

Unit 1

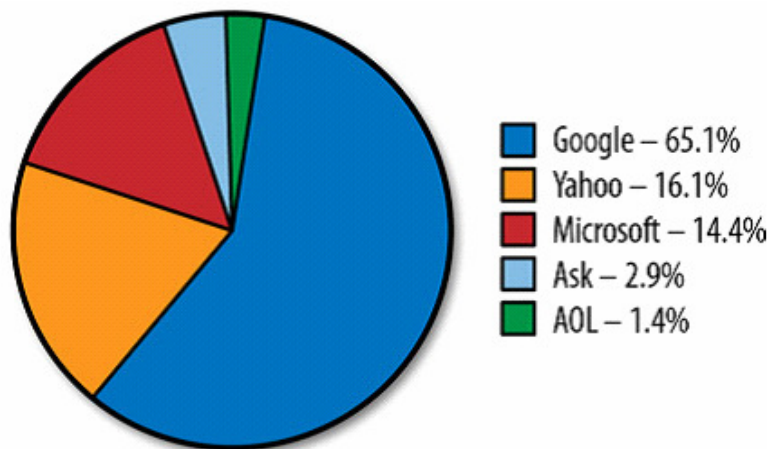
The Mission of Search Engines

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 constructing PR and marketing campaigns.

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. 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 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 spamming efforts by the search engines, as this would carry the risk of resulting in penalties for the websites they work on.

The Market Share of Search Engines

Following table shows the US market share for search engines in July 2011, 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 an 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 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. Most search processes comprise the following steps:

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).
2. Formulate that need in a string of words and phrases (the query). Most people formulate their queries in one to three words.

NATIONAL COMPUTER COLLEGE

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

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 understands 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 browser's address bar, clicking on a bookmark, or picking a link on your start page to go to a website; it is not the same as 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.

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 a site as well as search engine compatibility. Data from comScore provides some great insight into the types of things that people tend to search for. Following table shows a breakdown of many of the major categories that people's Internet searches fall into, based on comScore data for August 2011.

Parent category name	Total searches
Directories/Resources	2,789,625,911
Entertainment	1,750,928,801
Retail	1,686,123,715
Services	1,288,400,837
Conversational Media	837,067,182
Community	653,405,269
Travel	462,129,796
Health	435,860,663
News/Information	421,756,642
Sports	297,503,391

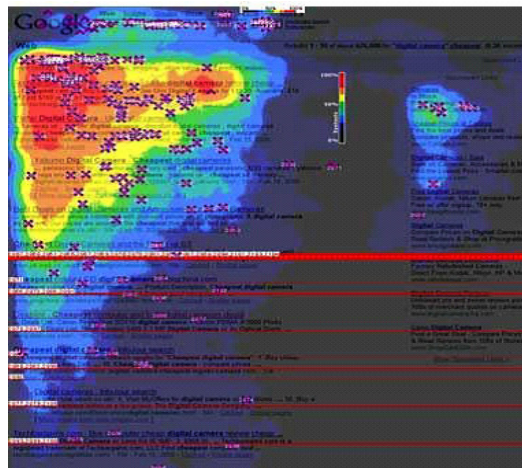
How Search Engines Drive Commerce on the Web

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. Ecommerce sales reported by the US Census Bureau were a healthy \$47.5 billion in the second quarter of 2011. 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 Google study from 2011 showed that each \$1 of online ad spend drives anywhere from \$4 to \$15 in offline sales.

Eye Tracking: How Users Scan Results Pages

Research firms Enquiro, Eyetools, and Didit conducted heat-map testing with search engine users that produced fascinating results related to what users see and focus on when engaged in search activity. Following Figure 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.

NATIONAL COMPUTER COLLEGE



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-left corner of the screen, then moving down vertically through the first two or three results, across the page to the first paid page result, down another few vertical results, and then across again to the second paid result. In May 2008, Google introduced the notion of Universal Search. This was a move from simply showing the 10 most relevant web pages 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, Following Figure shows what the eye-tracking pattern on a Blended Search page looks like.



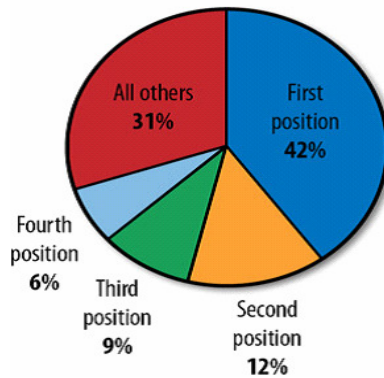
Click Tracking: How Users Click on Results, Natural Versus Paid

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 contrast, 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

NATIONAL COMPUTER COLLEGE

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 righthand column. The AdGooroo survey showed that having a prominent paid ad on the same search results page makes your #1 natural ranking receive 20% more clicks.

CTR by SERP position



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 search 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

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.

Horizontal navigation

The search engines also offer other types of navigation elements.

Search query box

All of the engines show the query you've performed and allow you to edit that query or enter a new query from the search results page next to the search query box.

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. *PPC (a.k.a. paid search) advertising* Companies purchase text ads from either Google AdWords or Microsoft adCenter. The results are ordered by a variety of factors, including relevance and bid amount.

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

Query refinements are offered by Google, Bing, and Yahoo!. The goal of these links is to let users search with a more specific and possibly more relevant query that will satisfy their intent.

Shopping search results

All three search engines do this as well. Shopping results incorporate offers from merchants into the results so that searchers that are looking to buy something can do so quite easily.

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 and Bing operate, and does not necessarily apply to other search engines that are popular, such as Yandex (Russia), Baidu (China), Seznam (Czechoslovakia), and Naver (Korea).

- **Crawling and Indexing**

To offer the best possible results, search engines must attempt to discover all the public pages on the World Wide Web and then present the ones that best match up with the user's search query. The first step in this process is **crawling** the Web. The search engines start with a seed set of sites that are known to be very high quality sites, and then visit the links on each page of those sites to discover other web pages. The link structure of the Web serves to bind together all of the pages that have been made public as a result of someone linking to them. Through links, search engines' automated robots, called *crawlers* or *spiders*, can reach the many billions of interconnected documents. This process repeats over and over again until the crawling process is complete. This process is an enormously complex one as the Web is a large and complex place. Search engines do not attempt to crawl the entire Web every day. In fact, they may become aware of pages that they choose not to crawl because they are not likely to be important enough. *Bing result for Megan Fox videos* to return in a search result. We will discuss the role of *importance* in the next section, "Retrieval and Rankings". Once the engines have retrieved a page during a crawl, 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. The first step in this process is to build a dictionary of terms. This is a massive database that catalogs all the significant terms on each page crawled by a search engine. A lot of other data is also recorded, such as a map of all the pages that each page links to, the anchor text of those links, whether or not those links are considered ads, and more. To accomplish the monumental task of holding data on hundreds of billions (or trillions) of pages that can be accessed in a fraction of a second, the search engines have constructed massive data centers.

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 seed set of websites. Starting with a known trusted set of websites enables search engines to measure how much they trust the other websites that they find through the crawling process.

