

**THE BEGINNERS GUIDE TO SEO**

**PDF**

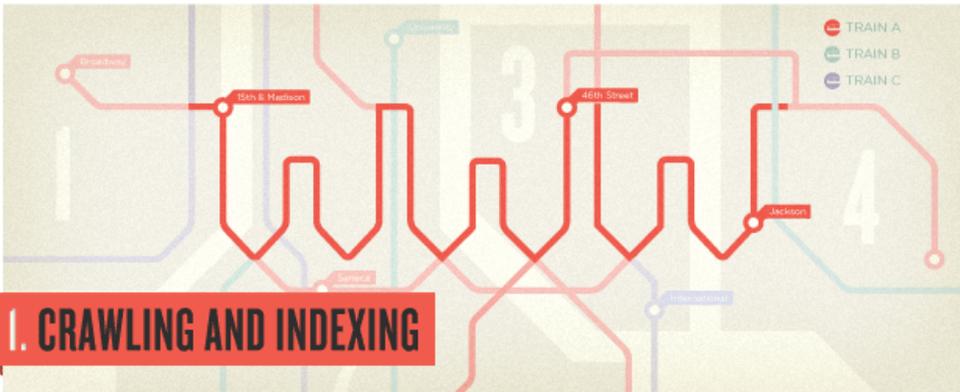
# 1

## CHAPTER ONE

# HOW SEARCH ENGINES OPERATE

Search engines have two major functions - crawling & building an index, and providing answers by calculating relevancy & serving results.

1. **Crawling and Indexing**  
Crawling and indexing the billions of documents, pages, files, news, videos and media on the world wide web.
2. **Providing Answers**  
Providing answers to user queries, most frequently through lists of relevant pages, through retrieval and rankings.



**Imagine the World Wide Web as a network of stops in a big city subway system.**

Each stop is its own unique document (usually a web page, but sometimes a PDF, JPG or other file). The search engines need a way to “crawl” the entire city and find all the stops along the way, so they use the best path available – links.

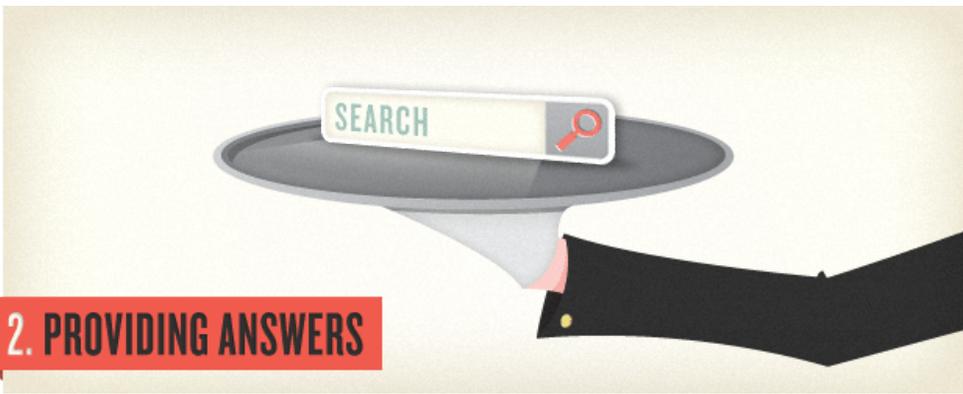
**“The link structure of the web serves to bind all of the pages together.”**

Through links, search engines’ automated robots, called “crawlers,” or “spiders” can reach the many billions of interconnected documents.

Once the engines find these pages, they next decipher the code from them and store selected pieces in massive hard drives, to be recalled later when needed for a search query. To accomplish the monumental task of holding billions of pages that can be accessed in a fraction of a second, the search engines have constructed datacenters all over the world.

These monstrous storage facilities hold thousands of machines processing large quantities of information. After all, when a person performs a search at any of the major engines, they demand results instantaneously – even a 1 or 2 second delay can cause dissatisfaction, so the engines work hard to provide answers as fast as possible.





## 2. PROVIDING ANSWERS

Search engines are **answer machines**. When a person looks for something online, it requires the search engines to scour their corpus of billions of documents and do two things – first, return only those results that are relevant or useful to the searcher’s query, and second, rank those results in order of perceived usefulness. It is both “**relevance**” and “**importance**” that the process of SEO is meant to influence.

To a search engine, relevance means more than simply finding a page with the right words. In the early days of the web, search engines didn’t go much further than this simplistic step, and their results suffered as a consequence. Thus, through evolution, smart engineers at the engines devised better ways to find valuable results that searchers would appreciate and enjoy. Today, 100s of factors influence relevance, many of which we’ll discuss throughout this guide.

### *How Do Search Engines Determine Importance?*

Currently, *the major engines typically interpret importance as popularity* – the more popular a site, page or document, the more valuable the information contained therein must be. This assumption has proven fairly successful in practice, as the engines have continued to increase users’ satisfaction by using metrics that interpret popularity.

Popularity and relevance aren’t determined manually. Instead, the engines craft careful, mathematical equations – algorithms – to sort the wheat from the chaff and to then rank the wheat in order of tastiness (or however it is that farmers determine wheat’s value).

These algorithms are often comprised of hundreds of components. In the search marketing field, we often refer to them as “ranking factors” SEOMoz crafted a resource specifically on this subject – [Search Engine Ranking Factors](#).



