

Chapter 1

Surveying the Search Engine Landscape

In This Chapter

- ▶ Discovering where people search
 - ▶ Understanding the difference between search sites and search systems
 - ▶ Distilling thousands of search sites down to three search systems
 - ▶ Understanding how search engines work
 - ▶ Gathering tools and basic knowledge
-

You've got a problem. You want people to visit your Web site; that's the purpose, after all — to bring people to your site to buy your product, or find out about your service, or hear about the cause you support, or for whatever other purpose you've built the site. So you've decided you need to get traffic from the search engines — not an unreasonable conclusion, as you find out in this chapter. But there are *so many* search engines! You have the obvious ones — Google, AOL, Yahoo!, and Bing (formerly MSN) — but you've probably also heard of others: HotBot, Dogpile, Ask.com, Netscape, and EarthLink. There's also Lycos, InfoSpace, Mamma.com, WebCrawler, and many more. To top it all off, you've seen advertising asserting that for only \$49.95 (or \$19.95, or \$99.95, or whatever sum seems to make sense to the advertiser), you, too, can have your Web site listed in hundreds, nay, thousands of search engines. You may have even used some of these services, only to discover that the flood of traffic you were promised turns up missing.

Well, I've got some good news. You can forget almost all the names I just listed — well, at least you can after you read this chapter. The point of this chapter is to take a complicated landscape of thousands of search sites and whittle it down into the small group of search systems that really matter. (Search sites? Search systems? Don't worry; I explain the distinction in a moment.)

If you really want to, you can jump to the “Where Do People Search?” section (near the end of the chapter) to see the list of search systems you need to worry about and ignore the details. But I’ve found that when I give this list to someone, he or she looks at me like I’m crazy because they know that some popular search sites aren’t on the list. This chapter explains why.

Investigating Search Engines and Directories

The term *search engine* has become the predominant term for *search system* or *search site*, but before reading any further, you need to understand the different types of search, um, things that you’re going to run across. Basically, you need to know about four things.

Search indexes or search engines



Search indexes or engines are the predominant type of search tools you’ll run across. Originally, the term *search engine* referred to some kind of search index, a huge database containing information from individual Web sites.



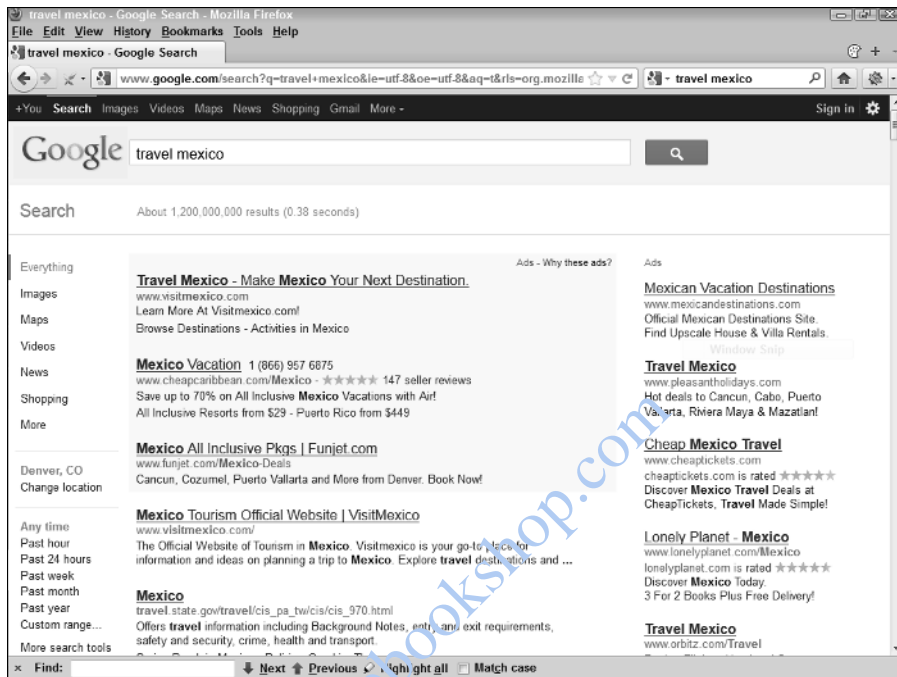
Large search-index companies own thousands of computers that use software known as *spiders* or *robots* (or just plain *bots*) to grab Web pages and read the information stored in them. These systems don’t always grab all the information on each page or all the pages in a Web site, but they grab a significant amount of information and use complex *algorithms* — calculations based on complicated formulae — to index that information. Google, shown in Figure 1-1, is the world’s most popular search engine, followed by Yahoo! and Bing, who round out the top three.