- **Retrieval and Rankings**

Start of a user's search quest The next step in this quest occurs when the search engine returns a list of relevant pages on the Web in the order it believes is most likely to satisfy the user. This process requires the search engine to scour its corpus of hundreds 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. It is the perception of 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 intention and terms of the user's query. The relevance of a document increases if the page contains terms relevant to the phrase queried by the user, 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 your site is not relevant to a query, the search engine does not consider it for inclusion in the search results for that query.

Importance refers to the relative importance, measured via *citation*, of a given document that matches the user's query. The importance of a given document increases with every other document that references it. In today's online environment, citations can come in the form of links to the document or references to it on social media sites. Determining how to weight these signals 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.

- **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 a 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 that it can use to determine whether or not to show that page in the results when a user enters a related search query. This map, often referred to as a *semantic map*, is built to help the search engine 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 in the results page. Therefore, the words you put on your page, and the "theme" of that page, play a huge role in ranking.

NATIONAL COMPUTER COLLEGE

Breaking up a web page The navigational elements of a web page are likely to be 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 focuses on the “real content” of the page.

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 used across many pages on a site, it does not help the search engine with the task of isolating how the content of a given web page differs from the content of other pages on the same site.

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

- 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.
- The information the search engine crawler is most interested in is the HTML text on the page.
- Although page source 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 engine read few other elements.
 - Page title
 - Meta keywords tag
 - Meta description tag
 - alt attribute for images
 - NoScript tag

What search engines cannot see?

- It is also worthwhile to review the types of content that search engines cannot “see” in the human sense.
 - Image
 - Flash
 - Audio and video
 - Ajax, JavaScript

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.

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 thesauruses and dictionaries 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

NATIONAL COMPUTER COLLEGE

Fuzzy Set Theory and certain equations 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 incorporates hundreds of technical and often difficult-to-comprehend terms, these can be broken down and understood even by an SEO novice. The following are some common types of searches in the IR field:

Measuring Content Quality and User Engagement

The search engines also attempt to measure the quality and uniqueness of a website's content. One method they may use for doing this is evaluating the document itself. For example, if a web page has lots of spelling and grammatical errors, that can be taken as a sign that little editorial effort was put into that page.

The search engines can also analyze the *reading level* of the document. One popular formula for doing this is the Flesch-Kincaid Grade Level Readability Formula, which considers things like the average word length and the number of words in a sentence to determine the level of education needed to be able to understand the sentence. Imagine a scenario where the products being sold on a page are children's toys, but the reading level calculated suggests that a grade level of a senior in college is required to read the page. This could be another indicator of poor editorial effort.

The other method that search engines can use to evaluate the quality of a web page is to measure actual user interaction. For example, if a large number of the users who visit the web page after clicking on a search result immediately return to the search engine and click on the next result, that would be a strong indicator of poor quality.

Google has access to a large number of data sources that it can use to measure how visitors interact with your website. Some of those sources include:

Interaction with web search results

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. Alternatively, 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, and how much weight they might put on it, is not known.

Google Analytics

It is hard to get a firm handle on just what percentage of websites run Google Analytics. A 2008 survey of websites by immeria.net showed that Google Analytics had a market share of 59%, and the Metric Mail Blog checked the top 1 million sites in Alexa and found that about 50% of those had Google Analytics. Suffice it to say that Google is able to collect detailed data about what is taking place on a large percentage of the world's websites. Google Analytics data provides Google with a rich array of data on those sites, including:

Bounce rate

The percentage of visitors who visit only one page on your website.

Time on site

The average amount of time spent by users on the site. Note that Google Analytics only receives information when each page is loaded, so if a visitor views only one page, it does not know how much time is spent on that page. More precisely, then, this metric tells you the average time between the loading of the first page and the loading of the last page, but it does not take into account how long visitors spent on the last page loaded.

Page views per visitor

The average number of pages viewed per visitor on your site.

Google Toolbar

It is not known how many users out there use the Google Toolbar, but the authors believe that they number in the millions. Google can track the entire web surfing behavior of these users. Unlike Google Analytics, the Google Toolbar can measure the time from when a user first arrives on a site to the time when that user loads a page from a different website. It can also get measurements of bounce rate and page views per visitor.

Google +1 Button

NATIONAL COMPUTER COLLEGE

In April 2011, Google began public testing of a new feature, the +1 button. This enables users to “vote” for a page, either directly in the search results or on the page itself, thereby identifying their favorite websites for a particular search query.

Chrome Blocklist Extension

In February 2011, Google released the Chrome Blocklist Extension. This provides users of the Chrome browser a way to indicate search results they don't like.

Google Instant Previews

Google also offers Instant Previews in its search results. This allows users to see a thumbnail view of the web page behind a search result before deciding to click on it. If a user looks at the preview for your page and then decides not to click on it, this can act as a negative vote for your site.

Google Reader

Google also provides the world's most popular RSS feed reader, which provides it with a lot of data on which content is the most engaging.

Goo.gl

In September 2010, Google released its own URL shortener. This tool allows Google to see what content is being shared and what content is being clicked on, even in closed environments where Google web crawlers are not allowed to go.

• 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 the 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 all 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 (one link can be worth 10 million times more than another). A website or page can be determined to be authoritative by combining an analysis of the linking patterns with semantic analysis. For example, perhaps you are interested in sites. Search engines can use semantic analysis to identify the collection of web pages that focus on the topic. The search engines can then determine which of these sites have the most links from the set 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 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 not authoritative because the right sites do not link to it. Such a grouping of relevant sites 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 links from non relevant 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 non relevant 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 is also a strong signal to the search engines. This is referred to as the *anchor text*, and if that text is keyword-rich, it will do more for your rankings in the search engines than if the link is not keyword-rich.

Analyzing Ranking Factors

SEOMoz periodically conducts surveys of leading search engine optimizers to determine what they think are the most important ranking factors. Here is a high-level summary of the top nine results:

Page Level Link Metrics

This refers to the links as related to the specific page, such as the number of links, the relevance of the links, and the trust and authority of the links received by the page.

Domain Level Link Authority Features

NATIONAL COMPUTER COLLEGE

Domain level link authority is based on a cumulative link analysis of all the links to the domain. Factors considered include the number of different domains linking to the site, the trust/authority of those domains, the rate at which new inbound links are added, the relevance of the linking domains, and more.

Page Level Keyword Usage

This describes use of the keyword term/phrase in particular parts of the HTML code on the page (title element, <h1>s, alt attributes, etc.).

Domain Level Keyword Usage

This refers to how keywords are used in the root or sub domain name, and how impactful that might be on search engine rankings.

Page Level Social Metrics

Social metrics considered include mentions, links, shares, Likes, and other social media site-based metrics. At the time of the survey, the considered sites were Facebook and Twitter. Since then Google has launched Google+, and Search, plus Your World, which would also be included in this definition.

Domain Level Brand Metrics

This factor includes search volume on the website's brand name, mentions, whether it has a presence in social media, and other brand-related metrics.

