

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 four 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 MSN — but you've probably also heard of others: HotBot, Dogpile, Ask.com, Netscape, EarthLink — even Amazon provides its A9 Web search box on many pages. There's also Lycos, InfoSpace, Mamma.com, and WebCrawler.

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 farther, 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, closely followed by Yahoo! and MSN Live Search.

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 — they reached 15 billion or so at one point — but decided not to play the “mine is bigger than yours” game with Yahoo!.

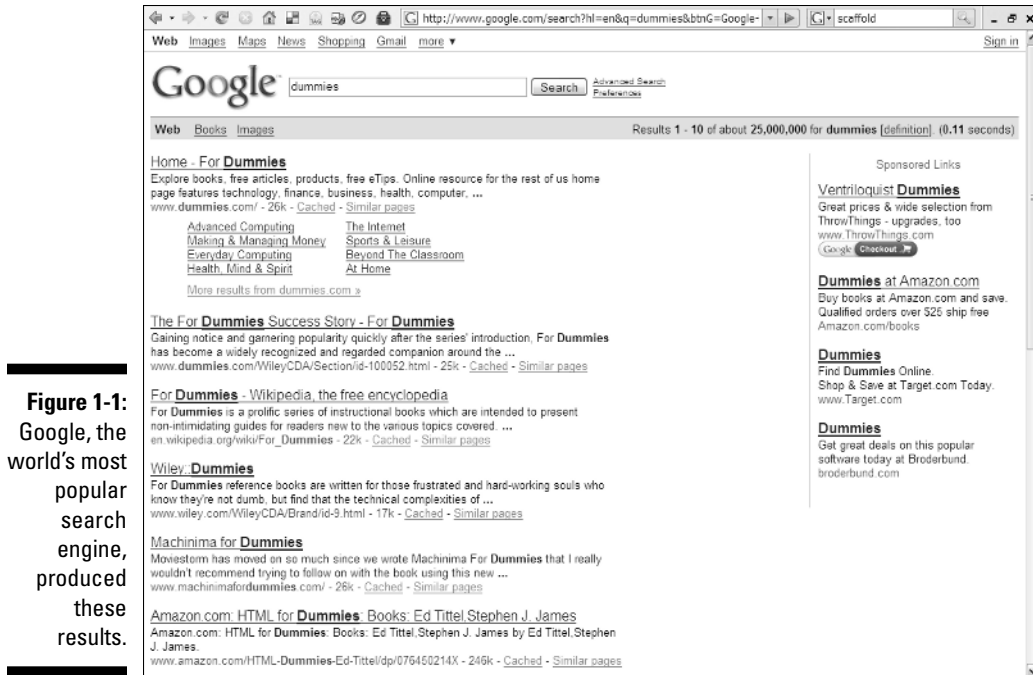


Figure 1-1:
Google, the world's most popular search engine, produced these results.

Search directories

A *search directory* is a categorized collection of information *about* Web sites instead of containing information *from* Web pages.

The most significant search directories are owned by Yahoo! (`dir.yahoo.com`) and the Open Directory Project (`www.dmoz.org`). (You can see an example of Open Directory Project information, displayed in Google — `dir.google.com` — in Figure 1-2.) Directory companies don't use 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 and description, 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.



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.**

Figure 1-2:
Google also has a search directory, but it doesn't create the directory itself; it gets it from the Open Directory Project.



3. Click the Search button.

The list of Web sites that appears is called the Yahoo! Search results.

4. Look for Directory above the Search box.

Above the Search box you'll 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'll end up in Yahoo! directory. (You can also go directly to the directory by using `dir.yahoo.com`.)

Non-spidered indexes

I wasn't sure what to call these things, so I made up a name: *non-spidered 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 2.) In other cases, the person who enters the site into the index provides this information. A number of the smaller systems discussed in Chapter 12 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 more detail in Chapter 18.



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 index or categorize Web sites — they build the index or directory you search through 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; see Figure 1-3.

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



- ✔ **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 from some other search system.
- ✔ **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, 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, published in March of 2008, Internet users in the United States carry out almost 10 billion searches a month at the major search engines.

- ✓ comScore also reported that in August of 2007, 750 million people around the world carried out a total of 61 billion searches on the world's top search sites; that's almost three searches every day by 10 percent of the world's population.
- ✓ 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.
- ✓ 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 \$700,000 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 are in slightly different positions.