Index envy

Late in 2005, Yahoo! (www.yahoo.com) claimed that its index contained information for about 20 billion pages, along with almost 2 billion images and 50 million audio and video pages. Google (www.google.com) used to actually state on its home page how many pages it indexed — it reached 15 billion or so at

one point — but decided not to play the “mine is bigger than yours” game with Yahoo! and removed the “mine is bigger than yours” boast. Still, in 2008 Google reported that it had one *trillion* URLs in its index! Since then, I haven’t seen an update. Just assume that it’s more than you can shake the proverbial stick at.

Figure 1-1:
Google,
the world's
most popular
search engine,
produced these
results.



Search directories

A *directory* is a categorized collection of information about Web sites. Rather than containing information *from* Web pages, it contains information *about* Web sites. The most significant search directories are owned by Yahoo! (<http://dir.yahoo.com>) and the Open Directory Project (www.dmoz.org; see Figure 1-2). Directories are not created using spiders or bots to download and index pages on the Web sites in the directory; rather, for each Web site, the directory contains information, such as a title, description, and category, submitted by the site owner. The two most important directories, Yahoo! and Open Directory, have staff members who examine all the sites in the directory to make sure they're placed into the correct categories and meet certain quality criteria. Smaller directories often accept sites based on the owners' submission, with little verification.

Search directories are simply nowhere near as important today as in the past. In 2011, in fact, Google gave up on its own directory; until then <http://dir.google.com> led to a Google directory based on the Open Directory Project data.



Figure 1-2:
The Open
Directory
Project.



Here's how to see the difference between Yahoo!'s search results and the Yahoo! directory:

1. **Go to** www.yahoo.com.
2. **Type a word into the Search box.**
3. **Click the Search button.**

The list of Web sites that appears is called the Yahoo! Search results. (The results are actually provided by Bing.)

4. **Look for Directory above the Search box.**

Above the Search box, you see either a Directory link or a More link that opens a drop-down menu with the Directory option inside the menu; either way, click Directory, and you end up in Yahoo! Directory. (You can also go directly to the directory by using <http://dir.yahoo.com>.)

Nonspidered indexes

I wasn't sure what to call these things, so I made up a name: *nonspidered indexes*. A number of small indexes, less important than major indexes (such as Google), don't use spiders to examine the full contents of each page in the index. Rather, the index contains background information about each page,

such as titles, descriptions, and keywords. In some cases, this information comes from the meta tags pulled off the pages in the index. (I tell you about meta tags in Chapter 3.) In other cases, the person who enters the site into the index provides this information. A number of the smaller systems discussed in Chapter 13 are of this type.

Pay-per-click systems

Some systems provide *pay-per-click* listings. Advertisers place small ads into the systems, and when users perform their searches, the results contain some of these sponsored listings, typically above and to the right of the free listings. Pay-per-click systems are discussed in an additional chapter posted at www.SearchEngineBulletin.com.



Keeping the terms straight

Here are a few additional terms that you'll see scattered throughout the book:

✓ **Search site:** This Web site lets you search through some kind of index or directory of Web sites, or perhaps both an index and directory. (In some cases, search sites known as *meta indexes* allow you to search through multiple indices.) Google.com, AOL.com, and EarthLink.com are all search sites. Dogpile.com and Mamma.com are meta-index search sites.

✓ **Search system:** This organization possesses a combination of software, hardware, and people that indexes or categorizes Web sites — the system builds the index or directory you search at a search site. The distinction is important because a search site might not actually own a search index or directory. For instance, Google is a search system — it displays results from the index that it creates for itself — but AOL.com and EarthLink.com aren't. In fact, if you search at AOL.com or EarthLink.com, you actually get Google search results.

Google and the Open Directory Project provide search results to hundreds of search sites. In fact, most of the world's search sites get their search results from elsewhere (mostly Google these days); see Figure 1-3.