Page Level Keyword Agnostic Features

Factors included here are on-page elements such as the number of links on the page, number of internal links, number of followed links, number of NoFollowed links, and other similar factors.

Page Level Traffic/Query Data

Elements of this factor include the click-through rate (CTR) to the page in the search results, the bounce rate of visitors to the page, and other similar measurements.

Domain Level Keyword Agnostic Features

Major elements of this factor in the survey included the number of hyphens in the domain name, numeric characters in the domain name, and domain name length.

Negative Ranking Factors

The SEOMoz survey also identified a number of negative ranking factors. Some of the most significant ones included:

Malware being hosted on the site

The search engines will act rapidly to penalize sites that contain viruses or trojans.

Cloaking

Search engines want publishers to show the same content to the search engine as is shown to users.

Pages on the Site That Sell Links

Google has a strong policy against paid links, and sites that sell them may be penalized. *Content that advertises Paid Links On The Site*

As an extension of the prior negative ranking factor, promoting the sale of paid links may be 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. However, it gets more complicated than that. If your site suddenly starts to get 300 new links per day, you have either abruptly become a lot more relevant or started to acquire links in a spammy way.

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 a user's approximate location and then show results tailored to that 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*. This is sometimes referred to as *adaptive search*. To avoid personalization, before searching users need to log out of their Google accounts and select "disable customizations based on search history" in the Google interface under Web History. This will allow the

NATIONAL COMPUTER COLLEGE

user to see Google results that are not personalized based on search history. However, the results will still be personalized to the user's location. You can also depersonalize your search results by performing your search query, and then appending `&pws=0` to the end of the search page URL and reloading the page. You also need to have Google Instant turned off in your preferences.

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 has become 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.

Vertical Search Engines

Vertical search is the term people sometimes use for specialty or niche search engines that focus on a limited data set. 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.

- **Image search**

All three of the big 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.

- **Video search**

As with image search, video search engines focus on searching specific types of files on the Web—in this case, video files in formats such as MPEG, AVI, and others.

- **News search**

News search is also unique. News search results operate on a different time schedule, as they must 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 nearly instantaneous responses. Modern consumers tend to want their news information.

- **Local search/maps**

Next up in our hit parade of major search verticals is local search (a.k.a. map search). Local search results are now heavily integrated into the traditional web search results, so a presence in local search can have a large impact on organizations that have one or more brick and mortar locations.

- **Blog search**

Google has implemented a search engine focused just on blog search called Google Blog Search. This search engine will respond to queries, but only search blogs (more accurately, feeds) to determine the results.

- **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.

- **Shopping search**

Microsoft also has some unique vertical search properties. One of the more interesting ones is its vertical shopping search solution.

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 engines that (are) the most important in that market. According to comScore data from June 2011, Google receives 68.9% of all searches performed worldwide. In addition, Google is the market share leader in every major regional market. In the Asia Pacific region.

China

NATIONAL COMPUTER COLLEGE

Baidu News reported in April 2011 that Baidu had more than 75% market share in China in.

Russia

According to figures announced by Yandex, the company's market share in Russia comprised about 65% of all searches.

South Korea

Naver was estimated to have about 70% market share in South Korea.

Czech Republic

The StartupMeme Technology blog reported Seznam as having more than 45% market share in the Czech Republic.

Unit 2

Developing an SEO Plan Prior to Site Development

It is widely understood in the industry that search engine optimization should be built in, as early as possible, to the entire site development strategy, from choosing a content management system (CMS) and planning site architecture to developing on-page content. SEO practitioners have significant input in both of these areas. Of course, many businesses learn about the need for SEO only after they have built their sites, in which case the time to start is now. SEO plans have many moving parts, and SEO decisions can have a significant impact on other departments, such as development, other marketing groups, and sales. Getting that input as soon as possible will bring the best results for a business at the least possible cost.

Business Factors That Affect the SEO Plan

Here are some examples of business issues that can impact SEO:

Revenue/business model

It makes a difference to the SEO practitioner if the purpose of the site is to sell products, sell advertising, or obtain leads. We will discuss this more in the later sections of this chapter.

Target customers

Who are you trying to reach? This could be an age group, a gender group, or as specific as people looking to buy a house within a 25-mile radius of Orlando, FL.

Competitor strategies

The competitive landscape is another big factor in your SEO plan. Competition may be strongly entrenched in one portion of the market online, and it may make sense to focus on a different segment. Or you may be the big dog in your market but you have specific competitors you want to fend off.

Branding goals

There may be terms that it is critical for you to own, for branding reasons. *Budget for content development* An important part of link building is ensuring the quality of your content, as well as your capacity to commit to the ongoing development of high-quality on-page site content.

How your potential customers search for products like yours

Understanding what customers do when they are searching for products or services like yours is one of the most basic functions of SEO. This involves mapping the actual search queries your target customers use when they go to a search engine to solve their current problem.

Understanding Your Audience and Finding Your Niche

A nontrivial part of an SEO plan is figuring out who you are targeting with your website. This is not always that easy to determine. As you will see in this section, many factors enter into this, including the competition, the particular strengths or weaknesses of your own company, and more.

Mapping Your Products and Services

Successful SEO requires a thorough understanding of the business itself. What products, services, and types of information and resources does your organization have to offer? As we outlined in the preceding section, a critical SEO activity understands who is searching for what you are trying to promote, which requires thoroughly understanding all aspects of your offering. You will also need to understand the broad market categories that your products fall into, as each of these categories might relate to sections of your website that you may want to create. By having sections of the site for those categories, you create an opportunity to obtain search traffic related to those categories. You also should consider business development and the company's expansion strategy at the outset of the SEO planning process. Consider Amazon, which began as a bookseller but has evolved into a general purpose e-tailer. Sites that go through these types of changes may need to be substantially restructured, and such restructurings can be a source of major SEO headaches. Anticipating those changes in advance provides the opportunity to recommend architectural approaches to dealing with those changes.

• **Content Is King**

One aspect of determining the desired audience for your website is determining who you want to reach, which requires an understanding of what you have to offer visitors to your site, both now and in the future. You may have a deep library of "how to" content, great videos, a unique photo gallery, or an awesome tool that people are interested in using. Each of these can be valuable in building a world-class website

NATIONAL COMPUTER COLLEGE

that does well in the search engines. The content you have available to you will affect your keyword research and site architecture, as your site content is the major source of information that search engines use to determine what your site is about. You need relevant content to even be “in the game” in search. On-site content also affects your link-building efforts. Link building is very similar to PR in that the success of your link-building efforts is integrally related to what you are promoting

• Segmenting Your Site’s Audience

Let’s not forget the audience itself! It is important for the SEO practitioner to understand the target audience. For example, Site A may be a website that sells gadgets. As a result, the site’s developers go out and implement a brilliant campaign to rank for the terms they consider relevant. Being young and energetic, they focus on the way their peers search for gadgets—but what if the target audience for the gadgets Site A sells are age 50 or older? Uh-oh, Site A is in trouble again. Why? The target audience for Site A (the over-50 crowd) may use different search terms than the younger generation to search for gadgets, which means

SEO for Raw Traffic

Optimizing for search engines and creating keyword-targeted content helps a site rank for key search terms, which typically leads to direct traffic and referring links as more and more people find, use, and enjoy what you’ve produced. Thousands of sites on the Web leverage this traffic to serve advertising, directly monetizing the traffic sent from the engines. Here are some factors to think about when considering SEO for raw traffic:

When to employ SEO for raw traffic

Use it when you can monetize traffic without actions or financial transactions taking place on your site.

