your site from others as long as those links are legitimate and relevant.
- Encourage web site traffic through many venues, including keyword advertising, reciprocal links, and marketing campaigns.
- Submit your web site to search engines manually, rather than waiting for them to pick up your site in the natural course of cataloging web sites.

You can't:

- Trick search engines by imbedding hidden keywords in your web site. This is a practice that will very likely get you banned by most search engines.
- Artificially generate links to your site from unrelated sites for the purpose of increasing your ranking based on link analysis. Most search engines have a built-in mechanism that will detect this type of deceptive practice.
- Artificially generate traffic to your web site so that it appears more popular than it is. Again, there are safeguards in place to prevent this from happening, and if you trip those safeguards, you could end up on the banned list for many search engines.
- Force your web site to appear in search engine rankings by submitting the site repeatedly for inclusion in the rankings. A good general rule of thumb is that you should submit your site once and then wait at least six weeks before submitting it again. Submitting it repeatedly will, again, only lead to something nasty like being banned from the search engine.
- Expect search engines to automatically rank you at the top of your topic, category, or keyword as soon as the site is picked up. It could take a little time to build the "status" that you need to reach a high search engine ranking. Remember, SEO is a process.

These are just basic rules for putting search engines to work for you. There are many more, which you will discover in the coming chapters. As you get started, however, keep these in mind, because you'll see them over and over again throughout the course of this book and any other research that you might be doing on search engine optimization.