✓ **Search term:** This is the word, or words, that someone types into a search engine when looking for information.

✓ **Search results:** Results are the information (the results of your search term) returned to you when you go to a search site and search for something. As just explained, in many cases the search results you see don't come from the search site you're using, but rather from some other search system.



Figure 1-3: Look carefully, and you'll see that many search sites get their search results from other search systems.

- ✓ **SERPs:** I don't use the term much, but you'll hear others in the business talking about the *serps*. It simply means *search engine results page*, the page that appears after you search.
- ✓ **Natural search results:** A link to a Web page can appear on a search results page two ways: The search engine may place it on the page because the site owner paid to be there (pay-per-click ads), or it may pull the page from its index because it thinks the page matches the search term well. These free placements are often known as *natural search results*; you'll also hear the term *organic search results* and sometimes even *algorithmic search results*.
- ✓ **Search engine optimization (SEO):** Search engine optimization (also known as *SEO*) refers to "optimizing" Web sites and Web pages to rank well in the search engines — the subject of this book, of course.

Why bother with search engines?

Why bother using search engines for your marketing? Because search engines represent the single most important source of new Web site visitors.

You may have heard that most Web site visits begin at a search engine. Well, this isn't true, though many people continue to use these outdated statistics because they sound good — "80 percent of all Web site visitors reach the site through a search engine," for instance. However, way back in 2003, that claim

was finally put to rest. The number of search-originated site visits dropped below the 50 percent mark. Most Web site visitors reach their destinations by either typing a *URL* — a Web address — into their browsers and going there directly or by clicking a link on another site that takes them there. Most visitors *don't* reach their destinations by starting at the search engines.

However, search engines are still *extremely important* for a number of reasons:

- ✔ The proportion of visits originating at search engines is still significant. Sure, it's not 80 percent, but with billions of searches each month, it's still a lot of traffic.
- ✔ According to a report by comScore, Internet users in the United States carried out 20 billion searches at major search engines in December 2011.
- ✔ Research shows that another 10 billion or more searches are carried out in other search sites, such as map sites (MapQuest), video sites (YouTube), retail sites (Amazon, eBay, Craigslist), and so on. It's likely that more than 25 billion searches are performed in the United States each month, 2 to 3 searches every day for every man, woman, child, and baby in the United States.
- ✔ Of the visits that don't originate at a search engine, a large proportion are revisits — people who know exactly where they want to go. This isn't new business; it's repeat business. Most *new* visits come through the search engines — that is, search engines are the single most important source of *new* visitors to Web sites.
- ✔ It's also been well established for a number of years that most people researching a purchase begin their research at the search engines. (Except for those who don't. As I discuss in Chapter 14, many, perhaps most, product searches actually begin in sites such as Amazon, eBay, and Craigslist. But then, I think it's important to understand that these sites are search engines; they are, in effect, product-search engines.)
- ✔ Search engines represent an inexpensive way to reach people. Generally, you get more bang for your buck going after free search-engine traffic than almost any other form of advertising or marketing.

Here's an example. One client of mine, selling construction equipment to the tune of \$10,000 a month, rebuilt his site and began a combined natural-search and paid-search campaign, boosting sales to around \$500,000 a month in less than two years. It's hard to imagine how he could have grown his company, with relatively little investment, so quickly without the search engines!

Where Do People Search?

You can search for Web sites at many places. Literally thousands of sites, in fact, provide the ability to search the Web. (What you may not realize, however, is that many sites search only a small subset of the World Wide Web.)

However, *most* searches are carried out at a small number of search sites. How do the world's most popular search sites rank? That depends on how you measure popularity:

- ✓ Percentage of site visitors (*audience reach*)
- ✓ Total number of visitors
- ✓ Total number of searches carried out at a site
- ✓ Total number of hours visitors spend searching at the site

Each measurement provides a slightly different ranking. Although all provide a similar picture with the same sites generally appearing on the list, some search sites are in slightly different positions.

The following list shows the United States' top general search sites in November 2011 — around 20 billion searches — according to comScore:

Google sites: 65.4 percent

Yahoo! sites: 15.1 percent

Microsoft sites (Bing): 15 percent

Ask Network: 2.9 percent

AOL, Inc.: 1.6 percent

Remember that this is a list of search sites, not search systems. In fact, the preceding list shows *groups* of sites — the Microsoft entry, for instance, includes Bing.com, MSN.com, MSNBC.com, Microsoft.com, and so on.