Keyword targeting

Keyword targeting in this scenario can be very broad. The goal here isn’t typically to select specific keywords, but rather to create lots of high-quality content that naturally targets interesting/searched-for terms. Instead of singular optimization on specific terms, the focus is on accessibility and best practices throughout the site to earn traffic through both high-volume and long-tail queries. Concentrate efforts on great content, and use keyword-based optimization only as a secondary method to confirm the titles/headlines of the works you create.

Page and content creation/optimization

A shallow, highly crawlable link structure is critical to getting all of your content indexed—follow good information architecture practices and use intelligent, detailed category and subcategory structures to get the most benefit out of your work. You’ll also need to employ good on-page optimization and make your articles easy to share and optimized for viral spreading.

SEO for Ecommerce Sales

One of the most direct monetization strategies for SEO is driving relevant traffic to an ecommerce shop to boost sales. Search traffic is among the best quality available on the Web, primarily because a search user has expressed a specific goal through her query, and when this matches a product or brand the web store carries, conversion rates are often extremely high. Here are some factors to think about when considering SEO for ecommerce sales:

When to employ SEO for ecommerce sales

Use it when you have products/services that are directly for sale on your website.

Keyword targeting

Paid search advertising is an excellent way to test the efficacy and potential ROI of keyword targets. Find those that have reasonable traffic and convert well, and pursue them further. You’ll often find that the more specific the query is—brand-inclusive, product-inclusive, and so on—the more likely the visitors are to make the purchase. Of course, as noted earlier, you should have little difficulty ranking for your brand terms, so the best use of this tactic is for generic terms that you will find harder to win on so you can decide if they are worth the effort.

Page and content creation/optimization

You’ll typically need to do some serious link building, along with internal optimization, to achieve high rankings for competitive, high-value keywords that bring in conversion-focused traffic. Manual link building is an option here, but scalable strategies that leverage a community or customers can be equally, or even more, valuable.

SEO for Mindshare/Branding

A less popular but equally powerful application of SEO is its use for branding purposes. Bloggers, social media/community websites, content producers, news outlets, and dozens of other web publishing

NATIONAL COMPUTER COLLEGE

archetypes have found tremendous value in appearing atop the SERPs and using the resulting exposure to bolster their brand recognition and authority. Here are some factors to think about when considering SEO for mindshare/branding:

When to employ SEO for mindshare/branding

Using it when branding or communicating a message is your goal. If you do not have direct monetization goals for the moment or for the foreseeable future, this is the approach for you. This approach can also be used on portions of ecommerce sites that are not about conversion but more about long-term branding and mindshare.

Keyword targeting

A keyword focus is less critical here—you'll likely have a few broad terms that receive the high traffic you want, but the long tail may be far more achievable and the better target. Focus on keywords that are going to bring you visitors who are likely to be interested in and remember your brand.

Page and content creation/optimization

Make an accessible site, use good link structure, apply best practices, and focus on links for domain authority rather than chasing after specific keywords.

SEO for Lead Generation and Direct Marketing

Although lead generation via the Web is less direct than an ecommerce transaction, it is arguably just as valuable and important for building customers, revenue, and long-term value. Millions of search queries have commercial intents that can't be (or currently aren't) fulfilled directly online. These can include searches for services such as legal consulting, contract construction, commercial loan requests, alternative energy providers, or virtually any service or product people source via the Web. Here are some factors to think about when considering SEO for lead generation and direct marketing:

When to employ SEO for lead generation and direct marketing

Use it when you have a non-ecommerce product/service/goal that you want users to accomplish on your site or for which you are hoping to attract inquiries/direct contact over the Web.

Keyword targeting

As with ecommerce, choose phrases that convert well, have reasonable traffic, and have previously performed in PPC campaigns.

Page and content creation/optimization

Although you might think it would be easier to rank high in the SERPs for lead-generation programs than for ecommerce, it is often equally challenging. You'll need a solid combination of on-site optimization and external link building to many different pages on the site (with good anchor text) to be competitive in the more challenging arenas.

SEO for Reputation Management

Since one's own name—whether personal or corporate—is one's identity, establishing and maintaining the reputation associated with that identity is generally of great interest. Imagine that you search for your brand name in a search engine, and high up in the search results is a web page that is highly critical of your organization.

Here are some factors to think about when considering SEO for reputation management:

When to employ SEO for reputation management

If you're trying to either protect your brand from negative results appearing on page 1 or push down already existing negative content, reputation management SEO is the only path to success.

Keyword targeting

Chances are this is very easy—the keyword you are targeting is your personal name, your brand name, or some common variant. You might want to use keyword research tools just to see whether there are popular variants you're missing.

Page and content creation/optimization

Unlike the other SEO tactics, reputation management involves optimizing pages on many different domains to demote negative listings. This involves using social media profiles, public relations, press releases, and links from networks of sites you might own or control, along with classic optimization of internal links and on-page elements. It is certainly among the most challenging of SEO practices, especially in Google, where the use of the *query deserves diversity* (QDD) algorithm can mean you have to work much harder to push down negatives because of how it favors diverse content.

SEO for Ideological Influence

For those seeking to sway public (or private) opinion about a particular topic, SEO can be a powerful tool. By promoting ideas and content within the search results for queries likely to be made by those seeking information about a topic, you can influence the perception of even very large groups. Politicians and

NATIONAL COMPUTER COLLEGE

political groups and individuals are the most likely employers of this tactic, but it can certainly be applied to any subject, from the theological to the technical or civic. Here are some factors to think about when considering SEO for ideological influence:

When to employ SEO for ideological influence

Use it when you need to change minds or influence decisions/thinking around a subject—for example, a group of theoretical physicists attempting to get more of their peers to consider the possibility of alternative universes as a dark matter source.

Keyword targeting

It's tough to say for certain, but if you're engaging in these types of campaigns, you probably know the primary keywords you're chasing and can use keyword research query expansion to find others.

Page and content creation/optimization

This is classic SEO, but with a twist. Since you're engaging in ideological warfare in the SERPs, chances are you've got allies you can rally to the cause. Leverage your combined links and content to espouse your philosophical preferences.

The Major Elements of Planning

As any experienced SEO consultant will tell you, you should incorporate your SEO strategy into the site planning process long before your site goes live. Your strategy should be well outlined before you make even the most basic technology choices, such as the hosting platform and your CMS. However, this is not always possible—and in fact, more often than not an SEO professional will be brought in to work on a site that already exists. Regardless of when you start, there are a number of major components to any SEO plan that you need to address long before you research the first title tag.