*You can surmise that search engines believe that Ohio State is the most **relevant** and **popular** page for the query “Universities” while the result, Harvard, is less relevant/popular.*

---

## So How Do I Get Some Success Rolling In?

### *or "How Search Marketers Succeed"*

The complicated algorithms of search engines may appear at first glance to be impenetrable. The engines themselves provide little insight into how to achieve better results or garner more traffic. What information on optimization and best practices that the engines themselves do provide is listed below:





## SEO INFORMATION FROM GOOGLE WEBMASTER GUIDELINES

Googlers recommend the following to get better rankings in their search engine:

- \* *Make pages primarily for users, not for search engines. Don't deceive your users or present different content to search engines than you display to users, which is commonly referred to as cloaking.*
- \* *Make a site with a clear hierarchy and text links. Every page should be reachable from at least one static text link.*
- \* *Create a useful, information-rich site, and write pages that clearly and accurately describe your content. Make sure that your <title> elements and ALT attributes are descriptive and accurate.*
- \* *Use keywords to create descriptive, human friendly URLs. Provide one version of a URL to reach a document, using 301 redirects or the rel="canonical" element to address duplicate content.*



## SEO INFORMATION FROM BING WEBMASTER GUIDELINES

Bing engineers at Microsoft recommend the following to get better rankings in their search engine:

- \* *Ensure a clean, keyword rich URL structure is in place*
- \* *Make sure content is not buried inside rich media (Adobe Flash Player, JavaScript, Ajax) and verify that rich media doesn't hide links from crawlers.*
- \* *Create keyword-rich content based on research to match what users are searching for. Produce fresh content regularly.*
- \* *Don't put the text that you want indexed inside images. For example, if you want your company name or address to be indexed, make sure it is not displayed inside a company logo.*

*So what you're telling me is that this is just the tip of the search marketing iceberg and there's a ton more?*

*yes.*

## But Have No Fear Fellow Search Marketer!

Over the 15 plus years that web search has existed, search marketers have found methods to extract information about how the search engines rank pages. SEOs and marketers use that data

to help their sites and their clients achieve better positioning.

Surprisingly, the engines support many of these efforts, though the public visibility is frequently low. Conferences on search marketing, such as the [Search Marketing Expo](#), [Pubcon](#), [Search Engine Strategies](#), [Distilled](#) & SEOMoz's own [MozCon](#) attract engineers and representatives from all of the major engines. Search representatives also assist webmasters by occasionally participating online in blogs, forums & groups.



### « TIME FOR AN EXPERIMENT »

There is perhaps no greater tool available to webmasters researching the activities of the engines than the freedom to use the search engines to perform experiments, test theories and form opinions. It is through this iterative, sometimes painstaking process, that a considerable amount of knowledge about the functions of the engines has been gleaned.

1. Register a new website with nonsense keywords (e.g. *ishkabibbell.com*)
2. Create multiple pages on that website, all targeting a similarly ludicrous term (e.g. *yoogewgally*)
3. Test the use of different placement of text, formatting, use of keywords, link structures, etc by making the pages as uniform as possible with only a singular difference
4. Point links at the domain from indexed, well-spidered pages on other domains
5. Record the search engines' activities and the rankings of the pages
6. Make small alterations to the identically targeting pages to determine what factors might push a result up or down against its peers
7. Record any results that appear to be effective and re-test on other domains or with other terms – if several tests consistently return the same results, chances are you've discovered a pattern that is used by the search engines.

## AN EXAMPLE TEST WE WHIPPED UP

In this test, we started with the hypothesis that a link **higher up in a page's code** carries more weight than a page lower down in the code. We tested this by creating a nonsense domain linking out to three pages, all carrying the same nonsense word exactly once. After the engines spidered out the pages, we found that the page linked to from the highest link on the home page ranked first.

***This process is not alone in helping to educate search marketers.***

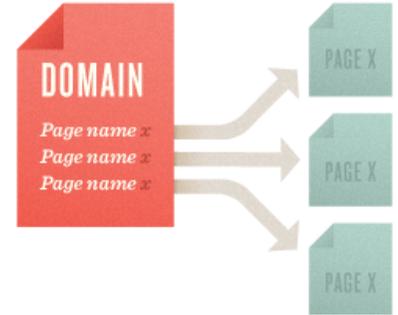
Competitive intelligence about signals the engines might use and how they might order results is

also available through patent applications made by the major engines to the United States Patent Office. Perhaps the most famous among these is the system that spawned Google's genesis in the Stanford dormitories during the late 1990's – PageRank – documented as [Patent #6285999](#) – *Method for node ranking in a linked database*. The original paper on the subject – [Anatomy of a Large-Scale Hypertextual Web Search Engine](#) – has also been the subject of considerable study. To those whose comfort level with complex mathematics falls short, never fear. Although the actual equations can be academically interesting, complete understanding evades many of the most talented search marketers. Remedial calculus isn't required to practice SEO!

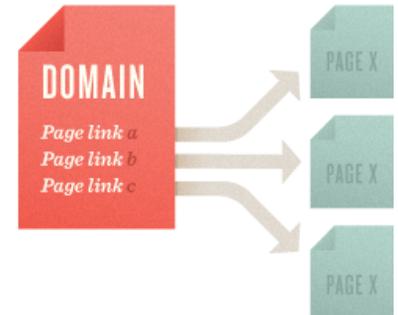
Through methods like patent analysis, experiments, and live testing, search marketers as a community have come to understand many of the basic operations of search engines and the critical components of creating websites and pages that earn high rankings and significant traffic.

***The rest of this guide is devoted to clearly explaining these practices. Enjoy!***

## *“Step” to “Domain”*



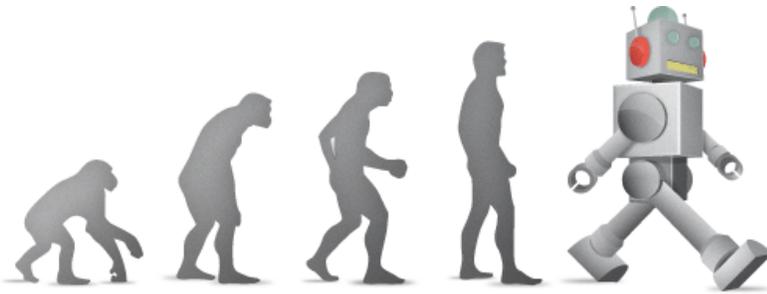
## *“Step” to “Domain”*



## CHAPTER TWO

# HOW PEOPLE INTERACT WITH SEARCH ENGINES

One of the most important elements to building an online marketing strategy around SEO is empathy for your audience. Once you grasp what the average searcher, and more specifically, your target market, is looking for, you can more effectively reach and keep those users.