In some cases, the sites own their own systems. Google provides its own search results, for instance, but AOL doesn't. (AOL gets its results from Google.) Yahoo! gets its results from Bing, thanks to a Yahoo!/Microsoft partnership — known as the *Yahoo! and Microsoft Search Alliance* — that was implemented in August 2010. (Look for the little *Powered by Bing* notice at the bottom of Yahoo! search pages.)

The fact that some sites get results from other search systems means two things:

- ✓ **The numbers in the preceding list are somewhat misleading.** They suggest that Google has 65.4 percent of all searches. But Google also feeds AOL its results — add AOL's searches to Google's, and you have 67 percent of all searches. Additionally, Google feeds search results to various other sites; comScore estimates Google's true share as 67.6 percent of all searches. Microsoft feeds not 15 percent of results but, when you add in the Yahoo! searches, powered by Microsoft, actually more like 30 percent.
- ✓ **You can ignore some of these systems.** At present, and for the foreseeable future, you don't need to worry about AOL.com. Even though it's one of the world's top search sites, as long as you remember that Google

feeds AOL, you need to worry about Google only. You don't really need to worry about Yahoo!, either; as long as Bing feeds Yahoo!, you can think of the two as essentially the same index.

Now reexamine the preceding list of the U.S.'s most important search sites and see what you can remove to get closer to a list of sites you care about. Check out Table 1-1 for the details.

<i>Search Site</i>	<i>Keep It On the List?</i>	<i>Description</i>
Google.com	Yes	The big kid on the block. Lots of people search the Google index on its own search site, <i>and</i> it feeds many sites. Obviously, Google has to stay on the list.
Yahoo.com	No	Yahoo! is obviously a large, important site, but it gets its search results from Bing, so as long as you're in the Bing index, you're in Yahoo!. (Yahoo! <i>does</i> have an important search <i>directory</i> , though, which I discuss in Chapter 13, but relatively few searches go through the directory.)
Bing	Yes	Bing creates its own index, gets many searches, <i>and</i> it feeds data to Yahoo!. So Bing is critical.
AOL.com	No	Fuggetaboutit — AOL gets search results from Google.
Ask.com (previously known as AskJeeves.com)	Yes	It has its own search engine and feeds some other systems — MyWay, Lycos, and Excite. Keep it in mind, though it's small and relatively unimportant compared to Google and Bing.

Based on the information in Table 1-1, you can whittle down your list of systems to three: Google, Bing, and Ask. The top two search systems are very important, accounting for 95 percent or more of all search results, with a small follower, Ask, which provides results to many smaller search sites. There's one more system I'm tempted to add to these three systems, though: the Open Directory Project (www.dmoz.org). This directory system feeds data to hundreds of search sites, including Google and AOL, so if you can get listed in here, it's a great thing, and, in fact, in earlier editions of this book, I *have* included it. However, whether you actually *can* get listed these days is another matter, so I'm going to leave it off the list, though I look at it in more detail in Chapter 13.



To summarize, three important systems are left:

- ✓ Google
- ✓ Bing
- ✓ Ask

That's not so bad, is it? You've just gone from thousands of sites to three, and only the top two are critical. (The only reason Ask.com gets included on such lists is that even though it has a tiny share of the search market, there's nothing *below* it on the list that comes close.)

Now, some of you may be thinking, "Aren't you missing some sites? What happened to HotBot, Mamma.com, WebCrawler, Lycos, and all the other systems that were so well known a few years ago?" A lot of them have disappeared or have turned over a new leaf and are pursuing other opportunities.

For example, Northern Light, a system well known in the late 1990s, now sells search software. And in the cases in which the search sites are still running, they're generally fed by other search systems. InfoSpace, Dogpile, and MetaCrawler, for instance, all get search results from Google, Yahoo!, and Bing, which means, in effect, from Google and Bing.

AltaVista, the first big search index, has been owned by Yahoo! for years, but currently displays Yahoo! search results, Powered by Bing. The same goes for AllTheWeb (for the geeks among you who remember it) — more Powered-By-Bing Yahoo! search results. If the search site you remember isn't mentioned here, it's either out of business, being fed by someone else, or simply not important in the big scheme of things.

