The Search Engines: Reflecting Consciousness and Connecting Commerce

SEARCH HAS BECOME INTEGRATED INTO THE FABRIC OF OUR SOCIETY. With more than 12 billion searches being performed each month as of January 2009 (according to comScore), approximately 400 million web searches are performed every day. This means that on average more than 4,500 searches are performed every single second of every day.

As Google owns approximately 65% of the search market share, Google's search technology handles more than 2,900 searches per second. In addition, users have grown to expect that the responses to their search queries will be returned in less than one second.

Now people can obtain information in mere seconds—information that 20 years ago would have required a trip to the library, a search through the Dewey Decimal System, and a foot search through halls of printed volumes, a process that could easily have consumed two hours or more. Through the new channel of search, people can also conduct much of their shopping, banking, and social transactions online—something that has changed the way our global population lives and interacts.

This dramatic shift in behavior represents what investors like to label a *disruptive event*—an event that has changed something in a fundamental way. Search engines are at the center of this disruptive event, and having a business's website rank well in the search engines when people are looking for the service, product, or resource it provides is critical to the survival of that business. However, like most paths to success, obtaining such prime search result real

estate is not a simple matter, but it is one that this book aims to deconstruct and demystify as we examine, explain, and explore the ever-changing art of search engine optimization (SEO).

The Mission of Search Engines

Search engines generate revenue primarily through paid advertising. The great majority of this revenue comes from a pay-per-click (or cost-per-click) model, in which the advertisers pay only for users who click on their ads.

Since web searchers are free to use any of the many available search engines on the Web to find what they are seeking, the burden is on the search engines to develop a relevant, fast, and fresh search experience. For the most part, search engines accomplish this by being *perceived* as having the most relevant results and delivering them the fastest, as users will go to the search engine they think will get them the answers they want in the least amount of time.

As a result, search engines invest a tremendous amount of time, energy, and capital in improving their relevance. This includes performing extensive studies of user responses to their search results, comparing their results against those of other search engines, conducting eye-tracking studies (discussed later in this chapter), and conducting PR and marketing campaigns.

Because the search engines' success depends so greatly on the relevance of their search results, manipulations of search engine rankings that result in non-relevant results (generally referred to as *spam*) are dealt with very seriously. Each major search engine employs a team of people who focus solely on finding and eliminating spam from their search results. This matters to SEO practitioners because they need to be careful that the tactics they employ will not be seen as spam by the search engines and carry the risk of resulting in penalties for the websites they work on.

The Market Share of Search Engines

Figure 1-1 shows the U.S. market share for search engines in January 2009, according to comScore. As you can see, Google is the dominant search engine on the Web in the United States.

In many European countries, the disparity is even greater. However, in some markets Google is not dominant. In China, for instance, Baidu is the leading search engine. The result is that in most world markets, a heavy focus on SEO is a smart strategy for Google.

The Human Goals of Searching

The basic goal of a human searcher is to obtain information relevant to her inquiry. However, searcher inquiries can take many different forms. One of the most important elements to building an online marketing strategy for a website around SEO and search rankings is
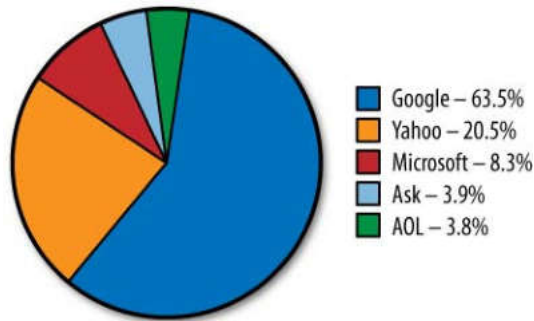
FIGURE 1-1. Search engine market share

developing a thorough understanding of the psychology of your target audience. Once you understand how the average searcher, and more specifically, your target market, uses search engines, you can more effectively reach and keep those users.

Search engine usage has evolved over the years, but the primary principles of conducting a search remain largely unchanged. The following steps comprise most search processes:

1. Experience the need for an answer, solution, or piece of information. For example, the user may be looking for a website (navigational query) to buy something (transactional query) or to learn something (informational query). We will discuss this in more detail in the following section.

2. Formulate that need in a string of words and phrases (the query). Most people formulate their queries in one to three words. ComScore data from March 2009 shows an average query length of 2.9 words. A more detailed look shows the following percentages of searches per word length (see Table 1-1).

TABLE 1-1. Searches by query length (comScore)

Words	Percent of searches
1	25.32%
2	24.96%
3	19.80%
4	13.17%
5	7.53%
6	4.04%
7	2.15%
8	1.19%

Data from Hitwise for this book shows a similar distribution of search query lengths (see Table 1-2).

TABLE 1-2. Searches by query length (Hitwise)

Percentage of U.S. clicks by number of keywords				
Subject	February 2008	January 2009	February 2009	Year-over-year % change
One word	21.04%	20.29%	20.48%	−3%
Two words	24.73%	23.65%	23.47%	−5%
Three words	21.84%	21.92%	21.68%	−1%
Four words	14.53%	14.89%	14.98%	3%
Five words	8.29%	8.68%	8.72%	5%
Six words	4.38%	4.65%	4.71%	8%
Seven words	2.29%	2.49%	2.51%	10%
Eight or more words	2.90%	3.43%	3.47%	20%
Note: data is based on four-week rolling periods (ending February 28, 2009; January 31, 2009; and March 1, 2008) from the Hitwise sample of 10 million U.S. Internet users.				
Source: Hitwise, an Experian company				

3. Execute the query, check the results, see whether you got what you wanted, and if not, try a refined query.

When this process results in the satisfactory completion of a task, a positive experience is created for the user, the search engine, and the site providing the information or result.

Who Searches and What Do They Search For?

ComScore reported that the number of search queries performed on the Web was approximately 12.6 billion across all engines in December 2008.

ComScore data shows that just under 79 million people in the United States were using a search engine on a given day in January 2009. Search engine users were slightly more likely to be women than men (50.4% versus 49.6%). According to comScore, Internet usage increases with household income, as per the data shown in Table 1-3 for January 2009.

TABLE 1-3. Internet users by household income

Household income	Internet users
$15,000–$24,999	5,792
$25,000–$39,999	16,108

Household income	Internet users
$40,000–$59,999	39,716
$60,000–$74,999	20,947
$75,000–$99,000	28,995
$100,000 or more	44,627

You can find additional data from studies, surveys, and white papers on Search Engine Land's Stats & Behaviors page.

All of this research data leads us to some important conclusions about web search and marketing through search engines. For example:

- Search is very, very popular. It reaches nearly every online American and billions of people around the world.
- Google is the dominant player in most world markets.
- Users tend to use short search phrases, but these are gradually getting longer.
- Search covers all types of markets.

Search is undoubtedly one of the best and most important ways to reach consumers and build a business, no matter the size, reach, or focus.

Determining Searcher Intent: A Challenge for Both Marketers and Search Engines

Good marketers are empathetic. Smart SEO practitioners and the search engines have a common goal of providing searchers with results that are relevant to their queries. Therefore, a crucial element to building an online marketing strategy around SEO and search rankings is to understand your audience. Once you grasp how your target market searches for your service, product, or resource, you can more effectively reach and keep those users.

Search engine marketers need to be aware that search engines are *tools*—resources driven by intent. Using the search box is fundamentally different from entering a URL into the address bar, clicking on a bookmark, or picking a link on your start page to go to a website; it is unique from a click on the "stumble" button in your StumbleUpon toolbar or a visit to your favorite blog. Searches are performed with *intent*; the user wants to find something in particular, rather than just land on it by happenstance.

What follows is an examination of the different types of queries, their categories, characteristics, and processes.

Navigational Queries

Navigational searches are performed with the intent of surfing directly to a specific website. In some cases, the user may not know the exact URL, and the search engine serves as the "White Pages." Figure 1-2 shows an example of a navigational query.

FIGURE 1-2. *Navigational query*

Opportunities: Pull searcher away from destination; get ancillary or investigatory traffic.

Average value: Generally low, with the exception of navigational searches on the publisher's own brand, where the value is very high as these types of searches tend to lead to very high conversion rates.

Informational Queries

Informational searches involve a huge range of queries—for example, local weather, maps and directions, details on the latest Hollywood awards ceremony, or just checking how long that trip to Mars really takes. Informational searches are primarily non-transaction-oriented (although they can include researching information about a product or service); the information itself is the goal and no interaction beyond clicking and reading is required. Figure 1-3 shows an example of an informational query.

Opportunities: Brand searchers with positive impressions of your site, information, company, and so on; attract inbound links; receive attention from journalists/researchers; potentially convert to sign up or purchase.

Average value: Middling. Note, though, that informational queries that are focused on researching commercial products or services can have high value.
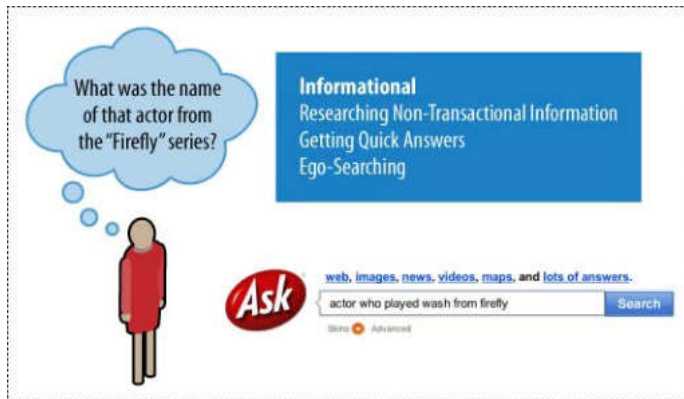
FIGURE 1-3. Informational query

Transactional Queries

Transactional searches don't necessarily involve a credit card or wire transfer. Signing up for a free trial account at CookIllustrated.com, creating a Gmail account, paying a parking ticket, or finding the best local Mexican cuisine for dinner tonight are all transactional queries. Figure 1-4 shows an example of a transactional query.

FIGURE 1-4. Transactional query

Opportunities: Achieve transaction (financial or other).

Average value: Very high.

Research by Pennsylvania State University and the Queensland University of Technology (*http://ist.psu.edu/faculty_pages/jjansen/academic/pubs/jansen_user_intent.pdf*) shows that more than

80% of searches are informational in nature and only about 10% of searches are navigational or transactional.

The researchers went further and developed an algorithm to automatically classify searches by query type. When they tested the algorithm, they found that it was able to correctly classify queries 74% of the time. The difficulty in classifying the remaining queries was vague user intent, that is, the query could have multiple meanings. Here are some URLs that point to additional academic research on this topic:

- *http://www.sigir.org/forum/F2002/broder.pdf*
- *http://www.strategynode.com/how-to-determine-if-a-search-query-is-navigational-informational-or-transactional*

When you are building keyword research charts for clients or on your own sites, it can be incredibly valuable to determine the intent of each of your primary keywords. Table 1-4 shows some examples.

TABLE 1-4. Sample search queries and intent

Term	Queries	Intent	$$ value
Beijing Airport	980	Nav	Low
Hotels in Xi'an	2,644	Info	Mid
7-Day China tour package	127	Trans	High
Sichuan jellyfish recipe	53	Info	Low

This type of analysis can help to determine where to place ads and where to concentrate content and links.

Hopefully, this data can help you to think carefully about how to serve different kinds of searchers, based on their individual intents, and then concentrate your efforts in the best possible areas.

Although informational queries are less likely to immediately convert into sales, this does not necessarily mean you should forego rankings on such queries. If you are able to build a relationship with users who find your site after an informational query, they may be more likely to come to you to make the related purchase at a later date.

One problem is that when most searchers frame their search query they provide very limited data to the search engine, usually in just one to three words. Since most people don't have a keen understanding of how search engines work, they can often provide a query that is too general or in a way that does not provide the search engine (or the marketer) with what it needs to determine their intent.

For this reason, general queries are important to most businesses because they often get the brand and site on the searcher's radar, and this initiates the process of building trust with the user. Over time, the user will move on to more specific searches that are more transactional or navigational in nature.

If, for instance, companies buying pay-per-click (PPC) search ads bought only the high-converting navigational and transactional terms and left the informational ones to competitors, they would lose market share to those competitors. During several days, a searcher may start with *digital cameras* and then hone in on *canon g10,* and buy from the store that showed up for digital cameras and pointed her in the direction of the Canon G10 model.

Given the general nature of how query sessions start, though, determining intent is quite difficult, and can result in searches being performed where the user does not find what she wants, even after multiple tries. An August 2007 Foresee/ACSI Report for eMarketer found that 75% of search engine and portal users were satisfied with their experiences. In a breakdown by property, 79% of Yahoo! users, 78% of Google users, and 75% of both Live Search (Microsoft's web search property, which has since been renamed to Bing) and Ask.com users reported being satisfied.

Based on this later study, more than 20% of users did not find what they were looking for. This suggests that there is plenty of room to improve the overall search experience. As an SEO practitioner, you should be aware that many of the visitors that you succeed in attracting to your site may have arrived for the wrong reasons (i.e., they were really looking for something else), and these visitors are not likely to help your business goals.

How People Search

Search engines invest significant resources into understanding how people use search, enabling them to produce better (i.e., faster, fresher, and more relevant) search engine results. For website publishers, the information regarding how people use search can be used to help improve the usability of the site as well as search engine compatibility.

Data from comScore provides some great insight into what people actually search for when they perform a search. Table 1-5 shows a breakdown of many of the major categories that people's Internet searches fall into, based on comScore data for October 2008.

TABLE 1-5. *Searches by market segment*

Parent category name	Percent of total searches
Directories/Resources	16.60%
Retail	11.86%
Entertainment	11.54%
Services	6.63%
Education	4.59%
Conversational Media	4.04%
Government	3.87%
Health	3.38%
Games	3.26%
News/Information	3.06%
Hobbies/Lifestyle	3.05%
Business/Finance	2.94%
Travel	2.21%
Community	1.94%
Regional/Local	1.87%
Sports	1.78%
Technology	1.73%
Automotive	1.67%
Real Estate	1.43%
Career Services and Development	1.12%
Telecommunications	0.78%
Auctions	0.57%
Portals	0.56%
ISP	0.38%
Gambling	0.27%
Business to Business	0.25%

This shows that people search across a very wide number of categories. Search engines are used to find information in nearly every portion of our lives. In addition, user interactions with

search engines can be a multistep process. Witness the user search session documented by Microsoft and shown in Figure 1-5.

Inside a real query "session"

Example decision: Which shoes to buy?
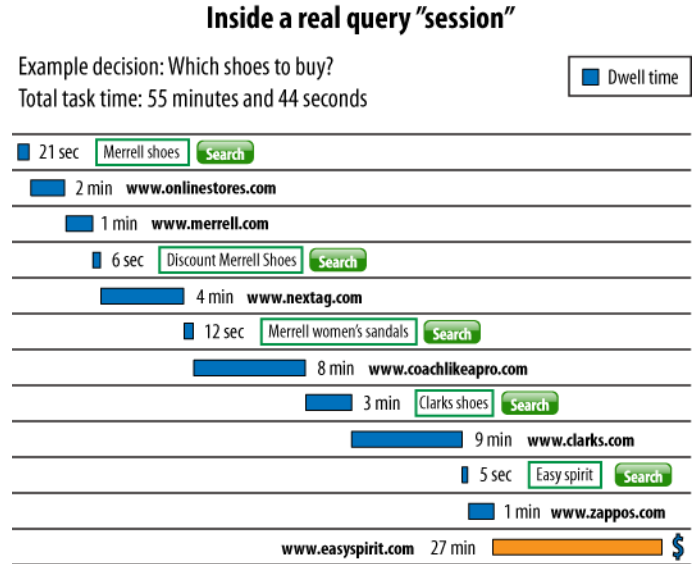Total task time: 55 minutes and 44 seconds

FIGURE 1-5. Merrell shoes user search session

In this sequence, the user performs five searches over a 55+ minute period before making a final selection. The user is clearly trying to solve a problem and works at it in a persistent fashion until the task is done.

However, it is increasingly common that search sessions of this type can take place over days. A 2007 study of e-commerce sites by ScanAlert showed that 30% of online transactions occurred more than 24 hours after the initial search (*http://searchenginewatch.com/3626363*).

This means people are thinking about their tasks in stages. As in our Merrell shoes example in Figure 1-5, people frequently begin with a general term and gradually get more specific as they get closer to their goal. They may also try different flavors of general terms. In Figure 1-5, it looks like the user did not find what she wanted when she searched on *Merrell shoes*, so she then tried *discount Merrell shoes*. You can then see her refine her search in the process, until she finally settles on Easy Spirit as the type of shoe she wants.

This is just one example of a search sequence, and the variety is endless. Figure 1-6 shows another search session, once again provided courtesy of Microsoft.

In this search session, the user has a health concern. This particular user starts with a five-word search, which suggests that she may have some experience using search engines. At 3:01 her search on *headache pregnant 3rd trimester* leads her to Answers.yahoo.com. After visiting this site, her search suddenly gets more specific.