• **Technology Choices**

As we already suggested, SEO is a technical process, and as such, it impacts major technology choices. For example, a CMS can facilitate—or, possibly, undermine—your SEO strategy. Some platforms do not allow you to have titles and meta descriptions that vary from one web page to the next, create hundreds (or thousands) of pages of duplicate content, or make a 302 (temporary) redirect the default redirect. All of these things could be disastrous for your website. This problem also exists with web servers. For example, if you use Internet Information Services (IIS), the default redirect choice is a 302. You can configure IIS to use a 301 redirect, but this is something you need to understand how to do and build into your SEO plan up front.

• **Market Segmentation**

Another critical factor to understand is the nature of the market in which you are competing. This tells you how competitive the environment is in general, and augmented with additional research, you can use this information to tell how competitive the SEO environment is. In some markets, natural search is intensively competitive.

Identifying the Site Development Process and Players

Before you start the SEO process, it is imperative to identify who your target audience is, what your message is, and how your message is relevant. There are no web design tools or programming languages that tell you these things. Your company's marketing, advertising, and PR teams have to set the objectives before you can implement them—successful SEO requires a team effort.

Your SEO team should be cross-functional and multidisciplinary, consisting of the team manager, the technical team, the creative team, the data and analytics team (if you have one), and the major stakeholders from marketing, advertising, and PR. In a smaller organization, you may have to wear all of those hats yourself. The team leader wants to know who the target audience is. What does the marketing team know about them? How did we find them? What metrics will we use to track them? All of this is key information that should have an impact on the project's technical implementation. Advertising messages need to be well thought out and prepared.

You do not want your team bickering over whether to optimize for “hardcore widget analysis” or “take your widgets to the next level.” Advertising serves multiple purposes, but its most fundamental purpose is to compel people to take a specific action. What action are you hoping to compel people to take? The PR team has to take your story to the media and entice them into writing and talking about it. What message do they want to deliver? You have to mirror that message in your content.

If they say you're relevant to organic cotton clothes but your project plan says you're relevant to yoga attire, the whole project is in trouble. When you're creating visibility, the people who build up your brand have to see a clear, concise focus in what you do. If you provide them with anything less, they'll find someone else to talk about. The technical and creative team is responsible for delivering the project. They take direction from marketing, advertising, and PR on what needs to be accomplished, but from there on out they have to put the pieces into place. As the project unfolds, marketing has to come back and say whether the target audience is being reached, advertising has to come back and say whether the

NATIONAL COMPUTER COLLEGE

message is clear, and PR has to come back and say whether the media like what they see. Ongoing feedback is essential because the success of your project is determined solely by whether you're meeting your goals. A successful SEO team understands all of these interactions and is comfortable relying on each team member to do his part. Establishing good communication among team members is essential. And even if you are a team of one, you need to understand all of these steps. Addressing all aspects of the marketing problem is a requirement for success.

Defining Your Site's Information Architecture

Whether you're working with an established website or not, you should plan to research the desired site architecture at the start of your SEO project. This task can be divided into two major components: technology decisions and structural decisions.

- **Technology Decisions**

As we outlined previously in this chapter, your technology choices can have a major impact on your SEO results. The following is an outline of the most important issues to address at the outset:

Dynamic URLs

Although Google now states that dynamic URLs are not a problem for the company, this is not entirely true, nor is it the case for the other search engines. Make sure your CMS does not end up rendering your pages on URLs with many convoluted parameters in them.

Session IDs or user IDs in the URL

It used to be very common for CMSs to track individual users surfing a site by adding a tracking code to the end of the URL. Although this worked well for this purpose, it was not good for search engines, because they saw each URL as a different page rather than variants of the same page. Make sure your CMS does not ever serve up session IDs. If you are not able to do this, make sure you use rel="canonical" on your URLs.

Superfluous flags in the URL

Related to the preceding two items is the notion of extra junk being present in the URL. This probably does not bother Google, but it may bother the other search engines, and it interferes with the user experience for your site.

Links or content based in JavaScript, Java, or Flash

Search engines often cannot see links and content implemented using these technologies. Make sure the plan is to expose your links and content in simple HTML text.

Content behind forms (including pull-down lists)

Making content accessible only after the user has completed a form (such as a login) or made a selection from an improperly implemented pull-down list is a great way to hide content from the search engines. Do not use these techniques unless you want to hide your content!

Temporary (302) redirects

This is also a common problem in web server platforms and CMSs. The 302 redirect blocks a search engine from recognizing that you have permanently moved the content and it can be very problematic for SEO as 302 redirects block the passing of Page Rank. You need to make sure the default redirect your systems use is a 301, or understand how to configure it so that it becomes the default. All of these are examples of basic technology choices that can adversely affect your chances for a successful SEO project. Do not be fooled into thinking that SEO issues are understood, let alone addressed, by all CMS vendors out there—unbelievably, many are still very far behind the SEO curve. It is also important to consider whether a “custom” CMS is truly needed when many CMS vendors are creating ever more SEO-friendly systems—often with much more flexibility for customization and a broader development base. There are also advantages to selecting a widely used CMS, including portability in the event that you choose to hire different developers at some point. Also, do not assume that all web developers understand the SEO implications of what they develop. Learning about SEO is not a requirement to get a software engineering degree or become a web developer. It is up to you, the SEO expert, to educate the other team members on this issue as early as possible in the development process.

- **Structural Decisions**

One of the most basic decisions to make about a website concerns internal linking and navigational structures, which are generally mapped out in a site architecture document. What pages are linked to from the home page? What pages are used as top-level categories that then lead site visitors to other related pages? Do pages that are relevant to each other link to each other? There are many, many aspects to determining a linking structure for a site, and it is a major usability issue because visitors make use of the links to surf around your website. For search engines, the navigation structure helps their crawlers determine what pages you consider the most important on your site, and it helps them establish the relevance of the pages on your site to specific topics.

- **Target keywords**

Keyword research is a critical component of SEO. What search terms do people use when searching for products or services similar to yours? How do those terms match up with your site hierarchy? Ultimately, the logical structure of your pages should match up with the way users think about products and services like yours.

- **Cross-link relevant content**

Linking between articles that cover related material can be very powerful. It helps the search engine ascertain with greater confidence how relevant a web page is to a particular topic. This can be extremely difficult to do well if you have a massive ecommerce site, but Amazon solves the problem very well, as shown in. The “Frequently Bought Together” and “What Do Customers Ultimately Buy After Viewing This Item?” sections are brilliant ways to group products into categories that establish the relevance of the page to certain topic areas, as well as to create links between relevant pages. In the Amazon system, all of this is rendered on the page dynamically, so it requires little day-to-day effort on Amazon’s part. The “Customers Who Bought...” data is part of Amazon’s internal databases, and the “Tags Customers Associate...” data is provided directly by the users themselves. Of course, your site may be quite different, but the lesson is the same. You want to plan on having a site architecture that will allow you to cross-link related items.