The following list runs down the world's most popular search sites, based on one month of searches during August of 2007 — almost 8 billion searches — according to a Nielsen/NetRatings study. These statistics are for U.S. Internet users:

Google.com	53.6%
Yahoo.com	19.9%
MSN.com	12.9%
AOL.com	5.6%
Ask.com	1.7%
My Web Search	0.9%
BellSouth Search	0.5%
Comcast Search	0.4%
My Way	0.4%
SBC Yellow Pages Search	0.4%

Remember: This is a list of search sites, not search systems. In some cases, the sites own their own systems. Google provides its own search results, but AOL doesn't. (AOL gets its results from Google.) My Web Search is a meta-search engine getting results from Google, Yahoo!, and Ask.com.

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 around 54 percent of all searches. But Google also feeds AOL its results — add AOL's searches to Google's, and you have 59.6 percent of all searches. Additionally, Google feeds Comcast and BellSouth (another 0.9 percent combined, according to NetRatings). Furthermore, My Web Search is a meta-search engine; therefore, if you search at My Web Search, you see results from Google, Yahoo!, MSN Live Search, and Ask.
- ✔ **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, you can forget about it. Sure, keep it in the back of your mind, but as long as you remember that Google feeds AOL, you need to worry about Google only.

Now reexamine the preceding list of the world'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.

Table 1-1 The Top Search Sites		
<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	Yes	Yahoo! is obviously a large, important site; keep it.
MSN.com (MSN Live Search)	Yes	Ditto; MSN Live Search creates its own index, and gets many searches.
AOL.com	No	Fuggetaboutit — AOL gets search results from Google (although it manipulates their appearance) and from the Open Directory Project.
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, though it's small.
MyWebSearch.com	No	This system simply searches through other systems' search indexes (Google, Yahoo!, and Ask). Forget it.
BellSouth Search	No	Results come from Google.
Comcast Search	No	Again, results come from Google, so forget it.
MyWay.com	No	MyWay uses data from Ask, so forget about it.
SBC Yellow Pages	No	These results come from the Yellow Pages, something we talk about in Chapter 13. For now, we can forget about this.

Based on the information in Table 1-1, you can whittle down your list of systems to four: Google, Yahoo!, MSN Live Search, and Ask. The top three search systems are all very important, with a small follower, Ask, which provides results to many smaller search sites. There's one more system I want to add to these four systems, though. Very few people search at the Open Directory Project (www.dmoz.org). However, this directory system feeds data to hundreds of search sites, including Google and AOL.



To summarize, five important systems are left:

- ✓ Google
- ✓ Yahoo!
- ✓ MSN Live Search
- ✓ Ask
- ✓ Open Directory Project

That's not so bad, is it? You've just gone from thousands of sites to five. Note, by the way, that the top three positions may shift around a little. Google had already lost a large proportion of its share (when I wrote the first edition of this book, Google had around three-quarters of the market— now it's probably a little over one-half), and then regained about 8 percent of the market between the second and third editions of this book.

Of course we may soon see some real changes in this profile. Microsoft is trying to buy Yahoo! If it succeeds, it seems likely that the two search indexes will be merged at some point. (After all, the main point of the proposed merger is to share — and thus cut — costs.)

Then there's Ask.com, which recently announced that it was no longer going head to head with Google — years of trying to gain market share got them nowhere — and was planning to focus on the search needs of married women. Perhaps more significantly, there are rumors that Ask may stop managing its own search index — and use Google's! Ask already uses Google's Pay Per Click ads (see Chapter 18), so extending its partnership with Google wouldn't be such a surprise. So, who knows, we may soon see just two major search indexes: the Yahoo!/MSN index and the Google/Ask/and-everything-else index.

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 where the search sites are still running, they're generally fed by other search systems. Mamma.com, Dogpile, and MetaCrawler get search results from the top four systems, for instance, and HotBot gets results from MSN Live Search and LyGo, a new service from Lycos.

AltaVista, the first big search index, is now owned by Yahoo! and is really just a different Web design displaying Yahoo! search results. The same goes for AllTheWeb, for the geeks among you who remember it. 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 Chapters 12 and 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. But the preceding systems are the important ones for every Web site.



Google alone provides well over 60 percent of all search results (down from 75 percent just a year or two ago). Get into *all* the systems on the preceding list, and you're in front of probably more than 96 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:

```
Results 1 - 10 of about 19,800,000 for personal injury
lawyer
```

This means Google has found almost 20 million pages that contain these three words. Yet, somehow, it 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 19,800,000. (By the way, this has to be one of the wonders of the modern world: Search engines have tens of thousands of computers, evaluating 20 billion pages or more, and returning the information 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 search system, "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. Then it looks for pages that have the pages 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 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 5.)

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 14 and 15.