Typical session: health

10 unique queries
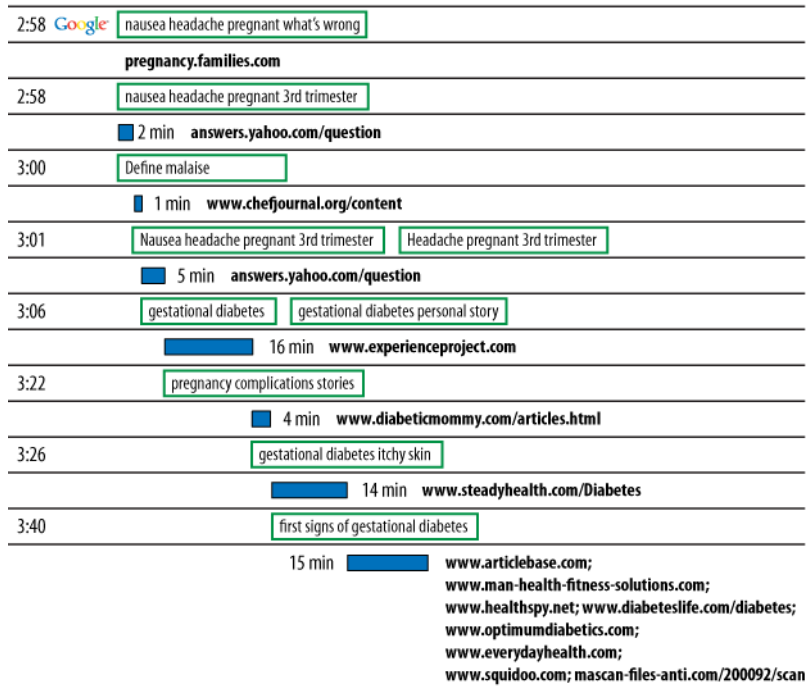7 partial re-queries and refinements
57 minutes

2:58 Google	nausea headache pregnant what's wrong	
	pregnancy.families.com	
2:58	nausea headache pregnant 3rd trimester	
	▪ 2 min **answers.yahoo.com/question**	
3:00	Define malaise	
	▌ 1 min **www.chefjournal.org/content**	
3:01	Nausea headache pregnant 3rd trimester	Headache pregnant 3rd trimester
	▪ 5 min **answers.yahoo.com/question**	
3:06	gestational diabetes	gestational diabetes personal story
	▬ 16 min **www.experienceproject.com**	
3:22	pregnancy complications stories	
	▪ 4 min **www.diabeticmommy.com/articles.html**	
3:26	gestational diabetes itchy skin	
	▬ 14 min **www.steadyhealth.com/Diabetes**	
3:40	first signs of gestational diabetes	

15 min ▬ **www.articlebase.com;**
www.man-health-fitness-solutions.com;
www.healthspy.net; www.diabeteslife.com/diabetes;
www.optimumdiabetics.com;
www.everydayhealth.com;
www.squidoo.com; mascan-files-anti.com/200092/scan

FIGURE 1-6. Health user search session

She begins to focus on gestational diabetes, perhaps because something she saw on Answers.yahoo.com led her to believe she may have it. The session culminates in a search for *first signs of gestational diabetes*, which suggests that she has concluded that this is quite possibly the issue she is facing.

The session stops there. It may be that at this point the user feels she has learned what she can. Perhaps her next step is to go to her doctor with her concerns, prepared to ask a number of questions based on what she learned.

Our next search session example begins with a navigational search, where the user simply wants to find the site for the travel website Orbitz (see Figure 1-7). The user's stay there is quite short, and she progresses to a search on *Cancun all inclusive vacation packages*. Following that she searches on a few specific resorts but finally settles on *cancun riviera maya hotels*, after which it appears she may have booked her hotel—the final site visited on that search is Bookings.occidentalhotels.com, and the direction of her searches changes after that.

Typical session: travel

11 unique queries
5 partial re-queries and refinements
33 minutes

Time	Engine	Query	Duration	URLs visited
19:14	Google	Orbitz travel		
			1 min	www.orbitz.com
19:15	Google	Cancun all inclusive vacation packages		
			14 min	cancun.bookit.com; www.travelocity.com; from bookit.com
19:29	Google	Allegro Playacar - All inclusive		
			2 min	from bookit.com; www.occidentalhotels.com
19:31	Google	Occidental Grand Xcaret - All inclusive resort		
			3 min	from bookit.com; search.bookit.com; www.occidentalhotels.com; from bookit.com
19:34	Google	NH Riviera Cancun - All inclusive		
			1 min	www.cancun.com/Hotels/NH-Riviera-Cancun; www.tripadvisor.com; www.travel-center.com
19:35	Google	cancun riviera maya hotels		
			7 min	www.nh-hotels.com; www.cancunandrivieramaya.com; www.nh-hotels.com; www.occidentalhotels.com; bookings.occidentalhotels.com
19:42	YAHOO!	Cancun theme park / Cancun theme park		
			1 min	www.cancun.com/Ecological_Parks/Xcaret
19:43	Live	xcaret		
			1 min	search.live.com/results; search.live.com/video
19:44	Live	xcaret		
			2 min	www.xcaret.com; www.xcaret.com/Gallery.aspx; bookings.occidentalhotels.com/travel/itinerary.cfm; www.xcaret.com/Tickets/Xcaret_Basic_Entrance.html
19:46	Live	xcaret video		
			1 min	search.live.com/video; www.orbitz.com/App/View; www.orbitz.com/App/ViewPackageDetails

FIGURE 1-7. Travel user search session

At that point, the user begins to look for things to do while she is in Cancun. She conducts a search for *cancun theme park* and then begins to look for information on *xcaret*, a well-known eco park in the area.

Users traverse through countless different scenarios when they are searching for something. Search engines do a lot of modeling of these scenarios to enable them to provide better results to users. The SEO practitioner can benefit from a basic understanding of searcher behavior as well. We will discuss this in more detail in Chapter 2.

How Search Engines Drive Commerce on the Web

People make use of search engines for a wide variety of purposes, with some of the most popular being to research, locate, and buy products. Even with the recession that (officially) hit the U.S. economy in late 2007, and which continues through the publication date of this book, e-commerce sales reported by the U.S. Census Bureau were a healthy $31.9 billion (*http://www.census.gov/mrts/www/data/pdf/08Q4.pdf*) through 2008.

It is important to note that search and offline behavior have a heavy degree of interaction, with search playing a growing role in driving offline sales. A Yahoo! study from 2007 (*http://searchengineland.com/yahoos-robo-study-search-has-big-impact-on-offline-purchases-11832*) showed the following:

- Online advertising drives $6 offline (in stores) for every $1 spent online.
- Search marketing has a greater impact on in-store sales lift than display advertising—three times greater, in fact.

There is also a significant amount of interaction between search and local offline commerce. WebVisible and Nielsen produced a 2007 report on local search (*http://searchengineland.com/survey-search-now-top-resource-for-local-information-12396*) that noted:

- 74% of respondents used search engines to find local business information versus 65% who turned to print Yellow Pages, 50% who used Internet Yellow Pages, and 44% who used traditional newspapers.
- 86% surveyed said they have used the Internet to find a local business, a rise from the 70% figure reported the year before.
- 80% reported researching a product or service online, then making that purchase offline from a local business.

Local search is an increasingly important component of SEO, and one which we will explore in detail in Chapter 2.

Eye Tracking: How Users Scan Results Pages

Research firms Enquiro, Eyetools, and Didit conducted heat-map testing with search engine users (*http://www.enquiro.com/research/eyetrackingreport.asp*) that produced fascinating results about what users see and focus on when engaged in search activity. Figure 1-8 depicts a heat map showing a test performed on Google. The graphic indicates that users spent the most amount of time focusing their eyes in the top-left area where shading is the darkest.

Published in November 2006, this particular study perfectly illustrates how little attention is paid to results lower on the page versus those higher up, and how users' eyes are drawn to bold keywords, titles, and descriptions in the natural ("organic") results versus the paid search listings, which receive comparatively little attention.
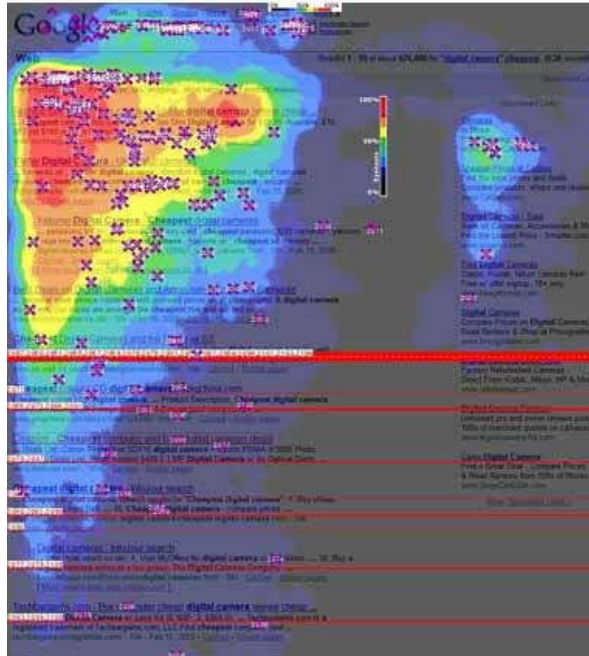
FIGURE 1-8. Enquiro eye-tracking results

This research study also showed that different physical positioning of on-screen search results resulted in different user eye-tracking patterns. When viewing a standard Google results page, users tended to create an "F-shaped" pattern with their eye movements, focusing first and longest on the upper-lefthand corner of the screen; moving down vertically through the first two or three results; moving across the page to the first paid page result; moving down another few vertical results; and then moving across again to the second paid result. (This study was done only on left-to-right language search results—Chinese, Hebrew, etc. results would be different.)

In May 2008, Google introduced the notion of Universal Search. This was a move from simply showing the 10 most relevant web pages (now referred to as "10 blue links") to showing other types of media, such as videos, images, news results, and so on, as part of the results in the base search engine. The other search engines followed suit within a few months, and the industry now refers to this general concept as *Blended Search*.

Blended Search, however, creates more of a chunking effect, where the chunks are around the various rich media objects, such as images or video. Understandably, users focus on the image first. Then they look at the text beside it to see whether it corresponds to the image or video thumbnail (which is shown initially as an image). Based on an updated study published

by Enquiro in September 2007, Figure 1-9 shows what the eye-tracking pattern on a Blended Search page looks like.

FIGURE 1-9. Enquiro eye-tracking results, Blended Search

Users' eyes then tend to move in shorter paths to the side, with the image rather than the upper-left-corner text as their anchor. Note, however, that this is the case only when the image is placed above the fold, so that the user can see it without having to scroll down on the page. Images below the fold do not influence initial search behavior until the searcher scrolls down.

This study is a vivid reminder of how important search engine results pages (SERPs) really are. And as the eye-tracking research demonstrates, "rich" or "personalized" search, as it evolves, will alter users' search patterns even more: there will be more items on the page for them to focus on, and more ways for them to remember and access the search listings. Search marketers need to be prepared for this as well.

Click Tracking: How Users Click on Results, Natural Versus Paid

By now, you should be convinced that you want to be on the top of the SERPs. It never hurts to be #1 in the natural search results.

In a bit of contrast to that, data shows that you may not want to be #1 in the paid search results, because the resulting cost to gain the #1 position in a PPC campaign can reduce the total net margin on your campaign. A study released by AdGooroo in June 2008 found that:

> Bidding for top positions usually makes financial sense only for high-budget, brand-name advertisers. Most other advertisers will find the optimal position for the majority of their keywords to lie between positions 5–7.

Of course, many advertisers may seek the #1 position in paid search results for a number of reasons. For example, if they have a really solid backend on their website and are able to make money when they are in the #1 position, they may well choose to pursue it. Nonetheless, the data from the survey suggests that there are many organizations for which being #1 in paid search does not make sense.

==Even if your natural ranking is #1, you can still increase the ranking page's click rate by having a sponsored ad above it or in the right column.== The AdGooroo survey showed that a prominent paid ad on the same search results page makes your #1 natural ranking receive 20% more clicks.

Distribution of Search Results and Traffic

To start breaking this down a bit, Figure 1-10 shows the screen real estate occupied by the two types of search results.
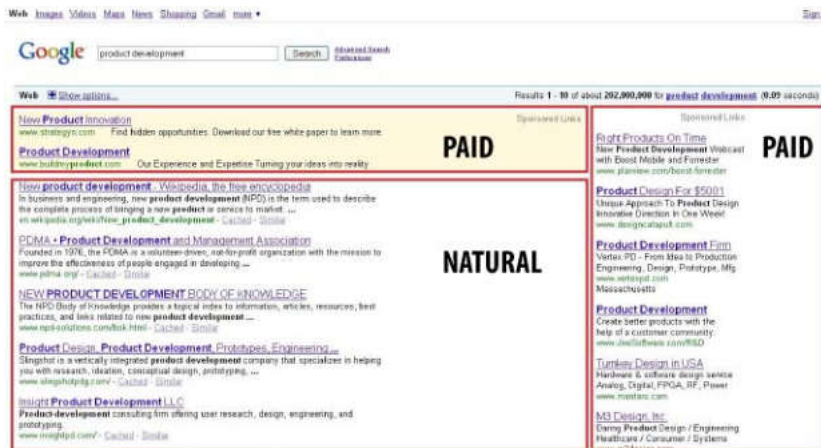

FIGURE 1-10. Paid and natural search results

This example from Google shows how the paid results appear above and to the right of the natural search results. Note that Google often does not show paid results above the natural results, in which case the paid results show up only on the right.

==Your position in the results has a huge impact on the traffic you will receive.== Here is some data about that:

- An April 2006 study by iProspect and Jupiter Research found that:
  — ==62% of search engine users click on a search result within the first page of results, and 90% within the first three pages.==
  — ==41% of search engine users who continue their search when not finding what they seek report changing their search term and/or search engine if they do not find what they're looking for on the first page of results; 88% report doing so after three pages.==
  — ==36% of users agree that "seeing a company listed among the top results on a search engine makes me think that the company is a top one within its field."==

- A study on data leaked from AOL's search query logs in August 2006 (*http://www.webuildpages.com/jim/click-rate-for-top-10-search-results/*) reveals the impact of position on the search results on the percentage of clicks received, as shown in Figure 1-11.

  In addition, the first 10 results received 89.71% of all click-through traffic; the next 10 results (normally listed on the second page of results) received 4.37%; the third page 2.42%; and the fifth page 1.07%. All other pages of results received less than 1% of total search traffic clicks.
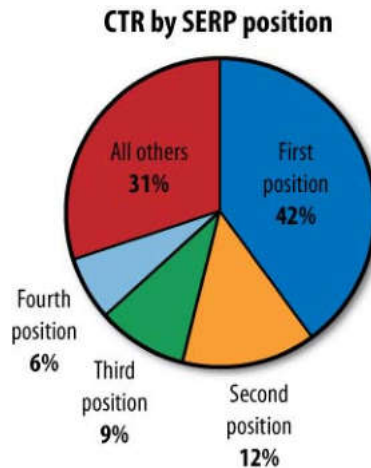
FIGURE 1-11. Click-through rate by SERP position

Why are searchers blind to relevant results farther down the page? Is this due to the "implied endorsement" effect, whereby searchers tend to simply trust the search engine to point them to the right thing? Or is it just the way humans are wired, to make snap decisions, as Malcolm Gladwell insightfully explains in his book *Blink* (Little Brown)?

According to the study, 72% of searchers click on the first link of interest, whereas 25.5% read all listings first and then decide which one to click. Both effects (implied endorsement and rapid cognition) most likely play a role in searcher behavior.

Different Intents and Effects of Listings in Paid Versus Natural Results

The AOL data in Figure 1-11 demonstrated that natural results get the lion's share of click results. Further data from the Enquiro, Didit, and Eyetools eye-tracking study shows which results users notice when looking at a search results page (see Table 1-6).

TABLE 1-6. Visibility of natural search results

Rank	Visibility
1	100%
2	100%
3	100%
4	85%
5	60%
6	50%
7	50%
8	30%
9	30%
10	20%

Table 1-7 shows the percentage of participants looking at a listing in this location in the paid results.

TABLE 1-7. Visibility of paid search results

Rank	Visibility
1	50%
2	40%
3	30%
4	20%
5	10%
6	10%
7	10%
8	10%

Notice how the visibility of a listing in the natural results is double or more (up to six times) of the visibility of the same position in the paid results. For example, in position 5 only 60% of users ever even notice the natural search result. But the paid search results fare even worse, with only 10% of users noticing the result in the fifth position.

Here are some additional takeaways from the Enquiro et al. study:

- 85% of searchers click on natural results.

- The top four sponsored slots are equivalent in views to being ranked at 7–10 in natural search in terms of visibility and click-through.
- This means if you need to make a business case for natural search, then (assuming you can attain at least the #3 rank in natural search for the same keywords you bid on) natural search could be worth two to three times your PPC results.

It is interesting to note that in spite of this data, companies are much more likely to spend money on PPC than SEO. For example, on June 30, 2008, Jupiter Research released a report (*http://news.cnet.com/8301-1023_3-9980927-93.html*) showing that search advertising should continue to be the largest category of online ad spending, growing from $9.1 billion in 2007 to $20.9 billion in 2013. That is significant growth.

During the same period, SEMPO data via Massimo Burgio shows that spending on SEO was $1.3 billion, with 11% of search-related budgets going to SEO and 87% to PPC (with another 1.4% to PPC technologies and <1% to paid inclusion).

Clearly, the PPC model is easier for companies to understand because it is more similar to traditional direct marketing methods than SEO is. The ROI of PPC can be tracked and demonstrated more reliably than SEO; thus, to date it has been considered more accountable as a marketing channel. However, as budgets are tightening and the focus is shifting to the highest ROI search investments, the focus is increasingly on SEO.