We like to say "**Build for users, not search engines.**" When users have a bad experience at your site, when they can't accomplish a task or find what they were looking for, this often correlates with poor search engine performance. On the other hand, when users are happy with your website, a positive experience is created, both with the search engine and the site providing the information or result.

What are users looking for? There are three types of search queries users generally perform:

- \* **"Do"** Transactional Queries - Action queries such as buy a plane ticket or listen to a song.
- \* **"Know"** Informational Queries - When a user seeks information, such as the name of the band or the best restaurant in New York City.
- \* **"Go"** Navigation Queries - Search queries that seek a particular online destination, such as Facebook or the homepage of the NFL.

When visitors type a query into a search box and land on your site, will they be satisfied with what they find? This is the primary question search engines try to figure out millions of times per day.

**The search engines' primary responsibility is to serve relevant results to their users.**

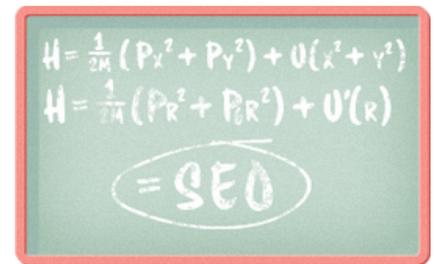
It all starts with the words typed into a small box.

Search engine usage has evolved over the years but the primary principles of conducting a search remain largely unchanged. Listed here are the steps that comprise most search processes:

1. Experience the need for an answer, solution or piece of information.
2. Formulate that need in a string of words and phrases, also known as "the query."
3. Enter the query into a search engine.
4. Browse through the results for a match.
5. Click on a result.
6. Scan for a solution, or a link to that solution.
7. If unsatisfied, return to the search results and browse for another link or...
8. Perform a new search with refinements to the query.

# A Broad Picture With Fascinating Data

Why invest time, effort and resources on SEO? When looking at the broad picture of search engine usage, fascinating data is available from several studies. We've extracted those that are recent, relevant, and valuable, not only for understanding how users search, but to help present a compelling argument about the power of search.



## Google leads the way in an October 2011 study by comScore:

- \* Google Sites led the U.S. core search market in April with 65.4 percent of the searches conducted, followed by Yahoo! Sites with 17.2 percent, and Microsoft Sites with 13.4 percent. (Microsoft powers Yahoo Search. In the real world, most webmasters see a much higher percentage of their traffic from Google than these numbers suggest.)
- \* Americans alone conducted a staggering 20.3 billion searches in one month. Google Sites accounted for 13.4 billion searches, followed by Yahoo! Sites (3.3 billion), Microsoft Sites (2.7 billion), Ask Network (518 million) and AOL LLC (277 million).
- \* Total search powered by Google properties equaled 67.7 percent of all search queries, followed by Bing which powered 26.7 percent of all search. (Microsoft powers Yahoo Search. In the real world, most webmasters see a much higher percentage of their traffic from Google than these numbers suggest.)

[VIEW ONLINE](#)

## Billions spent on online marketing from an August 2011 Forrester report:

- \* Interactive marketing will near \$77 billion in 2016.
- \* This spend will represent 26% of all advertising budgets combined.

[VIEW ONLINE](#)

## Search is the new Yellow Pages from a Burke 2011 report:

- \* 76% of respondents used search engines to find local business information vs. 74% who turned to print yellow pages.
- \* 57% who used Internet yellow pages, and 44% who used traditional newspapers.
- \* 67% had used search engines in the past 30 days to find local information, and 23% responded that they had used online social networks as a local media source.

[VIEW ONLINE](#)

## An August 2011 PEW Internet Study revealed:

- \* The percentage of Internet users who use search engines on a typical day has been steadily rising from about one-third of all users in 2002, to a new high of 59% of all adult Internet users.
- \* With this increase, the number of those using a search engine on a typical day is pulling ever closer to the 61 percent of Internet users who use e-mail, arguably the Internet's all-time killer app, on a typical day.

[VIEW ONLINE](#)

## StatCounter Global Stats Reports the top 5 Search Engines Sending Traffic Worldwide:

- \* Google sends 90.62% of traffic.
- \* Yahoo! sends 3.78% of traffic.
- \* Bing sends 3.72% of traffic.
- \* Ask Jeeves sends .36% of traffic.
- \* Baidu sends .35% of traffic.

[VIEW ONLINE](#)

## A 2011 Study by Slingshot SEO Reveals Click-through Rates for Top Rankings:

- \* A #1 position in Google's search results receives 18.2% of all click-through traffic.
- \* The second position receives 10.1%, the third 7.2%, the fourth 4.8%, and all others are under 2%.
- \* A #1 position in Bing's search results averages a 9.66% click-through rate.
- \* The total average CTR for first ten results was 52.32% for Google and 26.32% for Bing.

[VIEW ONLINE](#)

[view](#)



# That's Some Spicy Data You Got There!

*This particular study perfectly illustrated how little attention is paid to results on the page vs. those higher up, and how users' eyes are drawn to bolded keywords, titles, and descriptions in the organic results vs. The paid search listings.*



All of this impressive research data leads us to important conclusions about web search and marketing through search engines. In particular, we're able to make the following statements:

- \* *Search is very, very popular. Growing strong at nearly 20% a year, it reaches nearly every online American, and billions of people around the world.*
- \* *Search drives an incredible amount of both online and offline economic activity.*
- \* *Higher rankings in the first few results are critical to visibility.*
- \* *Being listed at the top of the results not only provides the greatest amount of traffic, but instills trust in consumers as to the worthiness and relative importance of the company/website.*

Learning the foundations of SEO is a vital step in achieving these goals.

*“As marketers, the Internet as a whole and search, specifically, are undoubtedly one of the best and most important ways to reach consumers and build a business, no matter the size, reach, or focus.”*



## CHAPTER THREE

# WHY SEARCH ENGINE MARKETING IS NECESSARY

An important aspect of Search Engine Optimization is making your website easy for both users and search engine robots to understand. Although search engines have become increasingly sophisticated, in many ways they still can't see and understand a web page the same way a human does. SEO helps the engines figure out what each page is about, and how it may be useful for users.