- **Use anchor text**

Anchor text is one of the golden opportunities of internal linking. As an SEO practitioner, you need to have in your plan from the very beginning a way to use keyword-rich anchor text in your internal links. Avoid using text such as “More” or “Click here,” and make sure the technical and creative teams understand this. You also need to invest time in preparing an anchor text strategy for the site.

- **Use breadcrumb navigation**

Breadcrumb navigation is a way to show the user where he is in the navigation hierarchy. This is helpful to both users and search engines.

- **Minimize link depth**

Search engines (and users) look to the site architecture for clues as to what pages are most important. A key factor is how many clicks from the home page it takes to reach a page. A page that is only one click from the home page is clearly important. A page that is five clicks away is not nearly as influential. In fact, the search engine spider may never even find such a page (depending in part on the site’s link authority). Standard SEO advice is to keep the site architecture as flat as possible, to minimize clicks from the home page to important content. Do not go off the deep end, though; too many links on a page are not good for search engines either.

Auditing an Existing Site to Identify SEO Problems

Auditing an existing site is one of the most important tasks that SEO professionals encounter. SEO is still a relatively new field and many of the limitations of search engine crawlers are non intuitive. In addition, many web developers, unfortunately, are not well versed in SEO. Even more unfortunately, some stubbornly refuse to learn, or, worse still, have learned the wrong things about SEO. This includes those who have developed CMS platforms, so there is a lot of opportunity to find problems when conducting a site audit.

Elements of an Audit

Your website needs to be a strong foundation for the rest of your SEO efforts to succeed. An SEO site audit is often the first step in executing an SEO strategy. The following sections identify what you should look for when performing a site audit.

- **Usability**

Although this may not be seen as a direct SEO issue, it is a very good place to start. Usability affects many factors, including conversion rate as well as the propensity of people to link to a site.

- **Accessibility/spiderability**

Make sure the site is friendly to search engine spiders. We discuss this in detail in.

- **Search engine health check**

Here are some quick health checks:

- Perform a site:*yourdomain.com* search in the search engines to check how many of your pages appear to be in the index. Compare this to the number of unique pages you believe you have on your site.
- Test a search on your brand terms to make sure you are ranking for them.
- Check the Google cache to make sure the cached versions of your pages look the same as the live versions of your pages.
- Check to ensure major search engine “tools” have been verified for the domain.

- **Keyword health checks**

Are the right keywords being targeted? Does the site architecture logically flow from the way users search on related keywords? Does more than one page target the same exact keyword.

- **Duplicate content checks**

The first thing you should do is to make sure the non-*www* versions of your pages 301-redirect to the *www* versions of your pages, or vice versa. While you are at it, check that you don't have *https*: pages that are duplicates of your *http*: pages. You should check the rest of the content on the site as well. The easiest way to do this is to take unique strings from each of the major content pages on the site and search on them in Google. Make sure you enclose the string inside double quotes.

- **URL check**

Make sure you have clean, short, descriptive URLs. *Descriptive* means keyword-rich but not keyword-stuffed. You don't want parameters appended (have a minimal number if you must have any), and you want them to be simple and easy for users (and spiders) to understand.

- **Title tag review**

Make sure the title tag on each page of the site is unique and descriptive. If you want to include your company brand name in the title, consider putting it at the end of the title tag, not at the beginning, as placement of keywords at the front of a URL brings ranking benefits. Also check to make sure the title tag is fewer than 70 characters long.

- **Content review**

Do the main pages of the site have enough content? Do these pages all make use of header tags? A subtler variation of this is making sure the number of pages on the site with little content is not too high compared to the total number of pages on the site.

- **Meta tag review**

Check for a meta robots tag on the pages of the site. If you find one, you may have already spotted trouble. An unintentional NoIndex or NoFollow tag (we define these in “Content Delivery and Search Spider Control” on page 245) could really mess up your search ranking plans. Also make sure every page has a unique meta description. If for some reason that is not possible, consider removing the meta description altogether. Although the meta description tags are generally not a significant factor in ranking, they may well be used in duplicate content calculations, and the search engines frequently use them as the description for your web page in the SERPs; therefore, they affect click-through rate.

- **Sitemaps file and robots.txt file verification**

Use the Google Webmaster Tools “Test robots.txt” verification tool to check your *robots.txt* file. Also verify that your *Sitemaps* file is identifying all of your (canonical) pages.

- **Redirect checks**

Use a server header checker such as Live HTTP Headers to check that all the redirects used on the site return a 301 HTTP status code. Check all redirects this way to make sure the right thing is happening. This includes checking that the canonical redirect is properly implemented. Unfortunately, given the non intuitive nature of why the 301 redirect is preferred, you should verify that this has been done properly even if you have provided explicit direction to the web developer in advance. Mistakes do get made, and sometimes the CMS or the hosting company makes it difficult to use a 301.

NATIONAL COMPUTER COLLEGE

- **Internal linking checks**

Look for pages that have excessive links. Google advises 100 per page as a maximum, although it is OK to increase that on more important and heavily linked-to pages. Make sure the site makes good use of anchor text in its internal links. This is a free opportunity to inform users and search engines what the various pages of your site are about. Don't abuse it, though. For example, if you have a link to your home page in your global navigation (which you should), call it "Home" instead of picking your juiciest keyword. The search engines view that particular practice as spammy, and it does not engender a good user experience. Furthermore, the anchor text of internal links to the home page is not helpful for rankings anyway. Keep using that usability filter through all of these checks.

- **Avoidance of unnecessary sub domains**

The engines may not apply the entirety of a domain's trust and link juice weight to sub domains. This is largely due to the fact that a sub domain could be under the control of a different party, and therefore in the search engine's eyes it needs to be separately evaluated. In the great majority of cases, sub domain content can easily go in a subfolder.

- **Geolocation**

If the domain is targeting a specific country, make sure the guidelines for country geotargeting outlined in. If your concern is primarily about ranking for *chicago pizza* because you own a pizza parlor in Chicago, IL, make sure your address is on every page of your site. You should also check your results in Google Local to see whether you have a problem there. Additionally, you will want to register with Google Places.

- **External linking**

Check the inbound links to the site. Use a backlinking tool such as Open Site Explorer to collect data about your links. Look for bad patterns in the anchor text, such as 87% of the links having the critical keyword for the site in them. Unless the critical keyword happens to also be the name of the company, this is a sure sign of trouble. This type of distribution is quite likely the result of link purchasing or other manipulative behavior.

- **Page load time**

Is the page load time excessive? Too long a load time may slow down crawling and indexing of the site. However, to be a factor, this really does need to be excessive—certainly longer than five seconds, and perhaps even longer than that.

- **Image alt tags**

Do all the images have relevant, keyword-rich image alt attribute text and filenames? Search engines can't easily tell what is inside an image, and the best way to provide them with some clues is with the alt attribute and the filename of the image. These can also reinforce the overall context of the page itself.

- **Code quality**

Although W3C validation is not something the search engines require, checking the code itself is a good idea. Poor coding can have some undesirable impacts. You can use a tool such as SEO Browser to see how the search engines see the page.