Interaction Between Natural and Paid Search

ICrossing published a report (*http://www.icrossing.com/articles/Search%20Synergy%20Report.pdf*) that showed a strong synergy between natural and paid search. The study shows what happens when you incorporate natural search into an existing paid search campaign and compare its performance to the performance of the sole paid search campaign. Figure 1-12 summarizes the improvement in the results.

The marked improvement in click-through rate intuitively makes sense. For years marketers have known that the number of impressions a consumer is exposed to will have a dramatic effect on metrics such as retention and likelihood to buy.

A search page provides you with more than one opportunity to put your name in front of the user. You should take advantage of this if you can. It is also useful to understand the difference between natural and paid search. Although some users do not understand the distinction between natural search results and paid search results, it is a well-accepted belief in the industry that the majority of users recognize paid search results as advertisements.

Other Factors to Consider

There are many other aspects to consider when thinking about search and your business. Here are some interesting examples:

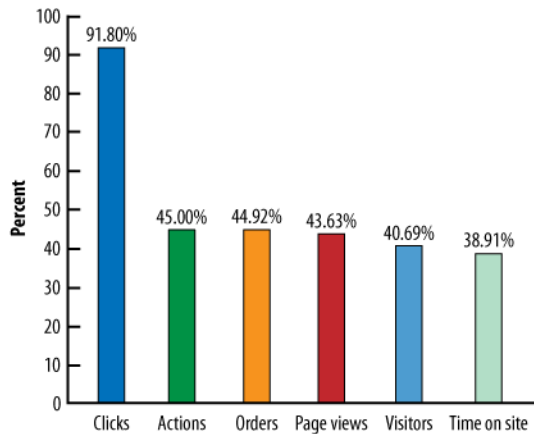**Online performance percentage change when integrating natural search and paid search**

FIGURE 1-12. Interaction between natural and paid search

Number of visits before purchase

A May 2007 study by ScanAlert showed that only 43% of users who made a purchase on a site made that purchase within an hour of their initial visit to the site. So, to translate that a bit, only 43% of users were likely to have made their purchase upon their first visit to the site. In addition, 26% of purchasers made their purchases after three days or more! This suggests that it is very important for publishers to plan on more complex interactions with users across multiple visits to the site.

Presentation changes made by the search engines

In February 2008, Yahoo! announced its SearchMonkey platform. This platform allows publishers and developers to directly influence the search results shown when your web pages are included in those results. Figure 1-13 shows a sample search result that demonstrates this.

Notice how the third listing (Yelp) looks different from the rest of the results. The difference in the look of the results really catches the eye, and can significantly impact click-through rates. And it works, as Amit Kumar of Yahoo! recently stated: "Our tests uncovered that users found these apps useful; in fact, in some cases, we saw a lift in click-through rate of as high as 15 percent."
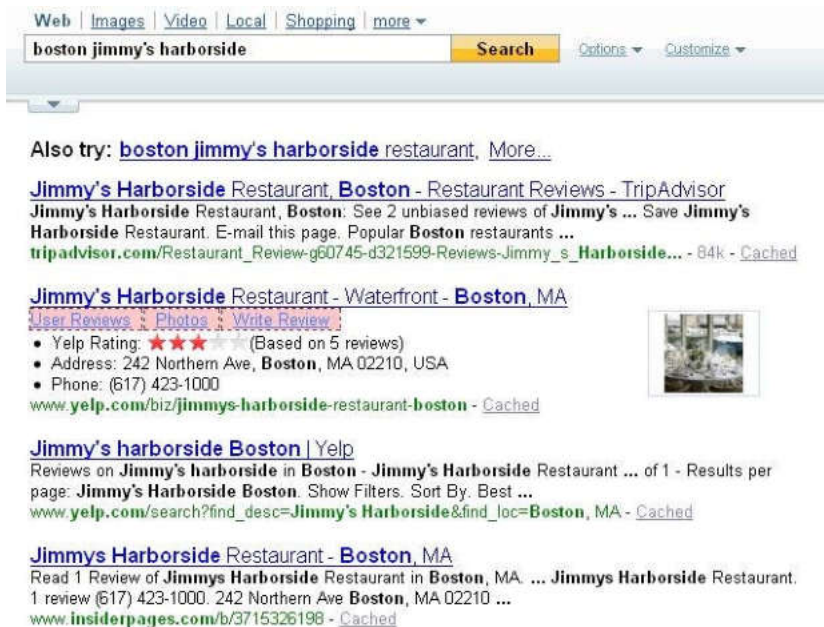
FIGURE 1-13. Yahoo!'s SearchMonkey

Conclusion

Search has penetrated the very fabric of global society. The way people work, play, shop, research, and interact has changed forever. Organizations of all kinds (businesses and charities), as well as individuals, need to have a presence on the Web—and they need the search engines to bring them traffic. As our society moves ever closer to a professional consumer ("prosumer") economy, the ways in which people create, publish, distribute, and ultimately find information and resources on the Internet will continue to be of great importance. This book will investigate further just how search, and therefore search engine optimization, is at the center of the Web and is your key to success in the new web economy.

Search Engine Basics

IN THIS CHAPTER, WE WILL BEGIN TO EXPLORE HOW SEARCH ENGINES WORK. Building a strong foundation on this topic is essential to understanding the SEO practitioner's craft.

As we discussed in Chapter 1, people have become accustomed to receiving nearly instantaneous answers from search engines after they have submitted a search query. In Chapter 1 we also discussed the volume of queries (more than 4,500 per second), and Google reported in July 2008 that it knew of about 1 trillion pages on the Web (*http://googleblog.blogspot .com/2008/07/we-knew-web-was-big.html*).

Underlying this enormous data processing task is the complex nature of the task itself. One of the most important things to understand about search engines is that the crawlers (or spiders) used to visit all the web pages across the Web are software programs. Software programs are only as smart as the algorithms used in implementing them, and although artificial intelligence is being increasingly used in those algorithms, web crawling programs still don't have the adaptive intelligence of human beings.

Software programs cannot adequately interpret each of the various types of data that humans can—videos and images, for example, are to a certain extent less readable by a search engine crawler than they are through the eyes of humans. In the section "Vertical Search Engines" on page 65, we will discuss how advances in image and video search have enabled search engines to inch even closer to human-like understanding.

Understanding Search Engine Results

In the search marketing field, the pages the engines return to fulfill a query are referred to as *search engine results pages* (SERPs). Each engine returns results in a slightly different format and will include *vertical results* (specific content targeted to a query based on certain triggers in the query, which we'll illustrate shortly).

Understanding the Layout of Search Results Pages

Figure 2-1 shows the SERPs in Google for the query *stuffed animals*.

1 Vertical navigation   **2** Search query box   **3** Results information
4 Paid search advertising   **5** Organic/algorithmic results

FIGURE 2-1. Layout of Google search results

Figure 2-2 shows Yahoo!'s results for the same query.

Figure 2-3 shows the layout of the results from Microsoft's Bing.

Each unique section represents a snippet of information provided by the engines. Here are the definitions of what each piece is meant to provide:

Vertical navigation

> Each engine offers the option to search different verticals, such as images, news, video, or maps. Following these links will result in a query with a more limited index. In Figure 2-3, for example, you might be able to see news items about stuffed animals or videos featuring stuffed animals.

Search query box

> All of the engines show the query you've performed and allow you to edit or reenter a new query from the search results page. Next to the search query box, the engines also

| 1 Vertical navigation | 2 Search query box | 3 Results information |
| 4 Paid search advertising | 5 Organic/algorithmic results | 6 Query refinement suggestions |

FIGURE 2-2. Layout of Yahoo! search results

| 1 Vertical navigation | 2 Search query box | 3 Results information |
| 4 Paid search advertising | 5 Organic/algorithmic results | 6 Query refinement suggestions |

FIGURE 2-3. Layout of Bing search results

offer links to the advanced search page, the features of which we'll discuss later in the book.

Results information

This section provides a small amount of meta information about the results that you're viewing, including an estimate of the number of pages relevant to that particular query (these numbers can be, and frequently are, wildly inaccurate and should be used only as a rough comparative measure).

PPC (a.k.a. paid search) advertising

The "Sponsored Results," to use Yahoo!'s terms, are text ads purchased by companies that use the various search ad platforms—Google AdWords, Yahoo! Search Marketing, and Microsoft adCenter. The results are ordered by a variety of factors, including relevance (for which click-through rate, use of searched keywords in the ad, and relevance of the landing page are factors in Google) and bid amount (the ads require a maximum bid, which is then compared against other advertisers' bids).

Natural/organic/algorithmic results

These results are pulled from the search engines' primary indexes of the Web and ranked in order of relevance and popularity according to their complex algorithms. This area of the results is the primary focus of this section of the book.

Query refinement suggestions

A relatively recent feature, query refinements are now offered by all three engines. The goal of these links is to let users search with a more specific and possibly more relevant query that will satisfy their intent.

In March 2009, Google enhanced the refinements by implementing Orion Technology, based on technology Google acquired in 2006. The goal of this enhancement is to provide a wider array of refinement choices. For example, a search on *principles of physics* will display refinements for the Big Bang, angular momentum, quantum physics, and special relativity.

Be aware that the SERPs are always changing as the engines test new formats and layouts. Thus, the images in Figures 2-1 through 2-3 may be accurate for only a few weeks or months until Google, Yahoo!, and Microsoft shift to new formats.

How Vertical Results Fit into the SERPs

These "standard" results, however, are certainly not all that the engines have to offer. For many types of queries, search engines show *vertical* results, or *instant answers* (which Google collectively refers to as *onebox results*), and include more than just links to other sites to help answer a user's questions. These types of results present many additional challenges and opportunities for the SEO practitioner.

Figure 2-4 shows an example of these types of results. The query in Figure 2-4 brings back a direct map with an address and the option to get directions. This result is drawn from a Google Maps search and attempts to provide the user with the answer he is seeking directly in the search results.

Figure 2-5 shows another example. The Google search in Figure 2-5 for weather plus a city name returns a direct answer. Once again, the user may not even need to click on a website if all he wanted to know was the temperature.

FIGURE 2-4. Local search result for a business

FIGURE 2-5. Weather search on Google

Figure 2-6 is an example of a search for a well-known painter. A Google search for the famous painter Edward Hopper returns image results of some of his most memorable works. This example is a little different from the "instant answers" type of result shown in Figures 2-4 and 2-5. If the user is interested in the first painting shown, he may well click on it to see the painting in a larger size or to get more information about it. For the SEO practitioner, getting placed in this onebox result could be a significant win.

Figure 2-7 shows an example from Yahoo!. A query on Yahoo! for *chicago restaurants* brings back a list of popular dining establishments from Yahoo!'s local portal. High placement in these results has likely been a good thing for Giordanos Pizzeria.

Figure 2-8 is an example of an instant answer result from Yahoo!. Searching Yahoo! to find the number of pounds in a ton brings back an instant answer of 2,000 pounds. Microsoft's Bing also presents vertical results and instant answer results.

FIGURE 2-6. Google search on an artist's name

FIGURE 2-7. Yahoo! search for Chicago restaurants

FIGURE 2-8. Yahoo! instant answer result

Figure 2-9 is an example of a celebrity search.

The results in Figure 2-9 include a series of images of the famous actor Charlie Chaplin. As a last example, Figure 2-10 is a look at the Bing search results for videos with Megan Fox.

At the top of the search results in Figure 2-10, you get a series of popular videos provided. Click on a video in the results and the video begins playing right there in the search results.

FIGURE 2-9. Bing result for Charlie Chaplin
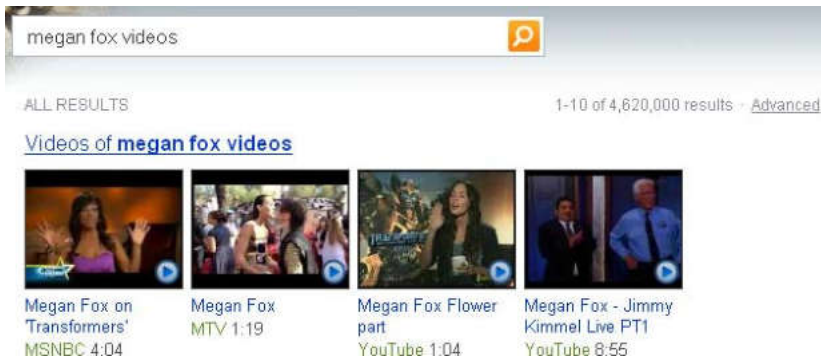
FIGURE 2-10. Bing result for Megan Fox videos

As you can see, the vast variety of vertical integration into search results means that for many popular queries, the standard set of 10 links to external pages is no longer the rule. Engines are competing by providing more relevant results and more targeted responses to queries that they feel are best answered by vertical results, rather than web results.

As a direct consequence, site owners and web marketers must take into account how this incorporation of vertical search results may impact their rankings and traffic. For many of the searches shown in the previous figures, a high ranking—even in position #1 or #2 in the algorithmic/organic results—may not produce much traffic because of the presentation of the vertical results above them.

The vertical results also signify an opportunity, as listings are available in services from images to local search to news and products. We will cover how to get included in these results in Chapter 8.

Algorithm-Based Ranking Systems: Crawling, Indexing, and Ranking

Understanding how crawling, indexing, and ranking works is helpful to SEO practitioners, as it helps them determine what actions to take to meet their goals. This section primarily covers the way Google, Yahoo!, and Microsoft operate, and does not necessarily apply to other search engines that are popular, such as Baidu (China) and Naver (Korea).

The search engines have several major goals and functions. These include:

- Crawling and indexing the billions of documents (pages and files) accessible on the Web
- Responding to user queries by providing lists of relevant pages

In this section, we'll walk through the basics of these functions from a nontechnical perspective. This section will start by discussing how search engines find and discover content.

Crawling and Indexing

Imagine the World Wide Web as a network of stops in a big city subway system. Each stop is its own unique document (usually a web page, but sometimes a PDF, JPEG, or other file). The search engines need a way to "crawl" the entire city and find all the stops along the way, so they use the best path available: the links between web pages, an example of which is shown in Figure 2-11.

In our representation in Figure 2-11, stops such as Embankment, Piccadilly Circus, and Moorgate serve as pages, while the lines connecting them represent the links from those pages to other pages on the Web. Once Google (at the bottom) reaches Embankment, it sees the links pointing to Charing Cross, Westminster, and Temple and can access any of those "pages."

The link structure of the Web serves to bind together all of the pages that were made public as a result of someone linking to them. Through links, search engines' automated robots, called *crawlers* or *spiders* (hence the illustrations in Figure 2-11), can reach the many billions of interconnected documents.

Once the engines find these pages, their next job is to parse the code from them and store selected pieces of the pages in massive arrays of hard drives, to be recalled when needed in a query. To accomplish the monumental task of holding billions of pages that can be accessed in a fraction of a second, the search engines have constructed massive data centers to deal with all this data.

One key concept in building a search engine is deciding where to begin a crawl of the Web. Although you could theoretically start from many different places on the Web, you would ideally begin your crawl with a trusted set of websites. You can think of a factor in evaluating the trust in your website as the click distance between your website and the most trusted
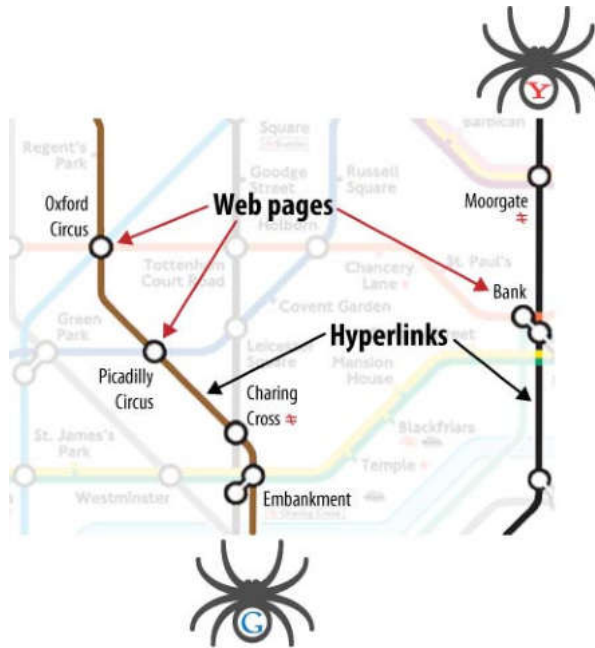
FIGURE 2-11. London's Tube used as an analogy for web crawling

websites. We will discuss the role of trust in search algorithms in more detail in "How Links Influence Search Engine Rankings" on page 279 in Chapter 7.

Retrieval and Rankings

For most searchers, the quest for knowledge begins as shown in Figure 2-12.

FIGURE 2-12. Start of a user's search quest

The next step in this quest for knowledge occurs when the search engine returns a list of relevant pages on the Web in the order most likely to satisfy the user. This process requires the search engines to scour their corpus of billions of documents and do two things: first, return only the results that are related to the searcher's query; and second, rank the results in order of perceived importance (taking into account the trust and authority associated with the site). It is both relevance and importance that the process of SEO is meant to influence.

Relevance is the degree to which the content of the documents returned in a search matches the user's query intention and terms. The relevance of a document increases if the terms or

phrase queried by the user occurs multiple times and shows up in the title of the work or in important headlines or subheads, or if links to the page come from relevant pages and use relevant anchor text.

You can think of relevance as the first step to being "in the game." If you are not relevant to a query, the search engine does not consider you for inclusion in the search results for that query. We will discuss how relevance is determined in more detail in "Determining Searcher Intent and Delivering Relevant, Fresh Content" on page 41.