### ***A Common Argument Against SEO***

We frequently hear statements like this:

*"No smart engineer would ever build a search engine that requires websites to follow certain rules or principles in order to be ranked or indexed. Anyone with half a brain would want a system that can crawl through any architecture, parse any amount of complex or imperfect code and still find a way to return the best and most relevant results, not the ones that have been "optimized" by unlicensed search marketing experts."*

### ***But Wait...***

Imagine you posted online a picture of your family dog. A human might describe it as "a black, medium-sized dog - looks like a Lab, playing fetch in the park." On the other hand, the best search engine in the world would struggle to understand the photo at anywhere near that level of sophistication. How do you make a search engine understand a photograph? Fortunately, SEO allows webmasters to provide "clues" that the engines can use to understand content. In fact, adding proper structure to your content is essential to SEO.

Understanding both the abilities and limitations of search engines allows you to properly build, format and annotate your web content in a way that search spiders can digest. Without SEO, many websites remain invisible to search engines.



## Limitations x3

### ***The Limits of Search Engine Technology***

The major search engines all operate on the same principles, as explained in [Chapter 1](#). Automated search bots crawl the web, follow links and index content in massive databases. They accomplish this with a type of dazzling artificial intelligence that is nothing short of amazing. That said, modern search technology is not all-powerful. There are [technical limitations of all kinds](#) that cause immense problems in both inclusion and rankings. We've listed the most common below:

## 1. Spidering and Indexing Problems

- \* Search engines aren't good at completing online forms (such as a login), and thus any content contained behind them may remain hidden.
- \* Websites using a CMS (Content Management System) often create duplicate versions of the same page - a major problem for search engines looking for completely original content.
- \* Errors in a website's crawling directives (robots.txt) may lead to blocking search engines entirely.
- \* Poor link structures lead to search engines failing to reach all of a website's content. In other cases, poor link structures allow search engines to spider content, but leave it so minimally exposed that it's deemed "unimportant" by the engine's index.

### Interpreting Non-Text Content

- \* Although the engines are getting better at reading non-HTML text, content in rich media format is traditionally difficult for search engines to parse.
- \* This includes text in Flash files, images, photos, video, audio & plug-in content.

## 3. The "Tree Falls in a Forest"

SEO isn't just about getting the technical details of search-engine friendly web development correct. **It's also about marketing.** This is perhaps the most important concept to grasp about the functionality of search engines. You can build a perfect website, but its content can remain invisible to search engines unless you promote it. This is due to the nature of search technology, which relies on the metrics of **relevance** and **importance** to display results.

The "tree falls in a forest" adage postulates that if no one is around to hear the sound, it may not exist at all - and this translates perfectly to search engines and web content. Put another way - if no one links to your content, the search engines may choose to ignore it.

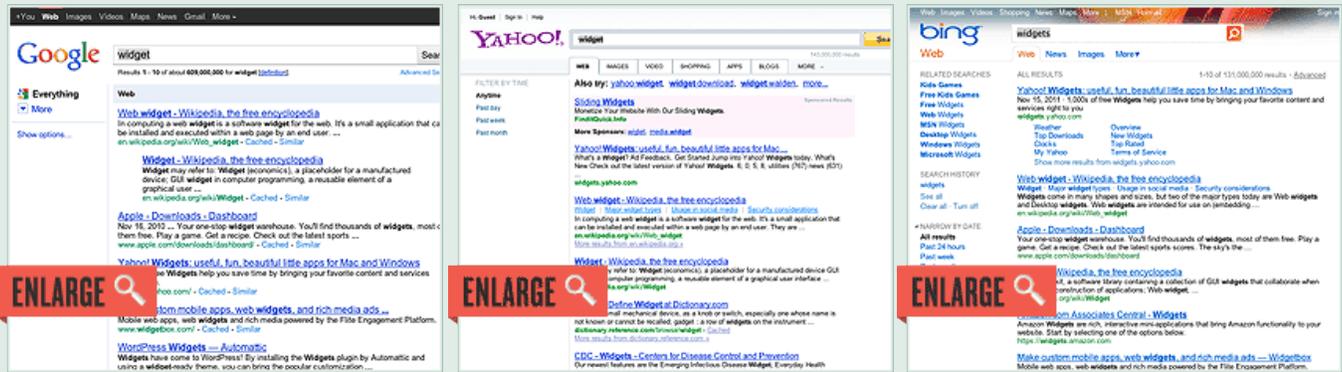
The engines by themselves have no inherent gauge of quality and no potential way to discover fantastic pieces of content on the web. Only humans have this power - to discover, react, comment and link to. Thus, great content cannot simply be created - it must be shared and talked about. Search engines already do a great job of promoting high quality content on websites that have become popular, but they cannot *generate* this popularity - this is a task that demands talented Internet marketers.

## 2. Content to Query Matching

- \* Text that is not written in common terms that people use to search. For example, writing about "food cooling units" when people actually search for "refrigerators".
- \* Language and internationalization subtleties. For example, color vs colour. When in doubt, **check what people are searching for** and use exact matches in your content.
- \* Location targeting, such as targeting content in Polish when the majority of the people who would visit your website are from Japan.
- \* Mixed contextual signals. For example, the title of your blog post is "Mexico's Best Coffee" but the post itself is about a vacation resort in Canada which happens to serve great coffee. These mixed messages send confusing signals to search engines.



Take a look at any search results page and you'll find the answer to why search marketing has a long, healthy life ahead.

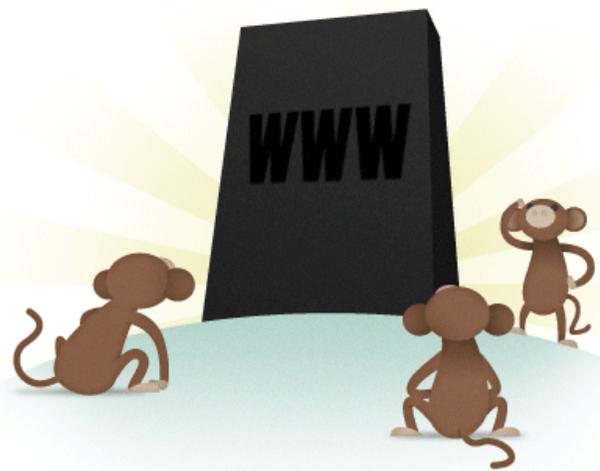


Ten positions, ordered by rank, with click-through traffic based on their relative position & ability to attract searchers. Results in positions 1, 2 and 3 receive much more traffic than results down the page, and considerably more than results on deeper pages. The fact that so much attention goes to so few listings means that there will always be a financial incentive for search engine rankings. No matter how search may change in the future, websites and businesses will compete with one another for this traffic, branding, and visibility it provides.