When you find a new search system, look carefully on the page near the search box, or on the search results page — perhaps at the bottom of the page in the copyright message — and you may find where the search results are coming from.

You'll also want to work with some other search systems, as you find out in Chapter 13. In some cases, you need to check out specialty directories and indexes related to the industry in which your Web site operates or submit your site to Web directories in order to build links back to your site. In addition, in Chapter 14, you find out about the product search sites — hugely important for those of you selling products. And in Chapter 19, I tell you about the video sites — YouTube, for instance, is the world's third most important search engine, after Google and Bing. However, the preceding systems — Google, Bing, and Ask.com — are the most important general-search systems. And again, only the first two are really critical.



Google alone provides almost 70 percent of all search results. Get into *all* three systems on the preceding list, and you're in front of probably more than 99 percent of all searchers. Well, perhaps you're in front of them. You have a chance of being in front of them, anyway, if your site ranks highly (which is what this book is all about).

Search Engine Magic

Go to Google and search for the term *personal injury lawyer*. Then look at the blue bar below the Google logo, and you see something like this:

About 30,600,000 results (0.29 seconds)

This means Google has found over 30 million pages that it believes match these three words in some way. Yet, somehow, Google has managed to rank the pages. It's decided that one particular page should appear first, and then another, and then another, and so on, all the way down to page 30,600,000. (By the way, this has to be one of the wonders of the modern world: Search engines have tens of thousands of computers, evaluating a trillion pages or more, in a fraction of a second.)

How do they do it?

How on earth does Google do it? How does it evaluate and compare pages? How do other search engines do the same? Well, I don't know *exactly*. Search engines don't want you to know how they work (or it would be too easy to create pages that exactly match the criteria of the search system for any given search term, "giving them what they want to see"). But I can explain the general concept.

When Google searches for your search term, it begins by looking for pages containing the exact phrase. Then it starts looking for pages containing the words close together, and for synonyms; search for *dog* and Google knows you may be interested in pages with the word *canine*, for instance. (One Google source claims that synonyms come into play in around 70 percent of all searches.) Then it looks for pages that have the words scattered around. This isn't necessarily the order in which a search engine shows you pages; in some cases, pages with words close together (but not the exact phrase) appear higher than pages with the exact phrase, for instance. That's because search engines evaluate pages according to a variety of criteria.

Search engines look at many factors. They look for the words throughout the page, both in the visible page and in the nonvisible portions of the HTML source code for the page. Each time they find the words, they are *weighted* in some way. A word in one position is worth more than a word in another position. A word formatted in one way is worth more than a word formatted in another. (You read more about this in Chapter 6.) There's more, though. Search engines also look at links pointing to pages and use those links to evaluate the referenced pages: How many links are there? How many are from popular sites? What words are in the link text? You read more about this in Chapters 15 through 17.

Stepping into the programmers' shoes

There's a lot of conflicting information out there about SEO. Some of it's good, some of it's not so good, and some of it's downright wrong. When evaluating a claim about what search engines do, I sometimes find it useful to step into the shoes of the people building the search engines; I try to think about what would make sense from the perspective of the programmers who write the code that evaluates all these pages.

Consider this: Say, you search for *personal injury lawyer*, and the search engine finds one page with the term in the page's title (between the <TITLE> and </TITLE> tags, which you read more about in Chapters 3 and 7), and another page with the term somewhere deep in the page text. Which do you think is likely to match the search term better? If the text is in the title, doesn't that indicate that page is likely to be related in some way to the term? If the text is deep in the body of the page, couldn't it mean that the page isn't directly related to the term, but that it's related to it in some incidental or peripheral manner?



Considering SEO from this point of view makes it easier to understand how search engines try to evaluate and compare pages. If the keywords are in the links that point to the page, the page is likely to be relevant to those keywords; if the keywords are in headings on the page, that must be significant; if the keywords appear frequently throughout the page, rather than just once, that must mean something. Suddenly, it all makes sense.

By the way, in Chapter 8, I discuss things that search engines don't like. You may hear elsewhere all sorts of warnings that may or may not be correct. Here's an example: I've read that using a refresh meta tag to automatically push a visitor from one page to another will get your site penalized and may even get your site banned from the search engine. You've seen this situation: You land on a page on a Web site, and there's a message saying something like, "We'll forward you to page *x* in five seconds, or you can click [here](#)." The theory is that search engines don't like this, and they may punish you for doing this.