Importance or *popularity* refers to the relative importance, measured via *citation* (the act of one work referencing another, as often occurs in academic and business documents) of a given document that matches the user's query. The popularity of a given document increases with every other document that references it. In the academic world, this concept is known as *citation analysis*.

You can think of importance as a way to determine which page, from a group of equally relevant pages, shows up first in the search results, which is second, and so forth. The relative authority of the site, and the trust the search engine has in it, are significant parts of this determination. Of course, the equation is a bit more complex than this and not all pages are equally relevant. Ultimately, it is the combination of relevance and importance that determines the ranking order.

So, when you see a search results page such as the one shown in Figure 2-13, you can surmise that the search engine (in this case, Yahoo!) believes the Superhero Stamps page on USPS.com has the highest combined score for relevance and popularity for the query *marvel superhero stamps*, while the Yahoo! Shopping page has a lower combined score for relevance and popularity.

Popularity and relevance aren't determined manually (those trillions of man-hours would require Earth's entire population as a workforce). Instead, the engines craft careful, mathematical equations—algorithms—to sort the wheat from the chaff and to then rank the wheat in order of quality. These algorithms often comprise hundreds of components. In the search marketing field, they are often referred to as *ranking factors* or *algorithmic ranking criteria*.

We discuss ranking factors or signals (*signals* is the term Google prefers) in more detail in "Analyzing Ranking Factors" on page 49.

Evaluating Content on a Web Page

Search engines place a lot of weight on the content of each web page. After all, it is this content that defines what a page is about, and the search engines do a detailed analysis of each web page they find during their crawl to help make that determination.

You can think of this as the search engine performing a detailed analysis of all the words and phrases that appear on a web page, and then building a map of that data for it to consider showing your page in the results when a user enters a related search query. This map, often
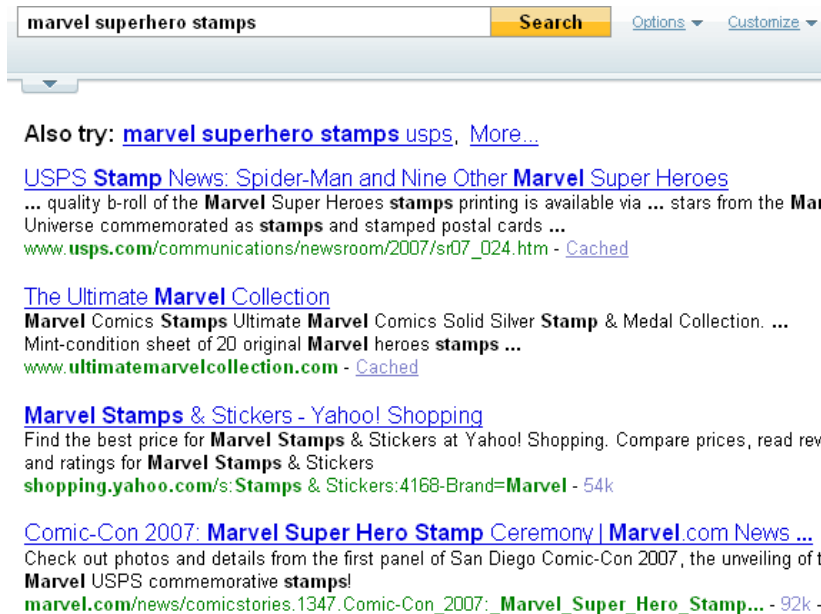
referred to as a *semantic map*, seeks to define the relationships between those concepts so that the search engine can better understand how to match the right web pages with user search queries.

If there is no semantic match of the content of a web page to the query, the page has a much lower possibility of showing up. Therefore, the words you put on the page, and the "theme" of that page, play a huge role in ranking.

Figure 2-14 shows how a search engine will break up a page when it looks at it, using a page on the Stone Temple Consulting website.

The navigational elements of a web page are likely similar across the many pages of a site. These navigational elements are not ignored, and they do play an important role, but they do not help a search engine determine what the unique content is on a page. To do that, the search engine gets very focused on the part of Figure 2-14 that is labeled "Real content".

Determining the unique content on a page is an important part of what the search engine does. It is this understanding of the unique content on a page that the search engine uses to determine the types of search queries for which the web page might be relevant. Since site navigation is generally not unique to a single web page, it does not help the search engine with that task.

FIGURE 2-14. Breaking up a web page

This does not mean navigation links are not important, because they most certainly are—however, they simply do not count when trying to determine the unique content of a web page because those navigation links are shared among many web pages.

One task the search engines face is judging the value of content. Although evaluating how the community responds to a piece of content using link analysis is part of the process, the search engines can also draw some conclusions based on what they see on the page.

For example, is the exact same content available on another website? Is the unique content the search engine can see two sentences long or 500+ words long? Does the content repeat the same keywords excessively? These are a few examples of things that the search engine can look at when trying to determine the value of a piece of content.

What Content Can Search Engines "See" on a Web Page?

Search engine crawlers and indexing programs are basically software programs. These programs are extraordinarily powerful. They crawl hundreds of billions of web pages, analyze the content of all these pages, and analyze the way all these pages link to each other. Then they organize this into a series of databases that can respond to a user search query with a highly tuned set of results in a few tenths of a second.

This is an amazing accomplishment, but it has its limitations. Software is very mechanical, and it can understand only portions of most web pages. The search engine crawler analyzes the raw HTML form of a web page. If you want to see what this looks like, you can do so by using your browser to view the source.

The two screen shots in Figure 2-15 show how to do that in Firefox (top) and Internet Explorer (bottom).

FIGURE 2-15. Viewing source in your browser

Once you view the source, you will be presented with the exact code for the website that the web server sent to your browser. This is what the search engine crawler sees (the search engine also sees the HTTP headers for the page). The crawler will ignore a lot of what is in the code. For example, search engines largely ignore code such as that shown in Figure 2-16, as it has nothing to do with the content of the web page.

```
<div id="wrapper">
  <div id="page-top">
    <div class="container">
      <ul id="site-nav">
              <li id="user_state">Aloha, <a href="/users/view/18040">Eric Enge</a></li>
              <li><a href="/messages"  id="messages">0 Messages</a></li>
        <li><a href="/users/my">My Account</a></li>
        <li><a href="/users/logout"  onclick="return confirm('Are you sure you want to logout?');">Logout</a>
</li>
              <!-- PRO Tour upgrade link in the nav -->
              <li>
        <form id="site-search" action="http://search.yahoo.com/search" method="get">
          <fieldset class="val">
            <input type="hidden" name="vs" value="seomoz.org" />
            <input type="text" name="p" id="site-search-text" tabindex="5" title="Search SEOmoz.org" />
            <input type="submit" name="site-search-go" id="site-search-go" value="SEARCH" />
          </fieldset>
        </form>
      </li>
    </ul>
  </div> <!-- end container -->
</div> <!-- end page top -->
```
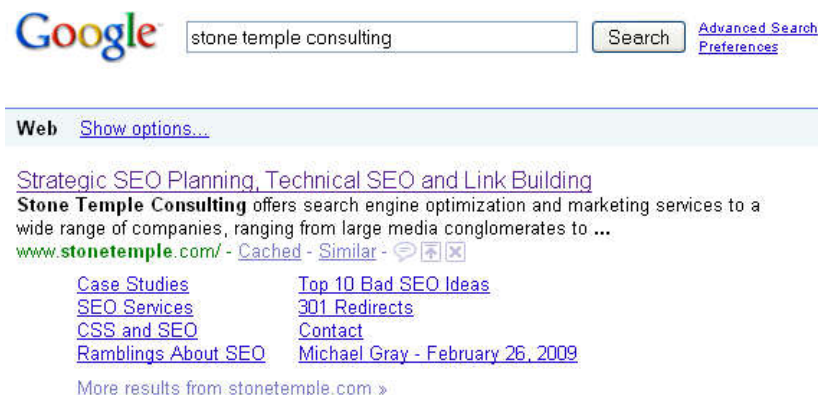
FIGURE 2-16. Sample web page source code

The information the search engine crawler is most interested in is in the HTML text on the page. Figure 2-17 is an example of HTML text for a web page (using the SEOmoz.org home page).

```
<li class="last">
<em><a hidefocus="true" id="money1" href="r/wq">Life</a></em>
<span class="pipe"></span>
</li>
</ul>
</div>
<div id="todayvideo">
</div>
<div id="todaybd" class="bd">
<cite class="timestamp"> </cite>
<span id="featured1ct" class="current">
<div id="featured1main">
<a href=s/1102897><img src="http://l.yimg.com/a/i/ww/news/2009/07/20/moon.jpg" width="154" height="115"
alt="In this July 20, 1969 file photo, Astronaut Edwin E. 'Buzz' Aldrin Jr. poses for a photograph beside the
U.S. flag deployed on the moon during the Apollo 11 mission. (AP Photo/Neil Armstrong, NASA, file)"></a>
<span>
<h3><a href=s/1102897>Astronauts' pitch to Obama</a></h3>
<p>The 40th anniversary of man's first steps on the moon sparks a bold request for the president.
<a class=more href=s/1102897>&#187; The future</a></p>
<ul>
<li><a class=search href=s/1102898>Find habitable planets</a></li>
<li><a class=bullet href=s/1102899>Lost word</a></li>
<li><a class=video href=s/1102900
onclick="window.open('s/1102900','playerWindow','width=793,height=666,scrollbars=no');return false;">Video</a>
```

FIGURE 2-17. Sample HTML text in the source code

Although Figure 2-17 still shows some HTML encoding, you can see the "regular" text clearly in the code. This is the unique content that the crawler is looking to find.

In addition, search engines read a few other elements. One of these is the page title. The page title is one of the most important factors in ranking a given web page. It is the text that shows in the browser's title bar (the blue line above the browser menu and the address bar).

Figure 2-18 shows the code that the crawler sees, using Trip Advisor as an example.

The first red circle in Figure 2-18 is for the title tag. The title tag is also often (but not always) used as the title of your listing in search engine results. Exceptions to this can occur when you obtain Yahoo! or DMOZ directory listings for your site. Sometimes the search engines may

choose to use a title for your page that was used in your listings in these directories, instead of the title tag on the page. There are also meta tags that allow you to block this from happening, such as the NOODP tag, which tells the search engine not to use DMOZ titles, and the NOYDIR tag, which tells Yahoo! not to use the Yahoo! directory listing. In any event, Figure 2-19 shows what happens when you search on *stone temple consulting* (the Stone Temple Consulting home page at *http://www.stonetemple.com*). Notice how the title of the search listing matches the title of the Stone Temple Consulting home page.

FIGURE 2-18. Meta tags in HTML source

FIGURE 2-19. Search result showing title tag

In addition to page titles, the search engines also read the "meta keywords" tag. This is a list of keywords that you wish to have associated with the page. *Spammers* (people who attempt to manipulate search engine results in violation of the search engine guidelines) ruined the SEO value of this tag many years ago, so its value is now negligible. Google does not use this tag for ranking at all, but Yahoo! and Bing seem to make reference to it (you can read about this in detail at *http://searchengineland.com/meta-keywords-tag-101-how-to-legally-hide-words-on-your -pages-for-search-engines-12099*). Spending a lot of time on meta keywords is not recommended because of the lack of SEO benefit.

The second red circle in Figure 2-18 shows an example of a meta keywords tag.

Search engines also read the meta description tag (the third red circle in the HTML in Figure 2-18). However, the meta description tag is not of any influence in search engine rankings (*http://searchengineland.com/21-essential-seo-tips-techniques-11580*).

Nonetheless, the meta description tag plays a key role as search engines often use it as the description for your page in search results. Therefore, a well-written meta description can have a significant influence on how many clicks you get on your search listing. Time spent on meta descriptions is quite valuable as a result. Figure 2-20 uses a search on *trip advisor* to show an example of the meta description being used as a description in the search results.
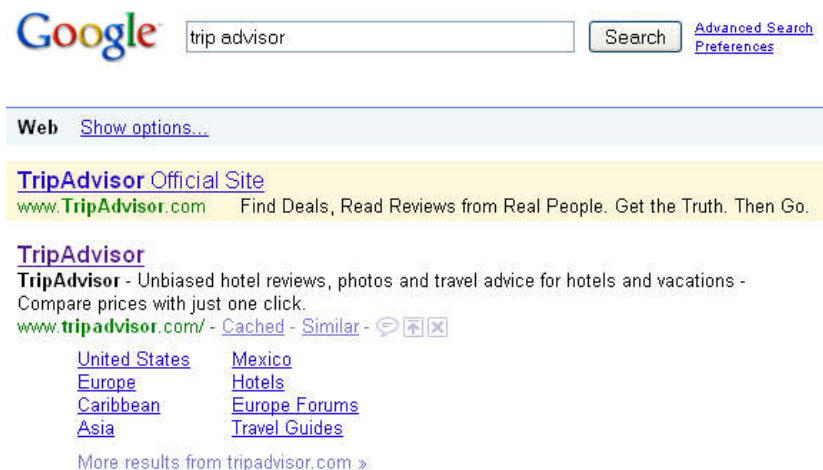
FIGURE 2-20. Meta description used in search results

> **N O T E**
>
> **The user's keywords are typically shown in boldface when they appear in the search results (sometimes close synonyms are shown in boldface as well). As an example of this, in Figure 2-20, *TripAdvisor* is in boldface at the beginning of the description.**

A fourth element that search engines read is the `alt` attribute for images. The `alt` attribute was originally intended to allow something to be rendered when viewing of the image is not possible. There were two basic audiences for this:

- Vision-impaired people who do not have the option of viewing the images.
- People who turn off images for faster surfing. This is generally an issue only for those who do not have a broadband connection.

Support for the vision-impaired remains a major reason for using the alt attribute. You can read more about this by visiting the Web Accessibility Initiative page on the W3C website.

Search engines also read the text contained in the alt attribute of an image tag. An image tag is an element that is used to tell a web page to display an image. Here is an example of an image tag from the Alchemist Media site:

```
<img src="http://www.alchemistmedia.com/img/btob2009.jpg" alt="BtoB Interactive
Marketing Guide" border="0" />
```

The src= part of the tag is where the image to be displayed is located. The part that starts with alt= and is then followed with "BtoB Interactive Marketing Guide" is considered the alt attribute.

The alt attribute is something that the search engines read. The search engines interpret it to help them determine what the image is about and to get more of a sense as to what the page is about.

A final element that search engines read is the noscript tag. In general, search engines do not try to interpret JavaScript that may be present on a web page (though this is already changing). However, a small percentage of users do not allow JavaScript to run when they load a web page (the authors' experience is that it is about 2%). For those users, nothing would be shown where the JavaScript is on the web page, unless the page contains a noscript tag.

Here is a very simple JavaScript example that demonstrates this:

```
<script type="text/javascript">
document.write("It is a Small World After All!")
</script>
<noscript>Your browser does not support JavaScript!</noscript>
```

The noscript portion of this is Your browser does not support JavaScript!. The search engines will read this text and see that as information about the web page. In this example, you could also choose to make the noscript tag contain the text "It is a Small World After All!". The noscript tag should be used only to represent the content of the JavaScript. (Placing other content or links in this tag could be interpreted as spammy behavior by the search engines.) In addition, the browser warning could end up as your search snippet, which would be a bad thing.

What search engines cannot see

It is also worthwhile to review the types of content that search engines cannot "see" in the human sense.

For instance, although search engines are able to detect that you are displaying an image, they have little idea what the image is a picture of, except for whatever information you provide them in the `alt` attribute, as discussed earlier. They can, however, determine pixel color and, in many instances, determine whether images have pornographic content by how much flesh tone there is in a JPEG image. So, a search engine cannot tell whether an image is a picture of Bart Simpson, a boat, a house, or a tornado. In addition, search engines will not recognize any text rendered in the image. The search engines are experimenting with technologies to use optical character recognition (OCR) to extract text from images, but this technology is not yet in general use within search.

In addition, conventional SEO wisdom has always held that the search engines cannot read Flash files, but this is a little overstated. Search engines are beginning to extract information from Flash, as indicated by the Google announcement at *http://googlewebmastercentral.blogspot .com/2008/06/improved-flash-indexing.html*. However, the bottom line is that it's not easy for search engines to determine what is in Flash. One of the big issues is that even when search engines look inside Flash, they are still looking for textual content, but Flash is a pictorial medium and there is little incentive (other than the search engines) for a designer to implement text inside Flash. All the semantic clues that would be present in HTML text (such as heading tags, boldface text, etc.) are missing too, even when HTML is used in conjunction with Flash.

A third type of content that search engines cannot see is the pictorial aspects of anything contained in Flash, so this aspect of Flash behaves in the same way images do. For example, when text is converted into a vector-based outline (i.e., rendered graphically), the textual information that search engines can read is lost. We discuss methods for optimizing Flash in "Redirects" on page 190 in Chapter 6.

Audio and video files are also not easy for search engines to read. As with images, the data is not easy to parse. There are a few exceptions where the search engines can extract some limited data, such as ID3 tags within MP3 files, or enhanced podcasts in AAC format with textual "show notes," images, and chapter markers embedded. Ultimately, though, a video of a soccer game cannot be distinguished from a video of a forest fire.

Search engines also cannot read any content contained within a program. The search engine really needs to find text that is readable by human eyes looking at the source code of a web page, as outlined earlier. It does not help if you can see it when the browser loads a web page— it has to be visible and readable in the source code for that page.

One example of a technology that can present significant human-readable content that the search engines cannot see is AJAX. AJAX is a JavaScript-based method for dynamically rendering content on a web page after retrieving the data from a database, without having to

refresh the entire page. This is often used in tools where a visitor to a site can provide some input and the AJAX tool then retrieves and renders the correct content.