## A Constantly Shifting Landscape

When search marketing began in the mid-1990's, manual submission, the meta keywords tag and keyword stuffing were all regular parts of the tactics necessary to rank well. In 2004, link bombing with anchor text, buying hordes of links from automated blog comment spam injectors and the construction of inter-linking farms of websites could all be leveraged for traffic. In 2011, social media marketing and vertical search inclusion are mainstream methods for conducting search engine optimization.

The future is uncertain, but in the world of search, change is a constant. For this reason, search marketing will remain a steadfast need for those who wish to remain competitive on the web. Others have claimed that SEO is dead, or that SEO amounts to spam. As we see it, there's no need for a defense other than simple logic - websites compete for attention and placement in the search engines, and those with the best knowledge and experience with these rankings receive the benefits of increased traffic and visibility.





# 4

## CHAPTER FOUR

# THE BASICS OF SEARCH ENGINE FRIENDLY DESIGN & DEVELOPMENT

Search engines are limited in how they crawl the web and interpret content. A webpage doesn't always look the same to you and I as it looks to a search engine. In this section, we'll focus on specific technical aspects of building (or modifying) web pages so they are structured for both search engines and human visitors alike. This is an excellent part of the guide to share with your programmers, information architects, and designers, so that all parties involved in a site's construction can plan and develop a search-engine friendly site.

## Indexable Content

In order to be listed in the search engines, your most important content should be in HTML text format. Images, Flash files, Java applets, and other non-text content are often ignored or devalued by search engine spiders, despite advances in crawling technology. The easiest way to ensure that the words and phrases you display to your visitors are visible to search engines is to place it in the HTML text on the page. However, more advanced methods are available for those who demand greater formatting or visual display styles:

1. Images in gif, jpg, or png format can be assigned "alt attributes" in HTML, providing search engines a text description of the visual content.
2. Search boxes can be supplemented with navigation and crawlable links.
3. Flash or Java plug-in contained content can be supplemented with text on the page.
4. Video & audio content should have an accompanying transcript if the words and phrases used are meant to be indexed by the engines.

### Seeing Like a Search Engine

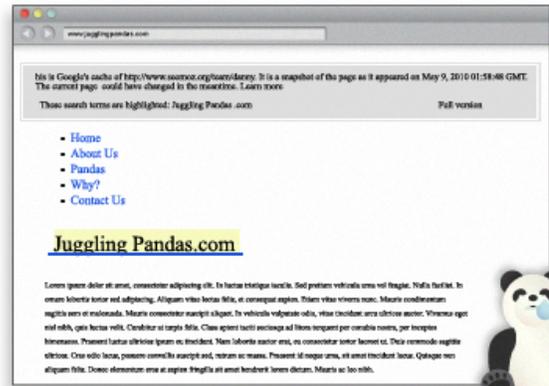
Many websites have significant problems with indexable content, so double-checking is worthwhile. By using tools like Google's cache, [SEO-browser.com](http://SEO-browser.com), or the [MozBar](http://MozBar) you can see what elements of your content are visible and indexable to the engines. Take a look at [Google's text cache of this page you are reading now](#). See how different it looks?



*"I think I have a problem with getting found. I built this huge flash site for juggling pandas and I'm showing up nowhere on Google. What's up?"*



Through Browser

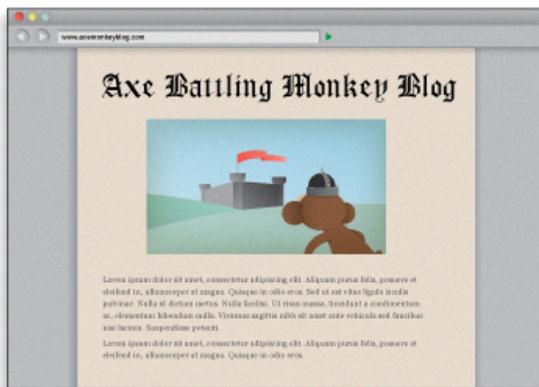


Through Google Cache

### Whoa! That's what we look like?

Using the Google cache feature, we're able to see that to a search engine, JugglingPandas.com's homepage doesn't contain all the rich information that we see. This makes it difficult for search engines to interpret relevancy.

### "I'm totally going to check out my Axe Battling Monkeys blog!"



What Humans See



What Search Engines See

### That's a lot of monkeys, and just headline text?

### Hey, where did the fun go?

Uh oh... via Google cache, we can see that the page is a barren wasteland. There's not even text telling us that the page contains the Axe Battling Monkeys. The site is entirely built in Flash, but sadly, this means that search engines cannot index any of the text content, or even the links to the individual games. Without any HTML text, this page would have a very hard time ranking in search results.

It's wise to not only check for text content but to also use SEO tools to double-check that the pages you're building are visible to the engines. This applies to your images, and as we see below, your links as well.