Identifying Current Server Statistics Software and Gaining Access

- **Web Analytics**

Analytics software can provide you with a rich array of valuable data about what is taking place on your site. It can answer questions such as:

- How many unique visitors did you receive yesterday?
- Is traffic trending up or down?
- What are the most popular search terms with which people find you?
- What are the most popular pages on your site?
- What are the best-converting pages on the site?

We strongly recommend that if your site does not currently have any measurement systems in place

NATIONAL COMPUTER COLLEGE

- **Logfile Tracking**

Logfiles contain a detailed click-by-click history of all requests to your web server. Make sure you have access to the logfiles and some method for analyzing them. If you use a third-party hosting company for your site, chances are it provides some sort of free logfile analyzer, such as AWStats, Webalizer, or something similar. Obtain access to whatever tool is in use as soon as you can. What these tools do that JavaScript-based web analytics software cannot is record search engine spider activity on your site. Although spidering will typically vary greatly from day to day, you can still see longer-term trends of search engine crawling patterns, and whether crawling activity is trending up (good) or down (bad). Although this web crawler data is very valuable, do not rely on these free solutions provided by hosting companies for all of your analytics data, as there is a lot of value in what traditional analytics tools can capture.

- **Google and Bing Webmaster Tools**

As mentioned earlier, other valuable sources of data include Google Webmaster Tools and Bing Webmaster Tools. From a planning perspective, you will want to get these tools in place as soon as possible. Both tools provide valuable insight into how the search engines see your site. This includes things such as external link data, internal link data, crawl errors, high-volume search terms, and much, much more.

- **Search Analytics**

Search analytics is a new and emerging category of tools. Search analytics tools specifically monitor how your website interacts with the search engines. Compete offers search-specific analytic tools, as do many smaller vendors. Although this category is in its infancy, it is worth monitoring closely to see what tools become available that can provide your organization with an advantage in competing for search traffic.

Determining Top Competitors

Understanding the competition should be a key component of planning your SEO strategy. The first step is to understand who your competitors in the search results really are. It can often be small players who give you a run for your money. For example, consider the previously mentioned credit card search in Google. Visa, MasterCard, American Express, and Discover Card all fail to reach the #1 position in the Google results. Instead, affiliate players dominate these results. Affiliates tend to be the most adept at search engine optimization and can be the most lax in abiding by the search engines' terms and conditions.

Assessing Historical Progress

Measuring the results of SEO changes can be challenging, partly because there are so many moving parts and partly because months can elapse between when changes are made to a site and when results are seen in search rankings and traffic. This difficulty only increases the importance of measuring progress and being accountable for results. This section will explore methods for measuring the results from your SEO efforts.

- **Maintain a Timeline of Site Changes**

Keeping a log of changes to your site is absolutely recommended. If you're not keeping a timeline, you will have a harder time executing your SEO plan and managing the overall SEO process. Sure, without one you can still gauge the immediate effects of content additions/revisions, link acquisitions, and development changes, but visibility into how technical modifications to the website might have altered the course of search traffic, whether positively or negatively, is obscured. If you can't map changes—both those intended to influence SEO and those for which SEO wasn't even a consideration—you'll be optimizing blind and could miss powerful signals that.

Google Trends for Websites could help dictate your strategy going forward. You should track more than just site changes as well. External factors that can have a big impact on your SEO results include confirmed search engine algorithm updates, competitor news events, and breaking news. Factors within your own business can have an impact too, such as major marketing or PR events, IPOs, or the release of earnings statements. There are many scenarios in which you will want to try to establish cause and effect, such as:

If search traffic spikes or plummets

Sudden changes in organic traffic are obviously notable events. If traffic plummets, you will be facing lots of questions about why, and having a log of site changes will put you in a better position to assess

NATIONAL COMPUTER COLLEGE

whether any changes you recommended could have been the cause. Of course, if traffic spikes you will want to be able to see whether an SEO-related change was responsible as well.

When gradual traffic changes begin

Changes do not always come as sudden spikes or drop-offs. If you see the traffic beginning a gradual climb, you will want to be able to assess the likely reasons.

To track and report SEO progress

Accountability is a key component of SEO. Budget managers will want to know what return they are getting on their SEO investment. This will inevitably fall into two buckets: itemizing specific work items worked on, and analyzing benefits to the business. Keeping an ongoing change log makes tracking and reporting SEO progress much easier to accomplish.

• **Types of Site Changes That Can Affect SEO**

Your log should track all changes to the website, not just those that were made with SEO in mind. Organizations make many changes that they do not think will affect SEO, but that have a big impact on it. Here are some examples:

- Adding content areas/features/options to the site.
- Changing the domain of the site.
- Modifying URL structures.
- Implementing a new CMS.
- Establishing new partnerships that either send links or require them.
- Acquiring new links to pages on the site other than the home page.
- Making changes to navigation/menu systems.
- Implementing redirects either to or from the site.
- Marketing activities that may drive upticks in usage/traffic and the source.

When you track these items, you can create an accurate storyline to help correlate causes with effects. If, for example, you've observed a spike in traffic from Yahoo! that started four to five days after you switched your menu links from the page footer to the header, it is likely that there is a causal relationship. Without such documentation it could be months before you noticed the surge—and there would be no way to trace it back to the responsible modification. Your design team might later choose to switch back to footer links, your traffic might fall, and no record would exist to help you understand why. Without the lessons of history, you are doomed to repeat the same mistakes.

• **Previous SEO Work**

When you are brought on to handle the SEO for a particular website, one of the first things you need to find out is which SEO activities have previously been attempted. There may be valuable data there, such as a log of changes that you can match up with analytics data to gauge impact.

If no such log exists, you can always check the Wayback Machine to see whether it has historical logs for your website. This offers snapshots of what the site looked like at various points in time. Even if a log was not kept, spend some time building a timeline of when any of the types of changes that affect SEO took place. In particular, see whether you can get copies of the exact recommendations the prior SEO consultant made, as this will help you with the timeline and the specifics of the changes made. You should also pay particular attention to understanding the types of link-building activities that took place. Were shady practices used that carry a lot of risk? Was there a particular link-building tactic that worked quite well? Going through the history of the link-building efforts can yield tons of information that you can use to determine your next steps.

Benchmarking Current Indexing Status

The search engines have an enormous task: that of indexing the world's online content—well, more or less. The reality is that they try hard to discover all of it, but they do not choose to include all of it in their indexes. There can be a variety of reasons for this, such as the page being inaccessible to the spider, being penalized, or not having enough link juice to merit inclusion. When you launch a new site or add new sections to an existing site, or if you are dealing with a very large site, not every page will necessarily make it into the index. To get a handle on this you will want to actively track the indexing level of your site. If your site is not fully indexed, it could be a sign of a problem (not enough links, poor site structure, etc.). Getting basic indexation data from search engines is pretty easy. All three major search engines support the same basic syntax for that: `site:yourdomain.com`. Short-term spikes are not a cause for concern, nor are periodic drops in levels of crawling. What is important is the general trend. For the other search engines,

NATIONAL COMPUTER COLLEGE

the crawl-related data can then be revealed using logfile analyzers, and then a similar timeline can be created and monitored.