The problem arises because the content is retrieved by a script running on the client computer (the user's machine) only after receiving some input from the user. This can result in many potentially different outputs. In addition, until that input is received the content is not present in the HTML of the page, so the search engines cannot see it.

Similar problems arise with other forms of JavaScript that don't render the content in the HTML until a user action is taken.

As of HTML 5, a construct known as the embed tag (`<embed>`) was created to allow the incorporation of plug-ins into an HTML page. Plug-ins are programs located on the user's computer, not on the web server of your website. This tag is often used to incorporate movies or audio files into a web page. The `<embed>` tag tells the plug-in where it should look to find the datafile to use. Content included through plug-ins is not visible at all to search engines.

Frames and iframes are methods for incorporating the content from another web page into your web page. Iframes are more commonly used than frames to incorporate content from another website. You can execute an iframe quite simply with code that looks like this:

```
<iframe src ="http://accounting.careerbuilder.com" width="100%" height="300">
  <p>Your browser does not support iframes.</p>
</iframe>
```

Frames are typically used to subdivide the content of a publisher's website, but they can be used to bring in content from other websites, as was done in Figure 2-21 with *http://accounting .careerbuilder.com* on the *Chicago Tribune* website.

Figure 2-21 is an example of something that works well to pull in content (provided you have permission to do so) from another site and place it on your own. However, the search engines recognize an iframe or a frame used to pull in another site's content for what it is, and therefore ignore the content inside the iframe or frame as it is content published by another publisher. In other words, they don't consider content pulled in from another site as part of the unique content of your web page.

Determining Searcher Intent and Delivering Relevant, Fresh Content

Modern commercial search engines rely on the science of information retrieval (IR). This science has existed since the middle of the twentieth century, when retrieval systems powered computers in libraries, research facilities, and government labs. Early in the development of search systems, IR scientists realized that two critical components comprised the majority of search functionality: relevance and importance (which we defined earlier in this chapter). To measure these factors, search engines perform document analysis (including semantic analysis of concepts across documents) and link (or citation) analysis.

FIGURE 2-21. Framed page rendered in a browser

Document Analysis and Semantic Connectivity

In *document analysis,* search engines look at whether they find the search terms in important areas of the document—the title, the metadata, the heading tags, and the body of the text. They also attempt to automatically measure the quality of the document based on document analysis, as well as many other factors.

Reliance on document analysis alone is not enough for today's search engines, so they also look at semantic connectivity. *Semantic connectivity* refers to words or phrases that are commonly associated with one another. For example, if you see the word *aloha* you associate it with Hawaii, not Florida. Search engines actively build their own thesaurus and dictionary to help them determine how certain terms and topics are related. By simply scanning their massive databases of content on the Web, they can use Fuzzy Set Theory and certain equations

(described at *http://forums.searchenginewatch.com/showthread.php?threadid=48*) to connect terms and start to understand web pages/sites more like a human does.

The professional SEO practitioner does not necessarily need to use semantic connectivity measurement tools to optimize websites, but for those advanced practitioners who seek every advantage, semantic connectivity measurements can help in each of the following sectors:

- Measuring which keyword phrases to target
- Measuring which keyword phrases to include on a page about a certain topic
- Measuring the relationships of text on other high-ranking sites/pages
- Finding pages that provide "relevant" themed links

Although the source for this material is highly technical, SEO specialists need only know the principles to obtain valuable information. It is important to keep in mind that although the world of IR has hundreds of technical and often difficult-to-comprehend terms, these can be broken down and understood even by an SEO novice.

Table 2-1 explains some common types of searches in the IR field.

TABLE 2-1. Common types of searches

Proximity searches	A proximity search uses the order of the search phrase to find related documents. For example, when you search for "*sweet German mustard*" you are specifying only a precise proximity match. If the quotes are removed, the proximity of the search terms still matters to the search engine, but it will now show documents that don't exactly match the order of the search phrase, such as *Sweet Mustard—German.*
Fuzzy logic	Fuzzy logic technically refers to logic that is not categorically true or false. A common example is whether a day is sunny (is 50% cloud cover a sunny day, etc.). In search, fuzzy logic is often used for misspellings.
Boolean searches	These are searches that use Boolean terms such as AND, OR, and NOT. This type of logic is used to expand or restrict which documents are returned in a search.
Term weighting	Term weighting refers to the importance of a particular search term to the query. The idea is to weight particular terms more heavily than others to produce superior search results. For example, the appearance of the word *the* in a query will receive very little weight in selecting the results because it appears in nearly all English language documents. There is nothing unique about it, and it does not help in document selection.

IR models (search engines) use Fuzzy Set Theory (an offshoot of fuzzy logic created by Dr. Lotfi Zadeh in 1969) to discover the semantic connectivity between two words. Rather than using a thesaurus or dictionary to try to reason whether two words are related to each other, an IR system can use its massive database of content to puzzle out the relationships.

Although this process may sound complicated, the foundations are simple. Search engines need to rely on machine logic (true/false, yes/no, etc.). Machine logic has some advantages over humans, but machine logic doesn't have a way of thinking like humans, and things that are

intuitive to humans can be quite hard for a computer to understand. For example, both oranges and bananas are fruits, but both oranges and bananas are not round. To a human this is intuitive.

For a machine to understand this concept and pick up on others like it, semantic connectivity can be the key. The massive human knowledge on the Web can be captured in the system's index and analyzed to artificially create the relationships humans have made. Thus, a machine knows an orange is round and a banana is not by scanning thousands of occurrences of the words *banana* and *orange* in its index and noting that *round* and *banana* do not have great concurrence, while *orange* and *round* do.

This is how the use of fuzzy logic comes into play, and the use of Fuzzy Set Theory helps the computer to understand how terms are related simply by measuring how often and in what context they are used together.

A related concept that expands on this notion is latent semantic analysis (LSA). The idea behind this is that by taking a huge composite (index) of billions of web pages, the search engines can "learn" which words are related and which noun concepts relate to one another.

For example, using LSA, a search engine would recognize that *trips* to the *zoo* often include *viewing wildlife* and *animals*, possibly as part of a *tour*.

Now, conduct a search on Google for *~zoo ~trips* (the tilde is a search operator; more on this later in this chapter). Note that the boldface words that are returned match the terms that are italicized in the preceding paragraph. Google is setting "related" terms in boldface and recognizing which terms frequently occur concurrently (together, on the same page, or in close proximity) in their indexes.

Some forms of LSA are too computationally expensive. For example, currently the search engines are not smart enough to "learn" the way some of the newer learning computers do at MIT. They cannot, for example, learn through their index that *zebras* and *tigers* are examples of striped animals, although they may realize that *stripes* and *zebras* are more semantically connected than *stripes* and *ducks*.

Latent semantic indexing (LSI) takes this a step further by using semantic analysis to identify related web pages. For example, the search engine may notice one page that talks about doctors and another one that talks about physicians, and determine that there is a relationship between the pages based on the other words in common between the pages. As a result, the page referring to doctors may still show up for a search query that uses the word *physician* instead.

Search engines have been investing in these types of technologies for many years. For example, in April 2003 Google acquired Applied Semantics, a company known for its semantic-text-processing technology. This technology currently powers Google's AdSense advertising program, and has most likely made its way into the core search algorithms as well.

For SEO purposes, this usage opens our eyes to realizing how search engines recognize the connections between words, phrases, and ideas on the Web. As semantic connectivity becomes

a bigger part of search engine algorithms, you can expect greater emphasis on the theme of pages, sites, and links. It will be important going into the future to realize the search engines' ability to pick up on ideas and themes and recognize content, links, and pages that don't fit well into the scheme of a website.

Link Analysis

In link analysis, search engines measure who is linking to a site or page and what they are saying about that site/page. They also have a good grasp on who is affiliated with whom (through historical link data, the site's registration records, and other sources), who is worthy of being trusted based on the authority of sites linking to them, and contextual data about the site on which the page is hosted (who links to that site, what they say about the site, etc.).

Link analysis goes much deeper than counting the number of links a web page or website has, as links are not created equal. Links from a highly authoritative page on a highly authoritative site will count more than other links of lesser authority. A website or page can be determined to be authoritative by combining an analysis of the linking patterns and semantic analysis.

For example, perhaps you are interested in sites about dog grooming. Search engines can use semantic analysis to identify the collection of web pages that focus on the topic of dog grooming. The search engines can then determine which of these sites about dog grooming have the most links from the set of dog grooming sites. These sites are most likely more authoritative on the topic than the others.

The actual analysis is a bit more complicated than that. For example, imagine that there are five sites about dog grooming with a lot of links from pages across the Web on the topic, as follows:

- Site A has 213 topically related links.
- Site B has 192 topically related links.
- Site C has 203 topically related links.
- Site D has 113 topically related links.
- Site E has 122 topically related links.

Further, it may be that Site A, Site B, Site D, and Site E all link to each other, but none of them link to Site C. In fact, Site C appears to have the great majority of its relevant links from other pages that are topically relevant but have few links to them. In this scenario, Site C is definitively not authoritative because it is not linked to by the right sites.

This concept of grouping sites based on their relevance is referred to as a *link neighborhood*. The neighborhood you are in says something about the subject matter of your site, and the number and quality of the links you get from sites in that neighborhood say something about how important your site is to that topic.

The degree to which search engines rely on evaluating link neighborhoods is not clear, and is likely to differ among search engines. In addition, links from nonrelevant pages are still believed to help the rankings of the target pages. Nonetheless, the basic idea remains that a link from a relevant site should count for more than a link from a nonrelevant site.

Another factor in determining the value of a link is the way the link is implemented and where it is placed. For example, the text used in the link itself (i.e., the actual text that will go to your web page when the user clicks on it) is also a strong signal to the search engines.

This is referred to as *anchor text*, and if that text is keyword-rich (with keywords relevant to your targeted search terms), it will do more for your rankings in the search engines than if the link is not keyword-rich. For example, anchor text of "Dog Grooming Salon" will bring more value to a dog grooming salon's website than anchor text of "Click here". However, take care. If you get 10,000 links using the anchor text "Dog Grooming Salon" and you have few other links to your site, this definitely does not look natural and could lead to problems in your rankings.

The semantic analysis of a link's value does go deeper than just the anchor text. For example, if you have that "Dog Grooming Salon" anchor text on a web page that is not really about dog grooming at all, the value of the link is less than if the page is about dog grooming. Search engines also look at the content on the page immediately surrounding the link, as well as the overall context and authority of the website that is providing the link.

All of these factors are components of link analysis, which we will discuss in greater detail in Chapter 7.

Problem Words, Disambiguation, and Diversity

On the opposite side of the coin are words that present an ongoing challenge for the search engines. One of the greatest challenges comes in the form of disambiguation. For example, when someone types in *boxers*, does he mean the prize fighter, the breed of dog, or the type of underwear? Another example of this is *jaguar*, which is at once a jungle cat, a car, a football team, an operating system, and a guitar. Which does the user mean?

Search engines deal with these types of ambiguous queries all the time. The two examples offered here have inherent problems built into them, but the problem is much bigger than that. For example, if someone types in a query such as *cars*, does he:

- Want to read reviews?
- Want to go to a car show?
- Want to buy one?
- Want to read about new car technologies?

The query *cars* is so general that there is no real way to get to the bottom of the searcher's intent based on this one query alone. (The exception is cases where prior queries by the same

searcher may provide additional clues that the search engine can use to better determine the searcher's intent.)

As an example, Figure 2-22 shows another generic search, this time using *gdp*.

FIGURE 2-22. Diverse results example

This brings up an important ranking concept. It is possible that a strict analysis of the relevance and link popularity scores in Figure 2-22 would not have resulted by itself in the Investopedia.com page being on the first page, *but* the need for diversity caused the ranking of the page to be elevated.

A strict relevance- and importance-based ranking system might have shown a variety of additional government pages discussing the GDP of the United States. However, a large percentage of users will likely be satisfied by the government pages already shown, and showing more of them is not likely to raise the level of satisfaction with the results.

Introducing a bit of variety allows Google to also provide a satisfactory answer to those who are looking for something different from the government pages. Google's testing has shown that this diversity-based approach has resulted in a higher level of satisfaction among its users.

For example, the testing data for the nondiversified results may have shown lower click-through rates in the SERPs, greater numbers of query refinements, and even a high percentage of related searches performed subsequently.

When Google wants to get really serious about disambiguation, it goes a different route. Check out the SERPs in Figure 2-23.

FIGURE 2-23. Disambiguating search queries

These "horizontal line," disambiguation-style results appear on many searches where Google thinks the searcher is probably seeking something that his query isn't producing. They're especially likely to appear for very general search phrases.

The idea to deliberately introduce diversity into the result algorithm makes sense and can enhance searcher satisfaction for queries such as:

- Company names (where searchers might want to get positive and negative press, as well as official company domains)
- Product searches (where e-commerce-style results might ordinarily fill up the SERPs, but Google tries to provide some reviews and noncommercial, relevant content)
- News and political searches (where it might be prudent to display "all sides" of an issue, rather than just the left- or right-wing blogs that did the best job of baiting links)

Where freshness matters

Much of the time, it makes sense for the search engines to deliver results from older sources that have stood the test of time. However, other times the response should be from newer sources of information.

For example, when there is breaking news, such as an earthquake, the search engines begin to receive queries within seconds, and the first articles begin to appear on the Web within 15 minutes.

In these types of scenarios, there is a need to discover and index new information in near-real time. Google refers to this concept as *query deserves freshness* (QDF). According to the *New York Times* (*http://www.nytimes.com/2007/06/03/business/yourmoney/03google.html?pagewanted=3*), QDF takes several factors into account, such as:

- Search volume
- News coverage
- Blog coverage
- Toolbar data (maybe)

QDF applies to up-to-the-minute news coverage, but also to other scenarios such as hot, new discount deals or new product releases that get strong search volume and media coverage.

A Few Reasons Why These Algorithms Sometimes Fail

As we've outlined in this chapter, the search engines do some amazing stuff. Nonetheless, there are times when the process does not work as well as you would like to think. Part of this is because users often type in search phrases that provide very little information about their intent (e.g., if they search on *car*, do they want to buy one, read reviews, learn how to drive one, learn how to design one, or something else?). Another reason is that some words have multiple meanings, such as *jaguar* (which is an animal, a car, a guitar, and in its plural form, a football team).

For more information on reasons why search algorithms sometimes fail, you can read the following SEOmoz article, which was written by Hamlet Batista:

> *http://www.seomoz.org/blog/7-reasons-why-search-engines-dont-return-relevant-results-100-of-the-time*

Analyzing Ranking Factors

SEOmoz conducted a survey of leading SEOs to determine what these SEOs thought were the most important ranking factors (*http://www.seomoz.org/article/search-ranking-factors*). Here is a high-level summary of the top nine results:

- Keyword use in title tag
- Anchor text of inbound link
- Global link authority of site
- Age of site
- Link popularity within the site's internal link structure
- Topical relevance of inbound links
- Link popularity of site in topical community
- Keyword use in body text
- Global link popularity of sites that link to the site

Here is a brief look at each of these:

Keyword use in title tag

The title tag was clearly considered the #1 ranking factor. The words in the title tag say a lot about the topic of a page (or they should), and it is a useful element for the search engines to key on. One reason is that it is a highly visible element on the web page and therefore is likely to be used, well, as a title for the page.

It is also generally accepted that search engines place the most weight on the words that appear at the start of the title. For that reason it is always a good idea to put your most important keywords at the start of the title tag, but make sure to avoid *keyword stuffing* (a concept that we will discuss more in "Keyword Targeting" on page 211 in Chapter 6).

Anchor text of inbound link

When one web page links to another, the anchor text used for that link is also used as a strong signal by the search engines. Webmasters and publishers who choose to link to other web pages from their site often will place keyword-rich anchor text in those links.

This has led to a tremendous focus by link builders in an attempt to get websites to provide keyword-rich text links to their clients' websites.

Global link authority of site

This metric factors the total link authority of the site, as opposed to the link authority of a given page on the site. As defined in the SEOmoz survey, it is intended to capture the quantity and quality of a site's links. There are three quick ways to get a general sense of this:

- Look at the PageRank for the home page of the domain. Although the original PageRank algorithm is flawed, it can still provide a general sense of a site's global link popularity (one area where this is weak is that it will not give you a sense of the relevance of any of the links). Note, however, that the home page is not always the highest PageRank page on a site.
- Use the Yahoo! Site Explorer search. This will give you a raw count of all inbound links to the domain and allow you to download up to 1,000 results into a spreadsheet.
- Use Linkscape's domain mozRank.

Age of site

This metric often surprises novice SEO practitioners, but there is no question that it is a factor. The reason this makes for a valuable signal is that a site that has been around for a long time must be doing something right in the eyes of the community it is trying to serve. Spammy or poor-quality sites ultimately don't do as well, so they don't last as long. In addition, spam was less prevalent in 1998, so sites started in 1998 are less likely to be spam compared to sites started more recently.

Link popularity within the site's internal link structure

Search engines rely on the publisher to tell them what they think is the most important part of their site. For example, a page that can be reached only after five clicks from the home page is probably not important.

On the other hand, a page that is included in the global navigation of the site, and hence ends up being linked to by every page of the site, is clearly one of the more important pages. The relevance of those internal links is also a factor (the more relevant they are the better).

Topical relevance of inbound links

The relevance of the sites and pages linking to the target page and the target keyword is a big factor as well. Links from a site on an unrelated topic are still beneficial, but highly relevant links to a site count for a lot more.