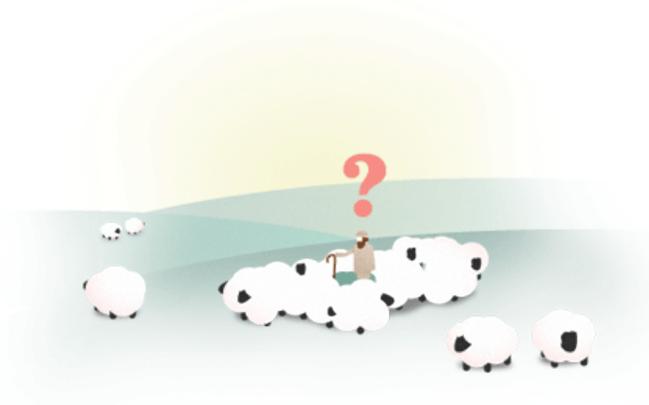
# Crawlable Link Structures

Just as search engines need to see content in order to list pages in their massive keyword-based indices, they also need to see links in order to find the content. A crawlable link structure - one that lets their spiders browse the pathways of a website - is vital in order to find all of the pages on a website. Hundreds of thousands of sites make the critical mistake of structuring their navigation in ways that search engines cannot access, thus impacting their ability to get pages listed in the search engines' indices.

Below, we've illustrated how this problem can happen:



In the example above, Google's spider has reached page "A" and sees links to pages "B" and "E". However, even though C and D might be important pages on the site, the spider has no way to reach them (or even know they exist.) This is because no direct, crawlable links point to those pages. As far as Google is concerned, they might as well not exist - great content, good keyword targeting, and smart marketing won't make any difference at all if the spiders can't reach those pages in the first place.



## ANATOMY OF A LINK

```
<a href="http://www.jonwye.com">Jon Wye's Custom Designed Belts</a>
```

start of link tag

link referral location

visible/anchor text of link

closure of link tag

In the above illustration, the "<a" tag indicates the start of a link. Link tags can contain images, text, or other objects, all of which provide a clickable area on the page that users can engage to move to another page. This is the original navigational element of the Internet - "hyperlinks". The link referral location tells the browser (and the search engines) where the link points to. In this example, the URL <http://www.jonwye.com> is referenced. Next, the visible portion of the link for visitors, called "**anchor text**" in the SEO world, describes the page the link points to. The page pointed to is about custom belts, made by my friend from Washington D.C., Jon Wye, so I've used the anchor text "Jon Wye's Custom Designed Belts". The </a> tag closes the link, so that elements later on in the page will not have the link attribute applied to them.

This is the most basic format of a link - and it is eminently understandable to the search engines. The spiders know that they should add this link to the engines' link graph of the web, use it to calculate query-independent variables (like Google's PageRank), and follow it to index the contents of the referenced page.

## Let's look at some common reasons why pages may not be reachable.

### Submission-required forms

If you require users to complete an online form before accessing certain content, chances are search engines may never see those protected pages. Forms can include a password protected login or a full-blown survey. In either case, search spiders generally will not attempt to "submit" forms and thus, any content or links that would be accessible via a form are invisible to the engines.

### Links in un-parseable Javascript

If you use Javascript for links, you may find that search engines either do not crawl or give very little weight to the links embedded within. Standard HTML links should replace Javascript (or accompany it) on any page where you'd like spiders to crawl.

### Links pointing to pages blocked by the meta robots tag or robots.txt

The [Meta Robots tag and the Robots.txt](#) file both allow a site owner to restrict spider access to a page. Just be warned that many a webmaster has unintentionally used these directives as an attempt to block access by rogue bots, only to discover that search engines cease their crawl.

### Frames or I-frames

Technically, links in both frames and I-Frames are crawlable, but both present structural issues for the engines in terms of organization and following. Unless you're an advanced user with a good technical understanding of how search engines index and follow links in frames, it's best to stay away from them.

### Robots don't use search forms

Although this relates directly to the above warning on forms, it's such a common problem that it bears mentioning. Some webmasters believe if they place a search box on their site, then engines will be able to find everything that visitors search for. Unfortunately, spiders don't perform searches to find content, and thus, it's millions of pages are hidden behind inaccessible walls, doomed to anonymity until a spidered page links to it.

### Links in flash, java, or other plug-ins

The links embedded inside the Panda site (from our above example) is a perfect illustration of this phenomenon. Although dozens of pandas are listed and linked to on the Panda page, no spider can reach them through the site's link structure, rendering them invisible to the engines (and un-retrievable by searchers performing a query).

### Links on pages with many hundreds or thousands of links

Search engines will only crawl so many links on a given page - not an infinite amount. This loose restriction is necessary to cut down on spam and conserve rankings. Pages with 100's of links on them are at risk of not getting all of those links crawled and indexed.

*If you avoid these pitfalls, you'll have clean, spiderable HTML links that will allow the spiders easy access to your content pages.*

## rel="nofollow"

Rel="nofollow" can be used with the following syntax:

```
<a href="http://www.seomoz.org" rel="nofollow">Lousy Punks!</a>
```

Links can have lots of attributes applied to them, but the engines ignore nearly all of these, with the important exception of the rel="nofollow" tag. In the example above, by adding the rel=nofollow attribute to the link tag, we've told the search engines that we, the site owners, **do not want this link to be interpreted as the normal, "editorial vote."**

Nofollow, taken literally, instructs search engines to not follow a link (although some do.) The nofollow tag came about as a method to help stop automated blog comment, guest book, and link injection spam ([read more about the launch here](#)), but has morphed over time into a way of telling the engines to discount any link value that would ordinarily be passed. Links tagged with

### Google

**Google states that in most cases, they don't follow nofollowed links, nor do these links transfer PageRank or anchor text values. Essentially, using nofollow causes us to drop the target links from our overall graph of the web. Nofollowed links carry no weight and are interpreted as HTML text (as though the link did not exist). That said, many webmasters believe that even a nofollow link from a high authority site, such as Wikipedia, could be**

nofollow are interpreted slightly differently by each of the engines, but it is clear they do not pass as much weight as normal "followed" links.

### Are nofollow Links Bad?

Although they don't pass as much value as their followed cousins, nofollowed links are a natural part of a diverse link profile. A website with lots of inbound links will accumulate many nofollowed links, and this isn't a bad thing. In fact, SEOMoz's [Ranking Factors](#) showed that high ranking sites tended to have a higher percentage of inbound nofollowed links than lower ranking sites.