Now, does this make any sense? Aren't there good reasons to sometimes use such forwarding techniques? Yes, there are. So why would search engines punish you for doing it? They don't. They probably won't index the page that is forwarding a visitor — based on the quite reasonable theory that if the site doesn't want the visitor to read the page, the search engine doesn't need to index it — but you're not going to get punished for using it.

Remember that the search engine programmers aren't interested in punishing anyone; they're just trying to make the best choices between billions of pages. Generally, search engines use their "algorithms" to determine how to rank a page, and they try to adjust the algorithms to make sure "tricks" are ignored. But they don't want to punish anyone for doing something for which there might be a good reason, even if the technique could also be used as a trick.

What would the programmers do? I like to use this as my “plausibility filter” when I hear someone make some unusual or even outlandish claim about how search engines function.

Gathering Your Tools

You need several tools and skills to optimize and rank your Web site. I talk about a number of these in the appropriate chapters, but I want to cover a few basics before I move on. It goes without saying that you need:

- ✓ Basic Internet knowledge
- ✓ A computer connected to the Internet
- ✓ A Web site
- ✓ One of these two things:
 - Good working knowledge of HTML
 - Access to a geek with a good working knowledge of HTML

Which path should you take? If you don't know what HTML means (HyperText Markup Language), you probably need to run out and find that geek. *HTML* is the code used to create Web pages, and you need to understand how to use it to optimize pages. Discussing HTML and how to upload pages to a Web site is beyond the scope of this book. If you're interested in finding out more, check out *HTML, XHTML, & CSS For Dummies*, by Ed Tittel and Jeff Noble, and *Creating Web Pages For Dummies*, 9th Edition, by Bud E. Smith (both published by John Wiley & Sons, Inc.).

- ✓ Web browser and SEO tools

In the past, I've recommended using Firefox, because of the plethora of SEO-related tools that can be installed into the browser. But all of the Big Three (Internet Explorer, Firefox, and Chrome) have a bunch of SEO-related tools now, and even the next two browsers on the popularity list (Safari and Opera) have some, too, though probably not as many. Look in your browser's add-on library for tools such as these:

- *NoFollow*: Lots of tools indicate the presence of “nofollow” links (see Chapter 15).
- *Whois*: These tools retrieve information about the domain of the site you're viewing. Great for digging up info on competitors.
- *Firebug*: A fantastic little tool for examining the code underlying a Web page. Right-click a component on the page you're looking at, select Inspect Element, and you see a frame that shows you how the component was created. Designed for Firefox, but with a “lite” version that works in other browsers.

- *Google Global*: Handy if you want to see Google search results in different countries.
- *Compete Browser Extension*: Provides information, in the status bar, about the popularity of the site you are visiting, from Compete.com. (Alexa and Quantcast are two other well-known page-popularity services.)
- *PageRank*: Various tools display the Google PageRank of the page currently displayed in the browser (see Chapter 15).
- *SEO plug-ins*: Search the add-on library for the term *SEO*, and you'll find a number of add-ons that are collections of tools that provide access to all sorts of data. For instance, the WebRank Toolbar shows Google PageRank, along with Alexa, Compete, and Quantcast rankings. SEOQuake provides all sorts of things, such as the number of pages on the displayed Web site that are indexed by Google and Bing, the number of links pointing to the site according to those search engines, a link to Whois information, a link to a list of similar sites, Alexa rank and PageRank, and so on.



A quick tip for Firefox users; don't upgrade as soon as there's a new version. The Firefox team tends to release new versions quicker than the add-on authors can keep up, so if you upgrade too soon, you'll find most of your add-ons are disabled.

Geek or no geek

Many readers of this book are business people who don't plan to do the search engine work themselves (or, in some cases, realize that it's a lot of work and need to find someone with more time or technical skills to do the work). However, having read the book, they understand far more about search engines and are

in a better position to find and direct someone else working on their site. As one reader-cum-client told me, "There's a lot of snake oil in this business," so his reading helped him understand the basics and ask the right questions of search engine optimization firms. (See Chapter 21 for more information on that subject.)