Link popularity of site in topical community

This metric speaks to establishing the authority of your site. If you have many links from members of the topical community, that is a very strong vote of confidence for your site.

Keyword use in body text

Search engines also look at the text within the body of a page. Putting the relevant keyword in your title tag is good, but it adds to the impact if the on-page content backs it up. Don't limit the content to the exact target keyword; the search engines also look for synonyms to further reinforce the relevance of a page to a topic.

Global link popularity of sites that link to the site

Another ranking factor is the importance of the linking site. Is the site a trusted authority in its field? If so, this will probably show up in its Global Link Popularity.

Negative Ranking Factors

The SEOmoz survey also identified the top five negative factors:

Server is often inaccessible to crawlers

Search engines want their users to have good experiences. If your site is subject to frequent outages, by definition it is not providing a good user experience. So, if the search engine crawler frequently is unable to access your web pages, the search engine will assume that it is dealing with a low-quality site.

Content very similar to or duplicate of other web pages

Search engines want to present lots of unique content (reference our discussion of diversity in "Determining Searcher Intent and Delivering Relevant, Fresh Content" on page 41). If your website is very thin on content, or is largely a duplicate of the content on other websites, this will potentially be seen as a negative ranking signal.

We discuss duplicate content in more detail in "Duplicate Content Issues" on page 226 in Chapter 6.

External links to low-quality/spam sites

One way to hurt your site's ranking is to make a practice of linking to low-quality sites or spam sites. Generally speaking, one bad outbound link might not hurt you, but if you link to lots of poor-quality sites, it certainly can hurt your rankings.

What makes this more complex is that you might link to a perfectly good domain that goes out of business. Whoever picks up the domain next might not be someone you want to link to, but you may not know that the site you originally linked to is gone and has been replaced by a spam site.

Participation in link schemes or actively selling links

Search engines also do not want publishers to sell links on their site for the purposes of passing *link authority* (or *link juice*, as many people in the industry call it). The search engines don't want publishers participating in shady link schemes, such as distributing hit counters with hidden links within them.

Duplicate titles/meta tags on many pages

Since the title tag is a very powerful signal, failure to create a distinct title tag can be a negative ranking signal. Correspondingly, the meta description, and meta keywords tags duplicated en masse, could potentially be seen as a negative ranking factor.

Other Ranking Factors

The ranking factors we've discussed so far are really just the basics. Search engines potentially factor in many more signals. Some of these include:

Rate of acquisition of links

If over time your site has acquired an average of 5 links per day, and then the links suddenly start to come in at a rate of 10 per day, that could be seen as a positive ranking signal.

On the other hand, if the rate of new links drops to two per day, that could be a signal that your site has become less relevant. Additionally, if your site suddenly starts to get 30 new links per day, you have either become a lot more relevant or started to acquire links in a spammy way. The devil is in the details here, with one of the most important details being from where those new links are coming. The concept of considering temporal factors in link analysis is documented in a U.S. patent held by Google and found at the following URL:

http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u= %2Fnetahtml%2FPTO%2Fsrchnum.html&r=1&f=G&l=50&s1=%2220050071741%22 .PGNR.&OS=DN/20050071741&RS=DN/20050071741

Alternatively, you can look it up by searching for patent number 20050071741.

Usage data

Other factors can be interpreted as a ranking signal. For example, if a user clicks through on a SERP listing and comes to your site, clicks the Back button, and then clicks on another

result in the same set of search results, that could be seen as a negative ranking signal. Or if the results below you in the SERPs are getting clicked on more than you are, that could be seen as a negative ranking signal for you and a positive ranking signal for them. Whether search engines use this signal or not, or how much weight they might put on it, is not known. At best, this is a noisy ranking signal and any use is likely to be small in scope. Any use may be primarily related to personalization and relevance feedback.

User data

Personalization is one of the most talked about frontiers in search. There are a few ways personalization can take place. For one, a search engine can perform a geolocation lookup to figure out where you are approximately located. Based on this, a search engine can show results tailored to your current location. This is very helpful, for example, if you are looking for a local restaurant.

Another way a search engine can get some data on a user is if the user creates a profile with the search engine and voluntarily provides some information. A simple example would be a language preference. If the user indicates he prefers Portuguese, the search engine can tailor the results to that preference.

Search engines can also look at the search history for a given user. Basically, the search engine maintains a log of all the searches you have performed when you are logged in. Based on this, it can see that you have been checking out luxury cars recently, and can use that knowledge to tweak the results you see after you search on *jaguar*.

Google sandbox

As we have discussed throughout this chapter, the search engines use a number of methods to fight spam. One technique that many people believe Google uses became known as the Google "sandbox." The sandbox is thought to be a filter where Google limits the rate of growth of the PageRank (or rankings) of new domains.

This approach could be useful in filtering out spam domains because they often don't stay around very long, so the spammer works hard to get them ranking and producing traffic as quickly as they can. The sandbox can potentially create a scenario where the website is caught by improved algorithms or manual review prior to becoming highly productive. At a minimum, it would increase the cost of the spammer's efforts.

Using Advanced Search Techniques

One of the basic tools of the trade for an SEO practitioner is the search engines themselves. They provide a rich array of commands that can be used to perform advanced research, diagnosis, and competitive analysis. Some of the more basic operators are:

`[-keyword]`

Excludes the keyword from the search results. For example, [`loans -student`] shows results for all types of loans *except* student loans.

[+keyword]

Allows for forcing the inclusion of a keyword. This is particularly useful for including *stopwords* (keywords that are normally stripped from a search query because they usually do not add value, such as the word *the*) in a query, or if your keyword is getting converted into multiple keywords through automatic stemming. For example, if you mean to search for the TV show *The Office*, you would want the word *The* to be part of the query. As another example, if you are looking for Patrick Powers, who was from Ireland, you would search for *patrick +powers Ireland* to avoid irrelevant results for Patrick Powers.

["key phrase"]

Shows search results for the exact phrase—for example, ["seo company"].

[keyword1 OR keyword2]

Shows results for *at least one* of the keywords—for example, [google OR Yahoo!].

These are the basics, but for those who want more information, what follows is an outline of the more advanced search operators available from the search engines.

Advanced Google Search Operators

Google supports a number of advanced search operators (*http://www.google.com/help/operators .html*) that you can use to help diagnose SEO issues. Table 2-2 gives a brief overview of the queries, how you can use them for SEO purposes, and examples of usage.

TABLE 2-2. Google advanced search operators

Operator	Short description	SEO application	Examples
(site:)—Domain-restricted search	Narrows a search to (a) specific domain(s)/ directories	Shows approximately how many URLs are indexed by Google	site:www.google.com
		From a directory	site:mit.edu/research/
		Includes all subdomains	site:google.com
		Shows sites of a specific top-level domain (TLD)	site:org
(inuRL:)/(allinurl:)— URL keyword restricted search	Narrows the results to documents containing (a) search term(s) in the URLs	Find web pages having your keyword in a file path	inurl:seo inurl:company = allinurl:seo company

Operator	Short description	SEO application	Examples
`(intitle:)`/ `(allintitle:)`—Title keyword restricted search	Restricts the results to documents containing (a) search term(s) in a page title	Find web pages using your keyword in a page title	`intitle:seo intitle:company` `=` `allintitle:seo company`
`(inanchor:)`/ `(allinanchor:)`—Anchor text keyword restricted search	Restricts the results to documents containing (a) search term(s) in the anchor text of backlinks pointing to a page	Find pages having the most backlinks/the most powerful backlinks with the keyword in the anchor text	`inanchor:seo inanchor:company` `=` `allinanchor:seo company`
`(intext:)`—Body text keyword restricted search	Restricts the results to documents containing (a) search term(s) in the body text of a page	Find pages containing the most relevant/most optimized body text	`intext:seo`
`(ext:)`/`(filetype:)`— File type restricted search	Narrows search results to the pages that end in a particular file extension	A few possible extensions/file types: PDF (Adobe Portable Document Format) HTML or *.htm* (Hypertext Markup Language) *.xls* (Microsoft Excel) *.ppt* (Microsoft PowerPoint) *.doc* (Microsoft Word)	`filetype:pdf` `ext:pdf`
`(*)`—Wildcard search	Means "insert any word here"	Search for a phrase "partial match"	seo * directory returns "seo *free* directory", "seo*friendly* directory", etc.
`(Related:)`—Similar URLs search	Shows *related pages* by finding pages linking to	Evaluate how relevant the site's "neighbors" are	Compare: `related:www.linkdiy.com`

Operator	Short description	SEO application	Examples
	the site and looking at what else they tend to link to (i.e., "co-citation"); usually 25 to 31 results are shown.		and `related:www.alchemistmedia.com`
`(Info:)`—Information about a URL search	Gives information about the given page	Learn whether the page has been indexed by Google; provides links for further URL information; this search can also alert you to possible site issues (duplicate content or possible DNS problems)	`info:netconcepts.com` will show you the page title and description, and invite you to view its related pages, incoming links, and page cached version
`(Cache:)`—See what the page looked like when Google crawled it	Shows Google's saved copy of the page	Google's *text* version of the page works the same way as SEO Browser	`cache:www.stonetemple.com`
(~keyword)	Shows keywords Google thinks are related to *keyword*	Can be very useful in uncovering related words that you should include on your page about *keyword*	*~zoo ~trip* will show you keywords related to *zoo* and *trip*

N O T E

When using the site: operator, some indexed URLs might not be displayed (even if you use the "repeat the search with omitted results included" link to see the full list). The site: query is notoriously inaccurate. You can obtain a more accurate count of the pages of your site indexed by Google by appending &start=990&filter=0 to the URL of a Google result set for a site: command.

This tells Google to start with result 990, which is the last page Google will show you since it limits the results to 1,000. This must take place in two steps. First, enter a basic site:*yourdomain*.com command, and then get the results. Then go up to the address bar and append the &start=990&filter=0 parameters to the end of the URL. Once done with this, you can look at the total pages returned to get a more accurate count.

To see more results, you can also use the following search patterns:

- [site:yourdomain.com/subdirectory1] + [site:yourdomain.com/subdirectory2] + etc. (the "deeper" you dig, the more/more accurate results you get)
- [site:yourdomain.com inurl:keyword1] + [site:yourdomain.com inurl:keyword2] + etc. (for subdirectory-specific keywords)
- [site:yourdomain.com intitle:keyword1] + [site:yourdomain.com intitle:keyword2] + etc. (for pages using the keywords in the page title)

Combined Google queries

To get more information from Google advanced search, it helps to learn how to effectively combine search operators. Table 2-3 illustrates which search patterns you can apply to make the most of some important SEO research tasks.

TABLE 2-3. Combined Google search operators

What for	Description	Format	Example
Competitive analysis	Search who mentions your competitor; use the date range operator within Google advanced search to find the most recent brand mentions; the following brand-specific search terms can be used: [domainname.com], [domain name], [domainname], [site owner name], etc.	[domainname.com - site:domainname.com] (+ add &as_qdr=d [past one day] to the query string); use d3 for three days, m3 for three months, etc.	seomoz -site:seomoz.org during past 24 hours
Keyword research	Evaluate the given keyword competition (sites that apply proper SEO to target the term)	[inanchor:*keyword* intitle:*keyword*]	inanchor:seo intitle:seo

What for	Description	Format	Example
	Find more keyword phrases	[key * phrase]	free * tools
SEO site auditing	Learn whether the site has canonicalization problems	[site:domain.com -inurl:www]	site:iceposter.com -inurl:www
	Find the site's most powerful pages	[www site:domain.com]	www site:alchemistmedia.com
		[tld site:domain.tld]	org site:netconcepts.com
		[inurl:domain site:domain.com]	inurl:stonetemple site:stonetemple.com
		[domain site:domain.com]	alchemistmedia site:alchemistmedia.com
	Find the site's most powerful page related to the keyword	[site:domain.com keyword]	site:netconcept.com seo
		[site:domain.com intitle:keyword]	site:stonetemple.com intitle:seo
		[site:domain inanchor:keyword]	site:seomoz.org inanchor:seo
Link building	Find authority sites offering a backlink opportunity	[site:org bookmarks/links/"favorite sites"/]	site:org donors
		[site:gov bookmarks/links/"favorite sites"/]	
		[site:edu bookmarks/links/"favorite sites"/]	
	Search for relevant forums and discussion boards to participate in discussions and probably link back to your site	[inurl:forum OR inurl:forums keyword]	inurl:forum OR inurl:forums seo

Firefox plug-ins for quicker access to Google advanced search queries

You can use a number of plug-ins with Firefox to make accessing these advanced queries easier:

- Advanced Dork, for quick access to (intitle:), (inurl:), (site:), and (ext:) operators for a highlighted word on a page, as shown in Figure 2-24

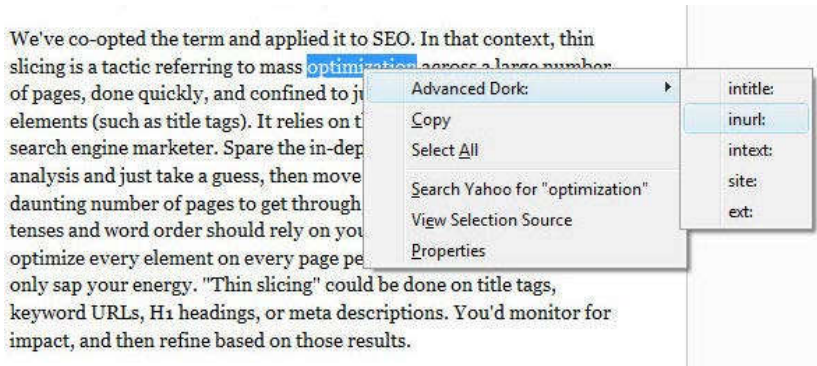- SearchStatus, for quick access to a (site:) operator to explore a currently active domain, as shown in Figure 2-25
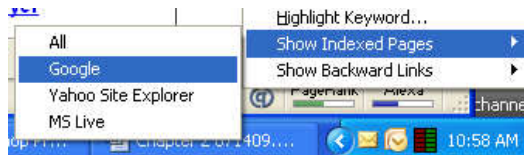

FIGURE 2-24. Advanced Dork plug-in for Firefox

FIGURE 2-25. SearchStatus plug-in for Firefox

Yahoo! Advanced Search Operators

The most popular of Yahoo!'s SEO advanced search operators is the one that checks the page or domain backlinks. Since Bing dropped its inbound link-checking ability except through its API, and while Google allows for only partial reports except for verified sites within Webmaster Central, Yahoo! has remained the main source for page backlink information.

Other Yahoo! queries are mostly used to enhance (link:) and linkdomain: searches. Table 2-4 lists the most essential Yahoo! advanced operators (that differ from the ones detailed for Google) and samples of their usage in SEO. Note, however, that it is unclear which of these will be supported in the future due to the announced plan for Yahoo! search to be powered

by Microsoft's Bing (*http://topics.nytimes.com/top/news/business/companies/yahoo_inc/yahoo-microsoft-deal/index.html*).

TABLE 2-4. Yahoo! advanced search operators

Operator	Short description	SEO application	Examples
`(Link:)`—URL backlinks restricted search	Finds any *page* backlinks (if used in isolation, redirects to Yahoo! Site Explorer)	Find most the relevant backlinks of any page	`link:http://www.seomoz.org/blog seo`
`(LinkDomain:)`—Domain backlinks restricted search	Finds any *domain* backlinks (if used in isolation, redirects to Yahoo! Site Explorer)	Find the most relevant backlinks of any domain	`linkdomain:www.alchemistmedia.com seo`
`(OriginUrlExtension:)`— File type restricted search `(Region:)`—Area restricted search	Narrows search results to the pages that end in a particular file extension Narrows search results to the sites belonging to a particular territory	A few possible extensions/file types: PDF (Adobe Portable Document Format) HTML or *.htm* (Hypertext Markup Language) *.xls* (Microsoft Excel) *.ppt* (Microsoft PowerPoint) *.doc* (Microsoft Word)	`originurlextension:pdf`
		Available regions: `region:europe` `region:africa` `region:asia` `region:centralamerica` `region:downunder` `region:mediterranean` `region:mideast` `region:northamerica` `region:southamerica` `region:southeastasia`	`region:europe`

Operator	Short description	SEO application	Examples
(Url:)—URL restricted search	Finds a specific document in Yahoo!'s index	Learn whether the page was indexed by Yahoo!	`url:http://www.netconcept.com`

Combined Yahoo! Queries

Yahoo! also allows you to created combined search queries. Table 2-5 shows some of the more unique ways to do this.

TABLE 2-5. Combined Yahoo! search operators

What for	Description	Format	Example
Keyword research	Explore territory-specific usage of your keyword	*keyword* region:europe	`seo region:europe`
SEO site auditing	Search for the best optimized page of the site	intitle:*keyword* inurl:*keyword* site:*www.site.org*	`intitle:seo inurl:seo site:www.stonetemple.com`
Link building	Find .edu or .gov sites linking to the page	link:http:// *www.domain.com* site:.edu OR site:.gov	`link:http://www.amazon.com site:.edu OR site:.gov`
	Find geospecific backlinks of a page/domain	link:http:// *www.domain.com* region:europe	`link:http://www.amazon.com region:europe`

Firefox extensions to help with Yahoo! (Link:) and (LinkDomain:) search operators

A couple of Firefox plug-ins help make it a bit easier to access these advanced Yahoo! search operators:

- SearchStatus, to show currently active page and domain backlinks in Yahoo!, as shown in Figure 2-26
- SEO for Firefox Toolbar Extension, which will show you detailed backlink information directly in your search results (see Figure 2-27)
- Seoquake Firefox Extension, which can be used to access Yahoo! backlink data (and a lot more) right from the search results page or from the currently active page (Seoquake), as shown in Figure 2-28
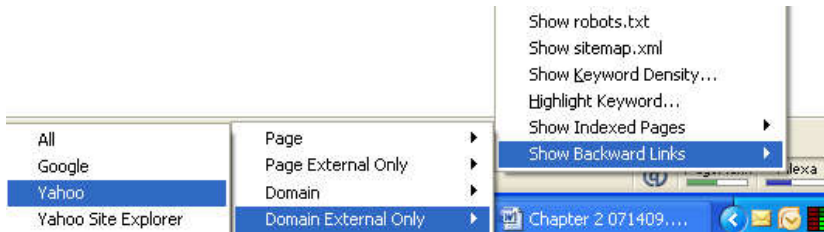
FIGURE 2-26. SearchStatus getting external backlinks from Yahoo!