*interpreted as a sign of trust.*

### Bing & Yahoo!

*Bing, which powers Yahoo search results, has also stated that they do not include nofollowed links in the link graph. In the past, they have also stated nofollowed links may still be used by their crawlers as a way to discover new pages. So while they "may" follow the links, they will not count them as a method for positively impacting rankings.*

## Keyword Usage & Targeting

Keywords are fundamental to the search process - they are the building blocks of language and of search. In fact, the entire science of information retrieval (including web-based search engines like Google) is based on keywords. As the engines crawl and index the contents of pages around the web, they keep track of those pages in keyword-based indices. Thus, rather than storing 25 billion web pages all in one database, the engines have millions and millions of smaller databases, each centered on a particular keyword term or phrase. This makes it much faster for the engines to retrieve the data they need in a mere fraction of a second.

Obviously, if you want your page to have a chance of ranking in the search results for "dog," it's wise to make sure the word "dog" is part of the indexable content of your document.

STEP 1: RECEIVE QUERY



STEP 2: FIND RELEVANT DATABASE

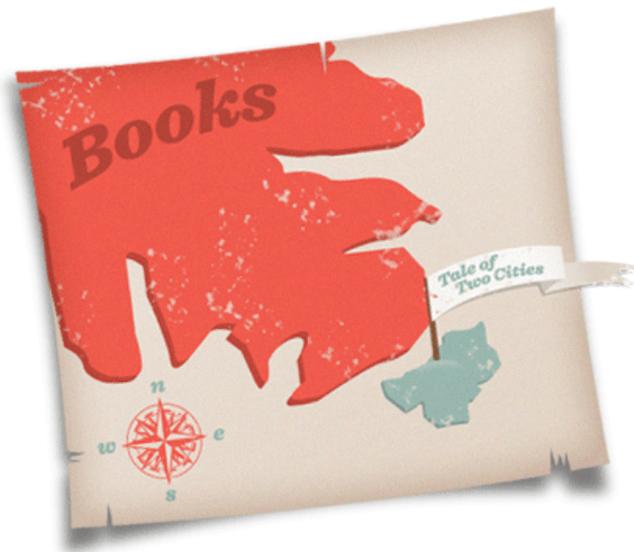
"DOG"	
http://www.thedog.com	PR 9
http://www.dogdog.com	PR 7
http://www.goddog.com	PR 6
http://www.dog.com	PR 3
http://www.dogyear.com	PR 5
http://www.baddog.com	PR 2

STEP 3: RANK DOCUMENTS



"DOG"	
http://www.thedog.com	PR 9
http://www.dogdog.com	PR 7
http://www.goddog.com	PR 6
http://www.dog.com	PR 3
http://www.dogyear.com	PR 5
http://www.baddog.com	PR 2

STEP 4: RETURN SEARCH RESULTS PAGE



## Keyword Domination

Keywords dominate our search intent and interaction with the engines. For example, a common search query pattern might go something like this:

When a search is performed, the engine matches pages to retrieve based on the words entered into the search box. Other data, such as the order of the words ("tanks shooting" vs. "shooting tanks"), spelling, punctuation, and capitalization of those keywords provide additional information that the engines use to help retrieve the right pages and rank them.

To help accomplish this, search engines measure the ways keywords are used on pages to help determine the "relevance" of a particular document to a query. One of the best ways to "optimize" a page's rankings is to ensure that keywords are prominently used in titles, text, and meta data.

Generally, the more specific your keywords, the better your chances of ranking based on less competition. The map graphic to the left shows the relevance of the broad term *books* to the specific title, *Tale of Two Cities*. Notice that while there are a lot of results (size of country) for the broad term, there are a lot less results and thus competition for the specific result.

# Keyword Abuse

Since the dawn of online search, folks have abused keywords in a misguided effort to manipulate the engines. This involves "stuffing" keywords into text, the url, meta tags and links. Unfortunately, this tactic almost always does more harm to your site.

In the early days, search engines relied on keyword usage as a prime relevancy signal, regardless of how the keywords were actually used. Today, although search engines still can't read and comprehend text as well as a human, the use of machine learning has allowed them to get closer to this ideal.

The best practice is to use your keywords naturally and strategically (more on this below.) If your page targets the keyword phrase "Eiffel Tower" then you might naturally include content about the Eiffel Tower itself, the history of the tower, or even recommended Paris hotels. On the other hand, if you simply sprinkle the words "Eiffel Tower" onto a page with irrelevant content, such as a page about dog breeding, then your efforts to rank for "Eiffel Tower" will be a long, uphill battle.

## On-Page Optimization

That said, keyword usage and targeting are still a part of the search engines' ranking algorithms, and we can leverage some effective "best practices" for keyword usage to help create pages that are close to "optimized." Here at SEOmoz, we engage in a lot of testing and get to see a huge number of search results and [shifts based on keyword usage tactics](#). When working with one of your own sites, this is the process we recommend:

- \* Use the keyword in the title tag at least once. Try to keep the keyword as close to the beginning of the title tag as possible. More detail on title tags follows later in this section.
- \* Once prominently near the top of the page.
- \* At least 2-3 times, including variations, in the body copy on the page - sometimes a few more if there's a lot of text content. You may find additional value in using the keyword or variations more than this, but in our experience, adding more instances of a term or phrase tends to have little to no impact on rankings.
- \* At least once in the alt attribute of an image on the page. This not only helps with web search, but also image search, which can occasionally bring valuable traffic.
- \* Once in the URL. Additional rules for URLs and keywords are discussed later on in this section.
- \* At least once in the meta description tag. Note that the meta description tag does NOT get used by the engines for rankings, but rather helps to attract clicks by searchers from the results page, as it is the "snippet" of text used by the search engines.
- \* Generally not in link anchor text on the page itself that points to other pages on your site or different domains (this is a bit complex - see this [blog post](#) for details).

## Keyword Density Myth

Keyword density is not a part of modern ranking algorithms, as demonstrated in Dr. Edel Garcia's [The Keyword Density of Non-Sense](#).