Benchmarking Current Rankings

People really love to check their search rankings. Many companies want to use this as a measurement of SEO progress over time, but it is a bit problematic, for a variety of reasons. Here is a summary of the major problems with rank checking:

- Google results are not consistent:
 - Different geographies often give different results.
 - Different data centers give different results.
 - Results are personalized for logged-in users based on their search histories.
 - No rank checker can monitor and report all of these inconsistencies.
- The Google API rarely matches up to what anyone sees in the search results:
 - It appears to match up only on very heavily trafficked, consistent search results; anything mid-tail or long tail is invariably inaccurate.
 - It is extremely slow to update, so even though news results or geographic results might be mixed in, the API won't update for days or sometimes weeks.
- Obsessing over rankings (rather than traffic) can result in poor strategic decisions:
 - When sites obsess over rankings for particular keywords, the time and energy they expend on those few key phrases often produces far less value than would have been produced if they had spent those resources on the site as a whole.
 - Long-tail traffic very often accounts for 70% to 80% of the demand curve, and it is much easier to rank in the long tail and get valuable traffic from there than it is to concentrate on the few rankings at the top of the demand curve.

Benchmarking Current Traffic Sources and Volume

The most fundamental objective of any SEO project should be to drive the bottom line. For a business, this means delivering more revenue with favorable ROI. As a precursor to determining the level of ROI impact, the SEO practitioner must focus on increasing the volume of relevant traffic to the site. This is a more important objective than anything related to rankings or number of links obtained. More relevant traffic should mean more revenue for the business. Today's web analytics tools make the gathering of such data incredibly easy. Two high-quality solutions are available that are completely free: Google Analytics and Yahoo! Web Analytics. These tools are sufficient for many smaller sites, though larger sites will probably need to consider a paid solution such as Omniture.com, IBM Unica NetInsight, or Webtrends.com. This type of data allows you to see which search engines are delivering the majority of the traffic to your site, and perhaps flag potential problems.

Leveraging Business Assets for SEO

Chances are your company/organization has a lot of valuable commodities beyond the website that can be put to good use to improve the quality and quantity of traffic you receive through search engine optimization efforts. We discuss some of these things in the subsections that follow.

• Other Domains You Own/Control

If you have multiple domains, the major items to think about are:

- Can you 301-redirect some of those domains back to your main domain or to a subfolder on the site for additional benefit?
- Do you own exact keyword match domain names that would make for effective microsites?
- If you're maintaining those domains as separate sites, are you linking between them intelligently?

If any of those avenues produce valuable strategies, pursue them—remember that it is often far easier to optimize what you're already doing than to develop entirely new strategies, content, and processes. Particularly on the link-building side, this is some of the lowest-hanging fruit around.

• Partnerships On and Off the Web

Partnerships can be leveraged in similar ways, particularly on the link-building front. If you have business partners that you supply or otherwise work with—or from whom you receive service—chances are good that you can implement link strategies between their sites and yours. Although reciprocal linking carries a

NATIONAL COMPUTER COLLEGE

bit of a bad reputation, there is nothing wrong with building a “partners,” “clients,” “suppliers,” or “recommended” list on your site, or with requesting that your organizational brethren do likewise for you. Just do this in moderation and make sure you link only to highly relevant, trusted sites.

- **Content or Data you’ve Never Put Online**

Chances are that you have content that you have never published on your website. This content can be immensely valuable to your SEO efforts. However, many companies are not savvy to the nuances of publishing that content in a manner that is friendly to search engines. Those hundreds of lengthy articles you published when you were shipping a print publication via the mail are a great fit for your website archives.

- **Customers Who Have Had a Positive Experience**

Customers are a terrific resource for earning links, but did you also know they can write? Customers and website visitors can contribute all kinds of content. Seriously, if you have user-generated content (UGC) options available to you and you see value in the content your users produce, by all means reach out to customers, visitors, and email list subscribers for both links and content opportunities.

- **Your Fans**

This principle applies equally to generic enthusiasts of your work. For many businesses that operate offline or work in entertainment, hard goods, or any consumer services, chances are good that if your business is worth its salt, there are people out there who’ve used your products or services and would love to share their experiences. Do you make video games? Reach out to your raving fans. Written a book? Mobilize your literary customers on the Web. Organize events? Like customers, fans are terrific resources for link acquisition, content creation, positive testimonials, and social media marketing.

Combining Business Assets and Historical Data to Conduct SEO/

Website SWOT Analysis

A classic staple of business school is the SWOT analysis—identifying the strengths, weaknesses, opportunities, and threats faced by a business or project. By combining data from your business asset assessment and historical tracking data (and visitor analytics), you can create some very compelling analyses of your organization and its marketplace.

Identifying **strengths** is typically one of the easier objectives:

- What sources of traffic are working well for your site/business?
- Which projects/properties/partnerships are driving positive momentum toward traffic/revenue goals?
- Which of your content sections/types produces high traffic and ROI?
- What changes have you made historically that produced significant value?

Determining the **weaknesses** can be tougher (and takes more intellectual honesty and courage):

- What content is currently driving low levels of search/visitor traffic?
- Which changes that were intended to produce positive results have shown little/no value?
- Which traffic sources are underperforming or under delivering?
- What projects/properties/partnerships are being leveraged poorly?

Parsing **opportunities** requires a combination of strength and weakness analysis. You want to find areas that are doing well but have room to expand, as well as those that have yet to be explored:

- What brainstormed but undeveloped or untested projects/ideas can have a significant, positive impact?
- What traffic sources currently sending good-quality traffic could be expanded to provide more value?
- What areas of weakness have direct paths to recovery?
- Which website changes have had positive results? Can these be applied more rigorously or to other areas for increased benefit?
- What new markets or new content areas are potentially valuable for expansion?
- What sources of new content/new links have yet to be tapped?

Determining **threats** can be the most challenging of the tasks. You’ll need to combine creative thinking with an honest assessment of your weaknesses and your competitors’ strengths, and consider the possibilities of macro-events that could shape your website/company’s future:

- In your areas of weakness, which players in your market (or other, similar markets) are strong? How have they accomplished this?

NATIONAL COMPUTER COLLEGE

- What shifts in human behavior, web usage, or market conditions could dramatically impact your business/site? (For example, consider the “what if people stopped searching and instead navigated the Web in different ways” perspective. It is a bit “pie in the sky,” but we have already seen Expedia partially destroy the travel agency business, Craigslist make classifieds obsolete, and Facebook start to take advertising market share from the search engines.)
 - Which competitors have had the most success in your arena? How have they accomplished this? Where do they intersect with your business/customers?
 - Are there any strategies implemented by start-ups in similar businesses that have had massive success in a particular arena that could be dangerous to your business if they were replicated in your market?
- Conducting SWOT analysis from a web marketing and SEO perspective is certainly one of the most valuable first steps you can take as an organization poised to expend resources. If you haven't taken the time to analyze the landscape from these bird's-eye-view perspectives, you might end up like a great runner who's simply gone off the course—sure, you'll finish fast, but where will it take you?