FIGURE 2-27. SEO for Firefox Toolbar Extension

FIGURE 2-28. Seoquake backlink data

Bing Advanced Search Operators

Although Bing is not as widely used for SEO purposes, it can still help you better understand
your site's SEO issues or effectively find valuable SEO data. Besides, Bing offers several unique
search operators worth looking into, as shown in Table 2-6.

TABLE 2-6. Bing advanced operators

Operator	Short description	SEO application	Examples
`(LinkFromDomain:)`—Domain outbound links restricted search	Finds all pages the given domain links out to	Find most relevant sites your competitor links out to	`linkfromdomain:seomoz.org seo`
`(Contains:)`—File type restricted search	Narrows search results to pages linking to a document of the specified file type	Find pages linking to a specific document type containing relevant information	`contains:wma seo`
`(IP:)`—IP address restricted search	Shows sites sharing one IP address	`ip:xxx.xxx.xxx.xxx`	`ip:207.182.138.245`
`(InBody:)`—Body text keyword restricted search	Restricts the results to documents containing query word(s) in the body text of a page	Find pages containing the most relevant/best optimized body text	`inbody:seo` (equivalent to Google's `intext:`)
`(Location:)/(loc:)`—Location specific search	Narrows search results to a specified location (multiple location options can be found under Bing's advanced search)	Find geospecific documents using your keyword	`seo loc:AU`
`(Feed:)`—Feed keyword restricted search	Narrows search results to terms contained in RSS feeds	Find relevant feeds	`feed:seo`
`(Hasfeed:)`—Feed keyword restricted search	Narrows search results to pages linking to feeds that contain the specified keywords	Find pages linking to relevant feeds	`hasfeed:seo site:cnn.com`

More Advanced Search Operator Techniques

You can also use more advanced SEO techniques to extract more information.

Keyword difficulty

When building a web page, it can be useful to know how competitive the keyword is that you are going after, yet this can be difficult to obtain. The `intitle:` operator shows pages that are more focused on your search term than the pages returned without that operator (e.g., `intitle:"dress boots"`).

You can use different ratios to give you a sense of how competitive a keyword market is (higher results mean that it is more competitive). For example:

> dress boots (20,900,000) versus "dress boots" (424,000) versus `intitle:"dress boots"` (37,000)
> Ratio: 20,900/37 = 565:1
> Exact phrase ratio: 424/37 = 11:1

Another significant parameter you can look at is the `inanchor:` operator; for example, `inanchor:"dress boots"`. You can use this operator in the preceding equation instead of the `intitle:` operator.

Using number ranges

The number range operator can help restrict the results set to a set of model numbers, product numbers, price ranges, and so forth. For example:

> `site:stevespanglerscience.com "product/1700..1750"`

Unfortunately, the number range combined with `inurl:` is not supported. So, the product number must be on the page. The number range operator is also great for copyright year searches (to find abandoned sites to acquire). Combine it with the `intext:` operator to improve the signal-to-noise ratio; for example, `intext:"copyright 1993..2005" -2008 blog`.

Advanced doc type search

The `filetype:` operator is useful for looking for needles in haystacks. Here are a couple of examples:

> `confidential business plan -template filetype:doc`
> `forrester research grapevine filetype:pdf`

Determine listing age

You can label results with dates that give a quick sense of how old (and thus trusted) each listing is; for example, by appending the `&as_qdr=m199` parameter to the end of a Google SERP URL, you can restrict results to those within the past 199 months.

Uncover subscriber-only or deleted content

You can get to subscriber-only or deleted content from the Cached link in the listing in the SERPs or by using the `cache:` operator. Don't want to leave a footprint? Add `&strip=1` to the end of the Google cached URL. Images on the page won't load.

If no Cached link is available, use Google Translate to take your English document and translate it from Spanish to English (this will reveal the content even though no Cached link is available):

```
http://translate.google.com/translate?prev=&hl=en&u=URL-GOES-HERE&sl=es&tl=en
```

Identify neighborhoods

The `related:` operator will look at the sites linking (the "Linking Sites") to the specified site, and then see which other sites are commonly linked to by the Linking Sites. This list is usually limited to between 25 and 31 results. These are commonly referred to as *neighborhoods*, as there is clearly a strong relationship between sites that share similar link graphs.

Find Creative Commons (CC) licensed content

Use the `as_rights` parameter in the URL to find Creative Commons licensed content. Here are some example scenarios to find CC-licensed material on the Web:

Permit commercial use
```
http://google.com/search?as_rights=(cc_publicdomain|cc_attribute|cc_sharealike|
cc_nonderived).-(cc_noncommercial)&q=KEYWORDS
```

Permit derivative works
```
http://google.com/search?as_rights=(cc_publicdomain|cc_attribute|cc_sharealike|
cc_noncommercial).-(cc_nonderived)&q=KEYWORDS
```

Permit commercial and derivative use
```
http://google.com/search?as_rights=(cc_publicdomain|cc_attribute|cc_sharealike).-
(cc_noncommercial|cc_nonderived)&q=KEYWORDS
```

Make sure you replace *KEYWORDS* with the keywords that will help you find content that is of relevance to your site. The value of this to SEO is an indirect one. Creative Commons content can potentially be a good source of content for a website.

Vertical Search Engines

Vertical search is the term people sometimes use for specialty or niche search engines that focus on a limited data set (as already mentioned, Google calls them onebox results). Examples of vertical search solutions provided by the major search engines are image, video, news, and blog searches. These may be standard offerings from these vendors, but they are distinct from the engines' general web search functions.

Vertical search results can provide significant opportunities for the SEO practitioner. High placement in these vertical search results can equate to high placement in the web search results, often above the traditional 10 blue links presented by the search engines.

Vertical Search from the Major Search Engines

The big three search engines offer a wide variety of vertical search products. Here is a partial list:

Google
> Google Maps, Google Images, Google Product Search, Google Blog Search, Google Video, Google News, Google Custom Search Engine, Google Book Search, Google US Gov't Search, etc.

Yahoo!
> Yahoo! News, Yahoo! Local, Yahoo! Images, Yahoo! Video, Yahoo! Shopping, Yahoo! Audio Search, etc.

Bing
> Bing Image, Bing Video, Bing News, Bing Maps, Bing Health, Bing Products, etc.

Image search

All three search engines offer image search capability. Basically, image search engines limit the data that they crawl, search, and return in results to images. This means files that are in GIF, TIF, JPG, and other similar formats. Figure 2-29 shows the image search engine from Bing.
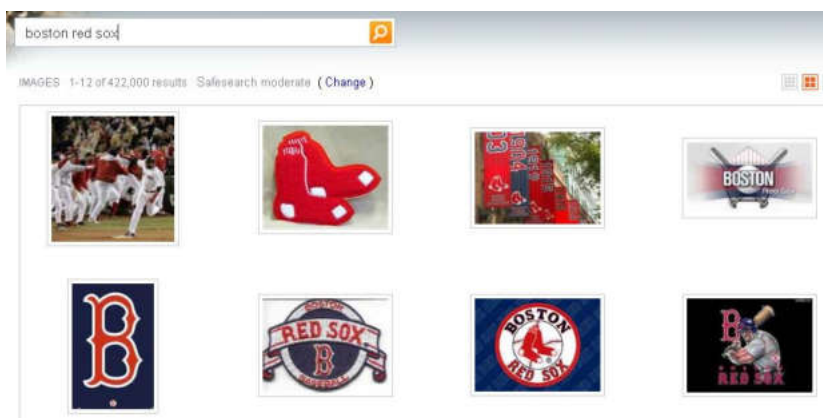
FIGURE 2-29. Image search results from Bing

Image search engines get a surprisingly large number of searches performed on them. According to comScore, more than 1 billion image searches were performed in October 2008, or a little more than 8.3% of all searches performed in that month. Similar data from Nielsen

Online shows image search comprised 6.0% of all search in January 2009. However, since an image is a binary file, it cannot be readily interpreted by a search engine crawler.

The search engine has to rely on text surrounding the image, the `alt` attribute within the `img` tag, and the image filename. Optimizing for image search is its own science, and we will discuss it in more detail in "Optimizing for Image Search" on page 346 in Chapter 8.

Video search

As with image search, video search engines focus on searching specific types of files on the Web, in this case video files, such as MPEG, AVI, and others. Figure 2-30 shows a quick peek at video search results from YouTube.
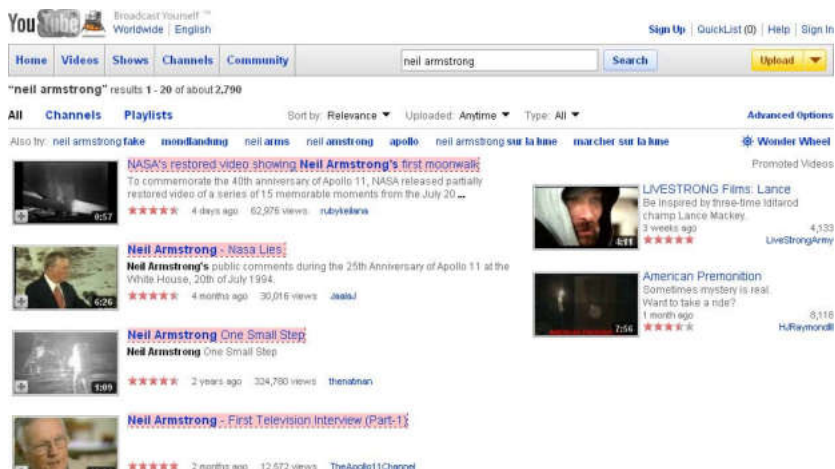
FIGURE 2-30. Video search results from YouTube

A very large number of searches are also performed in video search engines. Hitwise and comScore data shows approximately 125 million searches performed on video search on the major search engine properties (e.g., *http://video.google.com*, *http://video.yahoo.com*, and *http://video.bing.com*) in October 2008, and then this number balloons to 2.6 billion searches once you include YouTube (*http://www.youtube.com*), which has become the #2 search engine on the Web.

There is significant traffic to be gained by optimizing for video search engines and participating in them. Once again, these are binary files and the search engine cannot easily tell what is inside them.

This means optimization is constrained to data in the header of the video and on the surrounding web page. We will discuss video search optimization in more detail in "Others: Mobile, Video/Multimedia Search" on page 366 in Chapter 8.

However, each search engine is investing in technology to analyze images and videos to extract as much information as possible. For example, OCR technology is being used to look for text within images, and other advanced technologies are being used to analyze video content. Flesh-tone analysis is also in use to detect porn or recognize facial features. The application of these technologies is in its infancy, and is likely to evolve rapidly over time.

News search

News search is also unique. News search results operate on a different time schedule. News search results have to be very, very timely. Few people want to read the baseball scores from a week ago when several other games have been played since then.

News search engines must be able to retrieve information in real time and provide near instantaneous responses. Modern consumers tend to want their news information now. Figure 2-31 is a quick look at the results from a visit to Yahoo! News.
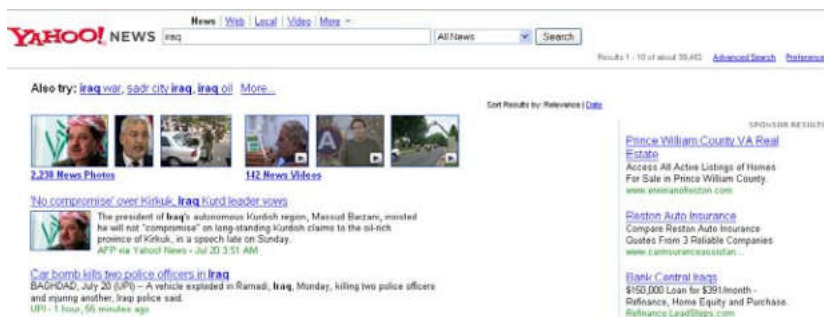
FIGURE 2-31. News search results from Yahoo!

As with the other major verticals, there is a lot of search volume here as well. To have a chance of receiving this volume, you will need to become a news source. This means timely, topical news stories generated on a regular basis. There are other requirements as well, and we will discuss them further in "Optimizing for News, Blog, and Feed Search" on page 355 in Chapter 8.

Local search/maps

Next up in our hit parade of major search verticals is local search (a.k.a. map search). Local search engines search through databases of locally oriented information, such as the name, phone number, and location of local businesses around the world, or just provide a service, such as offering directions from one location to another. Figure 2-32 shows Google Maps local search results.
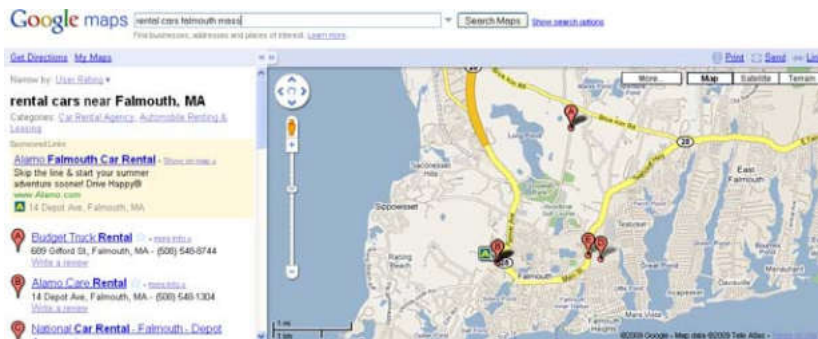
FIGURE 2-32. Local search results from Google

The integration of local search results into regular web search results has dramatically increased the potential traffic that can be obtained through local search. We will cover local search optimization in detail in "Optimizing for Local Search" on page 338 in Chapter 8.

Blog search

Google has implemented a search engine focused just on blog search called Google Blog Search (misnamed because it is an RSS feed engine and not a blog engine). This search engine will respond to queries, but only search blogs (more accurately, feeds) to determine the results. Figure 2-33 is an example search result for the search phrase *barack obama*.
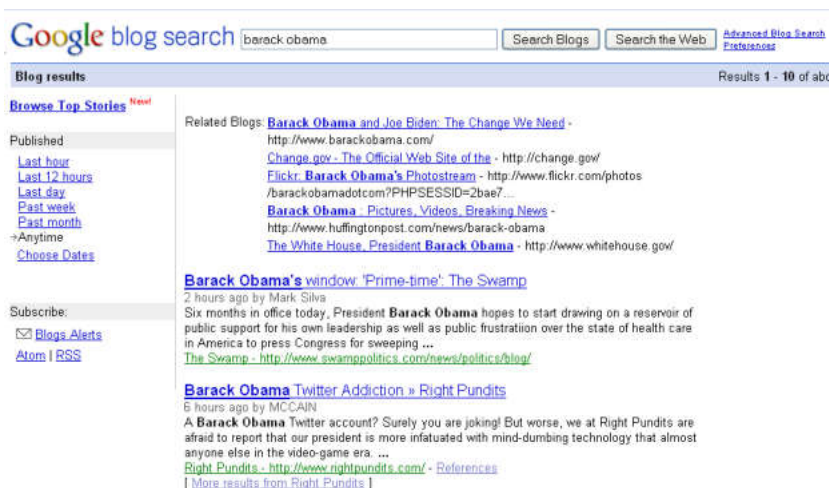
FIGURE 2-33. Results from Google Blog Search

We explore the subject of optimizing for Google Blog Search in "Optimizing for News, Blog, and Feed Search" on page 355 in Chapter 8.

Book search

The major search engines also offer a number of specialized offerings. One highly vertical search engine is Google Book Search, which specifically searches only content found within books, as shown in Figure 2-34.
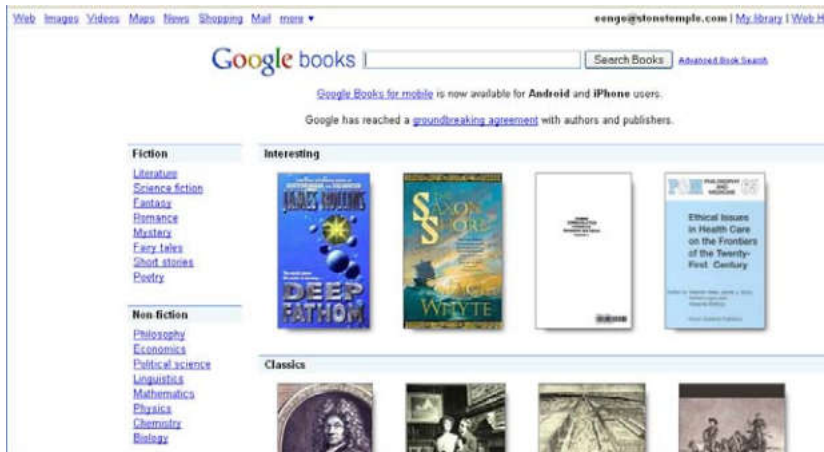

FIGURE 2-34. Google Book Search

Job search

Yahoo! also has a number of vertical search products. Yahoo! hotjobs is an example of a product designed to allow people to search for jobs (see Figure 2-35).