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, "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 2 and 6), 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, in some way, related 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, 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 pointing to the page, the page must 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 7, 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 a search engine 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, they don'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 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 4 For Dummies*, 5th Edition, by Ed Tittel and Mary Burmeister, and *Creating Web Pages For Dummies*, 8th Edition, by Bud E. Smith and Arthur Bebak (both published by Wiley Publishing, Inc.).

✔ Toolbars

Install the Google toolbar in your Web browser . . . and perhaps the Yahoo!, MSN Live Search, and Ask.com toolbars, too. And, maybe, the Alexa toolbar. (Before you complain about spyware, I explain in a few moments.) You may want to use these tools even if you plan to use a geek to work on your site. They're simple to install and open up a completely new view of the Web. The next two sections spell out the details.

Search toolbars

I definitely recommend the Google toolbar, which allows you to begin searching Google without going to the Web site first (if you use the Firefox Web browser, the built-in Search Engine box does the same). You might also want to use the Yahoo!, MSN Live Search, and Ask.com toolbars. Additionally, these toolbars have plenty of extras: auto form fillers, tabbed browsing, desktop search, spyware blockers, translators, spell checkers, and so on. I'm not going to describe all these tools, as they aren't directly related to SEO, but they're definitely useful.

You really don't need all of them, but hey, here they are if you really want to experiment. You can find these toolbars here:

✔ toolbar.google.com

✔ toolbar.yahoo.com

✔ toolbar.msn.com

✔ toolbar.ask.com

Unfortunately, all these toolbars require Microsoft Windows and the Internet Explorer browser or Firefox browser (the MSN toolbar only works in Internet Explorer); that's most of you but, I realize, not all. You can see the Google and Yahoo! toolbars, along with the Alexa toolbar (at the bottom left of the browser window), in Figure 1-4. Don't worry; you don't have to have all this clutter on your screen all the time. Right-click a blank space on any toolbar, and you can add and remove toolbars temporarily; simply open a toolbar when you need it. I leave the Google toolbar on all the time, and open the others now and then.

Figure 1-4: The toolbars provide useful information for search engine campaigns; the Google bar's PageRank button is shown open.



One thing I do find frustrating about these systems is the pop-up blockers. Yes, they can be helpful, but often they block pop-up windows that I want to see; if you find that you click a link and it doesn't open, try Ctrl+clicking (which may temporarily disable the pop-up blocker), or disable the blocker on the toolbar.

I refer to the Google toolbar here and there throughout this book because it provides you with the following useful features:

- ✓ A way to search Google without going to `www.google.com` first
- ✓ A quick view of the Google PageRank, an important metric that I explain in Chapter 14
- ✓ A quick way to see if a Web page is already indexed by Google
- ✓ A quick way to see some of the pages linking to a Web page



Geek or no geek

Many readers of this book's first edition 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. 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.

The Google toolbar has a number of other useful features, but the preceding features are the most useful for the purposes of this book. Turn on the PageRank button after installing the toolbar:

1. **Click the Settings button, and then choose Options from the menu that appears.**
2. **Click the More tab in the Toolbar Options dialog box.**
3. **Enable the PageRank and Page Info check box and click OK.**

Alexa toolbar

Alexa is a company owned by Amazon.com, and a partner with Google and Microsoft. It's been around a long time, and millions of people around the world use it. Every time someone uses the toolbar — Sparky, as it's now known — to visit a Web site, the toolbar sends the URL to Alexa, allowing the system to create an enormous database of site visits. The toolbar can provide traffic information to you; you can quickly see how popular a site is and even view a detailed traffic analysis, such as an estimate of the percentage of Internet users who visit the site each month.

Work with the Alexa toolbar for a while, and you'll quickly get a feel for site popularity. A site ranks 453? That's pretty good. 1,987,123? That's a sign that hardly anyone visits the site. In addition, it provides a quick way to find information about who owns the site on which the current page sits, and how many pages link to the current page. You can find the Alexa toolbar (refer to Figure 1-4), at toolbar.alexa.com.

I've been criticized for recommending the Alexa toolbar in earlier editions of this book: Some people claim it is spyware. Some anti-spyware programs even search for the toolbar and flag it as spyware, though others don't. As I mention, the toolbar sends the URLs you're visiting. However, Alexa states on the site (and I believe them) that "The Alexa Toolbar contains no advertising and does not profile or target you." I know for sure that it doesn't display ads. Alexa doesn't steal your usernames and passwords, as is occasionally claimed. Alexa does gather information about where you visit, but it doesn't know who you are, so does it matter? Decide for yourself.