Celebrity xRank

Microsoft also has some unique vertical search properties. One of the more interesting ones is Celebrity xRank, which offers data on celebrity rankings and search trends, as shown in Figure 2-36.

Universal Search/Blended Search

Google made a big splash in 2007 when it announced Universal Search. This was the notion of integrating images, videos, and results from other vertical search properties directly into the main web search results.
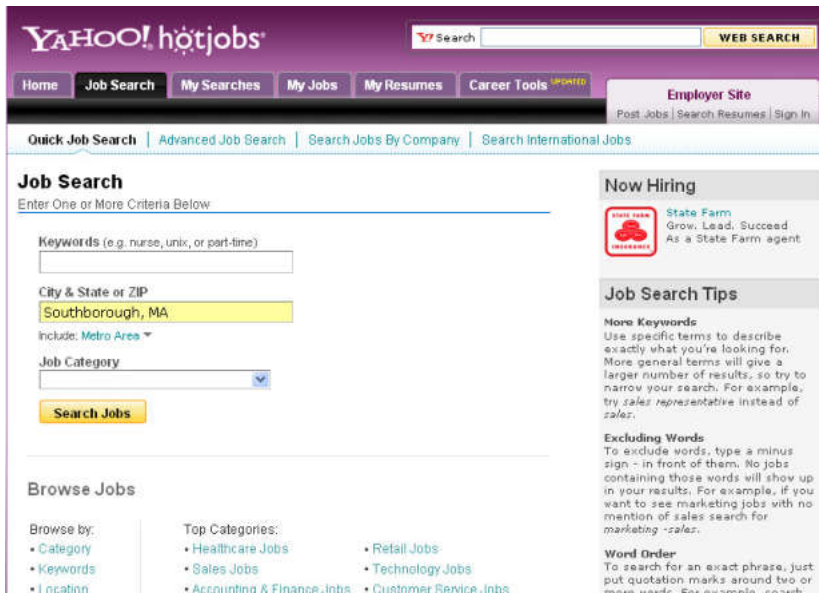
FIGURE 2-35. Yahoo! hotjobs

This quickly became a huge driver of traffic to sites because of their images, videos, news, local search information, and more. The other search engines quickly followed suit and began offering vertical search integration before 2007 was over. People now refer to this general concept as Blended Search (since Universal Search is specifically associated with Google). A look at some Universal Search results from Google can help illustrate the concept (see Figure 2-37).

Note the news results, along with an image at the very top of the results, along with more image results farther down. This information is coming from Google's news search index. If you look farther down in the search results, you will continue to see more vertical results, including video results and a timeline (see Figure 2-38).

A wide range of vertical data sets have been integrated into Google's Universal Search, as well as into the Blended Search results of the other search engines. In addition to the preceding examples, you can also see images, videos, and local data integrated into the traditional web search results.

The advent of Blended Search has significantly increased the opportunity for publishers with matching vertical data sets (such as a rich music library) to gain significant additional traffic to their sites by optimizing these data sets for the appropriate vertical search.
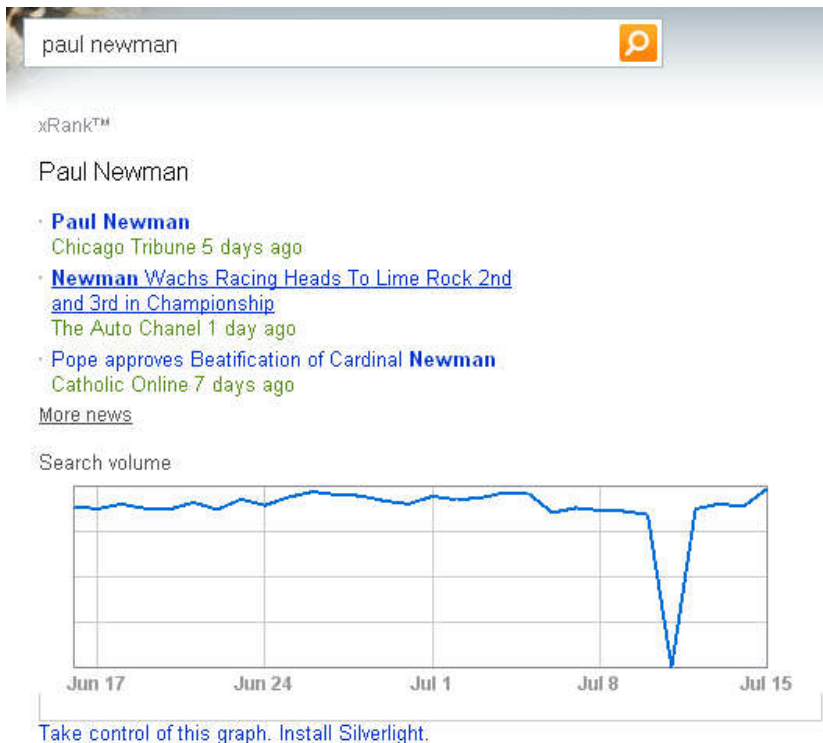
FIGURE 2-36. Bing xRank search

Meta search

Meta search engines are search engines that aggregate results from multiple search engines and present them to the user. The two most well-known ones are MetaCrawler and Dogpile. However, their cumulative search volume is quite small, and these do not factor into SEO strategies.

More specialized vertical search engines

Vertical search can also come from third parties. Here are some examples:

- Comparison shopping engines, such as PriceGrabber, Shopzilla, and NexTag
- Travel search engines, such as Expedia, Travelocity, Kayak, and Uptake
- Real estate search engines, such as Trulia and Zillow
- People search engines, such as Spock and Wink
- Job search engines, such as Indeed, CareerBuilder, and SimplyHired
- Music search engines, such as iTunes Music Store

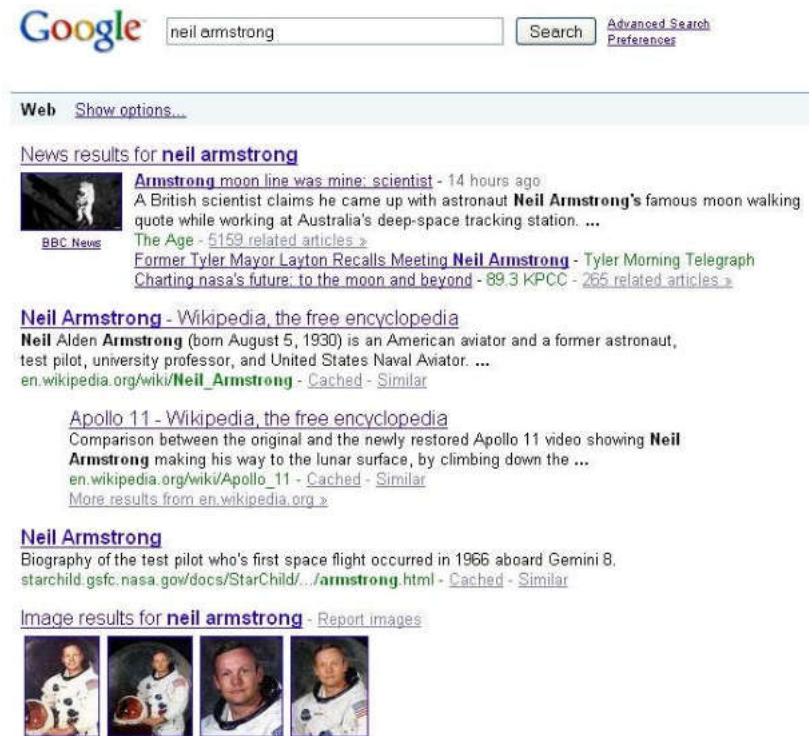- **B2B search engines**, such as Business.com, KnowledgeStorm, Kellysearch, and ThomasNet

FIGURE 2-37. Google Universal Search results

In addition, some companies offer products that allow anyone to build his own search engine, such as Google's Custom Search Engines, Eurekster, and Rollyo. Also, specialty search engines are offered by the major search engines not covered in this section.

There is an enormous array of different vertical search offerings from the major search engines, and from other companies as well. It is to be expected that this explosion of different vertical search properties will continue.

Effective search functionality on the Web is riddled with complexity and challenging problems. Being able to constrain the data types (to a specific type of file, a specific area of interest, a specific geography, or whatever) can significantly improve the quality of the results for users.

Video results for **neil armstrong**

First Moon Landing 1969
1 min 44 sec
www.youtube.com

Neil Armstrong, 60 Minutes
Interview
14 min
video.google.com

Astronaut Bio: **Neil Armstrong**
Neil Armstrong (NASA Photo) **Neil** A. **Armstrong** NASA Astronaut (former). PERSONAL
DATA: Born August 5, 1930 in Wapakoneta, Ohio. Married. Two sons. **...**
www.jsc.nasa.gov/Bios/htmlbios/**armstrong**-na.html - Cached - Similar

NASA - **Armstrong, Neil**
Born in 1930, **Neil** A. **Armstrong**, a United States astronaut, was the first person to set foot
on the moon. On July 20, 1969, **Armstrong** and Buzz Aldrin landed **...**
www.nasa.gov/worldbook/**armstrong_neil**_worldbook.html - Cached - Similar

NASA - Biography of **Neil Armstrong**
Neil A. **Armstrong**, the first man to walk on the moon, was born in Wapakoneta, Ohio, on
August 5, 1930. He began his NASA career in Ohio. **...**
www.nasa.gov/centers/glenn/about/bios/**neil**abio.html - Cached - Similar

Neil Armstrong Biography - Biography of **Neil Armstrong** - First Man **...**
Born on August 5, 1930 in Auglaize County, OH, **Neil Armstrong** was the eldest of 3 children.
Neil Armstrong worked at numerous jobs around town and at the **...**
space.about.com/od/astronautbiographies/.../**neilarmstrong**.htm - Cached - Similar

Never-before seen photo shows **Neil Armstrong's** face as he first **...**
Jul 22, 2009 **...** An amazing new photograph showing **Neil Armstrong's** face through his
space suit visor has come to light.
www.dailymail.co.uk/.../Never-seen-photo-shows-**Neil-Armstrongs**-face-walks-moon.html - 57
minutes ago - Similar

Timeline results for **neil armstrong**
1969   On **July 20, 1969**, Apollo 11 astronauts **Neil Armstrong** and Edwin "Buzz" Aldrin climbed
       inside their lunar lander, undocked from the command module **...**
       books.google.com
1969   On **July 20, 1969**, **Neil Armstrong** became the first human being to set foot on the moon.
       This book tells the story of **Armstrong's** life, with details **...**
       www.rif.org

FIGURE 2-38. More Google Universal Search results

Country-Specific Search Engines

At this stage, search is truly global in its reach. Google is the dominant search engine in many countries, but not all of them. How you optimize your website depends heavily on the target market for that site, and the search engine(s) that is (are) the most important in that market.

According to comScore, Google is receiving 62% of all searches performed worldwide. In addition, Google is the market share leader in every regional market. Most notable, though, is the Asia Pacific region, where Google holds a relatively narrow 36.1% to 26.8% edge over Baidu, the largest search engine in China. This is the only regional market in which Google

has less than 60% market share, and it also happens to be the largest market for search in the world (in terms of total searches performed).

Here is some data on countries where other search engines are the major players:

China

Baidu News reported in February 2009 that Baidu had more than 73% market share in China in 2008 (*http://baidu-news.blogspot.com/2009/02/baidus-market-share-at-732-domestic.html*).

Russia

According to figures announced by Yandex, the company's market share in Russia comprises about 50% of all searches (*http://www.searchengineworld.com/seo/3459183.htm*).

South Korea

Naver was estimated by the *International Herald Tribune* (*http://www.iht.com/articles/2007/07/04/business/naver.php*) to have about 77% market share in South Korea in mid-2007.

Czech Republic

The StartupMeme Technology blog reported that Seznam was reported to have more than 60% market share in Czechoslovakia in mid-2008 (*http://startupmeme.com/google-to-acquire-czech-search-engine-seznam-for-1-billion/*).

Optimizing for Specific Countries

One of the problems international businesses continuously need to address with search engines is identifying themselves as "local" in the eyes of the search engines. In other words, if a search engine user is located in France and wants to see where the wine shops are in Lyons, how does the search engine know which results to show?

Here are a few of the top factors that contribute to international ranking success:

- Owning the proper domain extension (e.g., .com.au, .co.uk, .fr, .de, .nl) for the country that your business is targeting
- Hosting your website in the country you are targeting (with a country-specific IP address)
- Registering with local search engines:
  — Google: *http://www.google.com/local/add/*
  — Yahoo!: *http://listings.local.yahoo.com/csubmit/index.php*
  — Bing: *https://ssl.bing.com/listings/BusinessSearch.aspx*
- Having other sites from the same country link to you
- Using the native language on the site (an absolute requirement for usability)
- Placing your relevant local address data on every page of the site
- Defining your preferred region in Google Webmaster Tools

All of these factors act as strong signals to the search engines regarding the country you are targeting, and will make them more likely to show your site for relevant local results.

The complexity increases when targeting multiple countries. We will discuss this in more depth in "Best Practices for Multilanguage/Country Targeting" on page 275 in Chapter 6.

Profile of China's Internet Usage

Based on the data that China released in July 2008, China has more Internet users than the United States, with 253 million users. Here is a summary of data from the Chinese report:

- China has the most Internet users in the world (see Figure 2-39).

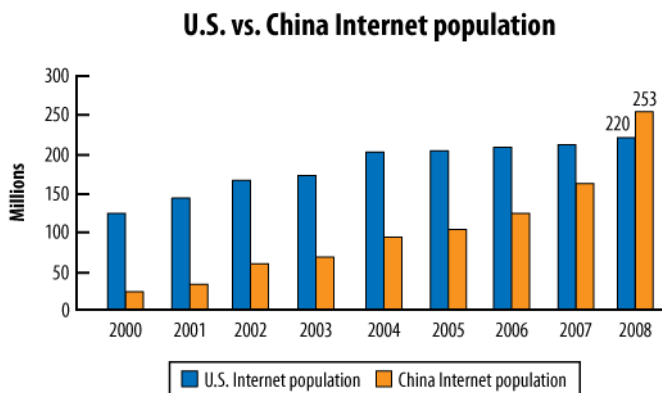

FIGURE 2-39. U.S. versus China Internet users

- China has the most broadband users in the world (see Figure 2-40). *"This report, the 22nd Statistical Report on the Internet Development in China, also indicates the number of broadband users has reached 214 million, which also tops the world."*
- China's Internet penetration rate continues to grow and grow and grow (see Figure 2-41):
  - U.S. Internet usage has hovered around a 70% penetration rate in the past five years, while Chinese Internet penetration has jumped from 7% to almost 20% in the same period.
  - China could plausibly reach a similar penetration rate to the United States within 20 years.
- Mobile phone use in China continues to grow:
  - China has *601 million* mobile phone users, according to the latest government report.
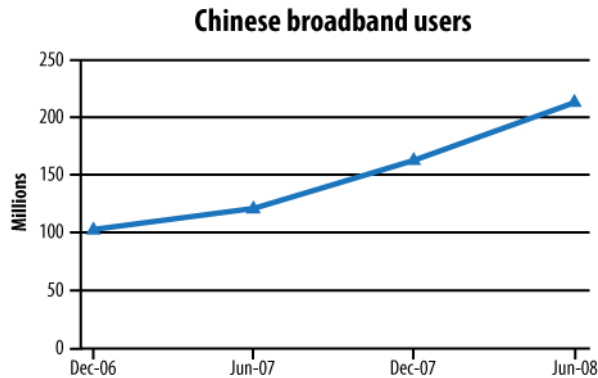 - From January 2008 to June 2008, there were *53.3 million* new mobile phone users.
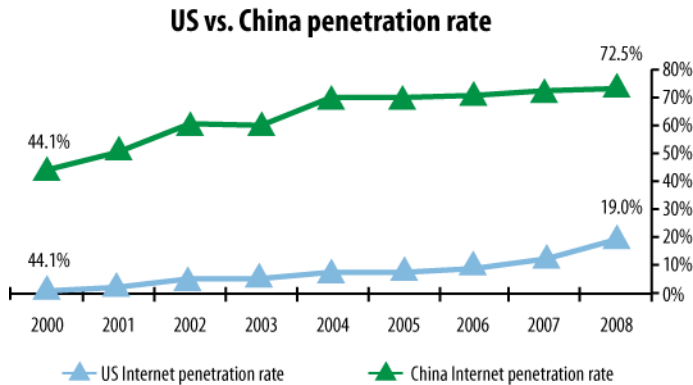
FIGURE 2-40. Chinese broadband users

FIGURE 2-41. U.S. versus China Internet penetration

— One carrier, China Mobile, has more than 414 million mobile subscribers and is ranked #1 in the world (*http://en.wikipedia.org/wiki/List_of_mobile_network_operators*).

— Only *12%* of these users have accessed the Internet from a mobile phone because of the lack of a proper 3G network (none of the Chinese telcos have a 3G license); an estimated 73 million have accessed the Internet from a mobile phone.

— However, the Chinese government is moving aggressively to issue 3G licenses to the major telcos.

Conclusion

Understanding how search engines work is an important component of SEO. The search engines are constantly tuning their algorithms. For that reason, the successful SEO professional is constantly studying search engine behavior and learning how they work.