If two documents,  $D_1$  and  $D_2$ , consist of 1000 terms ( $l = 1000$ ) and repeat a term 20 times ( $tf = 20$ ), then a keyword density analyzer will tell you that for both documents Keyword Density (KD)  $KD = 20/1000 = 0.020$  (or 2%) for that term. Identical values are obtained when  $tf = 10$  and  $l = 500$ . Evidently, a keyword density analyzer does not establish which document is more relevant. A density analysis or keyword density ratio tells us nothing about:

1. The relative distance between keywords in documents (proximity)
2. Where in a document the terms occur (distribution)
3. The co-citation frequency between terms (co-occurrence)
4. The main theme, topic, and sub-topics (on-topic issues) of the documents

The Conclusion:

Keyword density is divorced from content, quality, semantics, and relevancy.

What should optimal page density look like then? An optimal page for the phrase "running shoes" would thus look something like:

### PAGE TARGETING THE PHRASE "RUNNING SHOES"

---

**TITLE:** *Running Shoes* for Runners Who Love High Quality, Comfortable Shoes

---

**H1 HEADER TAG:** Find the Best **Running Shoes** to Fit Your Needs

---



**BODY COPY:** Some compelling text about the benefits of choosing the right **running shoes**, finding the brands and sizes that fit your feet and how to compare **running shoes** in an easy-to-follow, logical fashion. This would also be a good place to describe how different **running shoes** apply to different activities like trail running, long distance running, sprinting, etc.

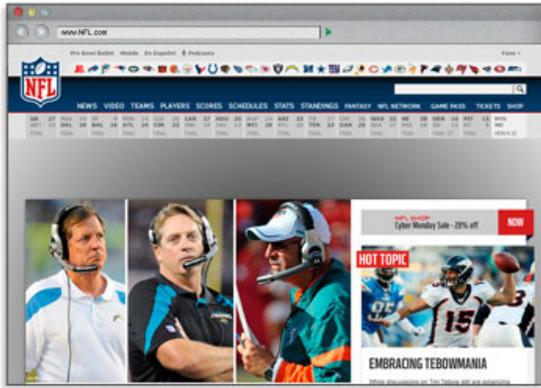
---

**ALT ATTRIBUTE FOR PHOTO:** The Famous **Running Shoes** of an Olympian

---

**URL:** <http://www.yourdomain.com/Running Shoes>

You can read more information about On-Page Optimization at [this post](#).



# Title Tags

The **title element of a page** is meant to be an accurate, concise description of a page's content. It is critical to both user experience and search engine optimization.

As title tags are such an important part of search engine optimization, the following best practices for title tag creation makes for terrific low-hanging SEO fruit. The recommendations below cover the critical parts of optimizing title tags for search engine and usability goals.

## Be mindful of length

Search engines display only the first 65-75 characters of a title tag in the search results. (After this length, the engines show an ellipsis - "...") to indicate when a title tag has been cut off) This is also the general limit allowed by most social media sites, so sticking to this limit is generally wise. However, if you're targeting multiple keywords (or an especially long keyword phrase) and having them in the title tag is essential to ranking, it may be advisable to go longer.

## Place important keywords close to the front

The closer to the start of the title tag your keywords are, the more helpful they'll be for ranking and the more likely a user will be to click them in the search results.

## Leverage branding

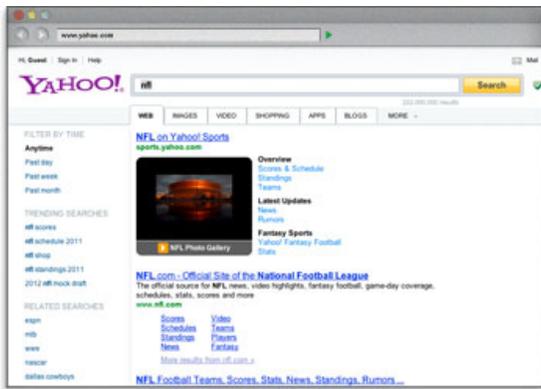
At SEOmoz, we love to end every title tag with a brand name mention, as these help to increase brand awareness, and create a higher click-through rate for people who like and are familiar with a brand. Sometimes it makes sense to place your brand at the beginning of the title tag, such as your homepage. Since words at the beginning of the title tag carry more weight, be mindful of what you are trying to rank for.

## Consider readability and emotional impact

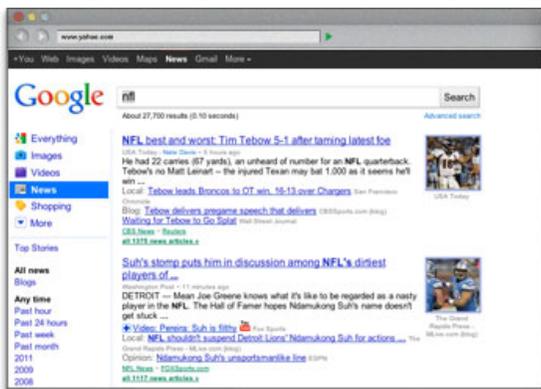
Title tags should be descriptive and readable. Creating a compelling title tag will pull in more visits from the search results and can help to invest visitors in your site. Thus, it's important to not only think about optimization and keyword usage, but the entire user experience. The title tag is a new visitor's first interaction with your brand and should convey the most positive impression possible.

## Best Practices for Title Tags

The title tag of any page appears at the top of Internet browsing software, and is often used as the title when your content is shared through social media or republished.



Using keywords in the title tag means that search engines will "bold" those terms in the search results when a user has performed a query with those terms. This helps garner a greater visibility and a higher click-through rate.



The final important reason to create descriptive, keyword-laden title tags is for ranking at the search engines. In [SEOmoz's biannual survey of SEO industry leaders](#), 94% of participants said that keyword use in the title tag was the most important place to use keywords to achieve high rankings.

# Meta Tags

Meta tags were originally intended to provide a proxy for information about a website's content. Several of the basic meta tags are listed below, along with a description of their use.

## Meta Robots

The [Meta Robots tag](#) can be used to control search engine spider activity (for all of the major engines) on a page level. There are several ways to use meta robots to control how search engines treat a page:

- \* **index/noindex** tells the engines whether the page should be crawled and kept in the engines' index for retrieval. If you opt to use "noindex", the page will be excluded from the engines. By default, search engines assume they can index all pages, so using the "index" value is generally unnecessary.
- \* **follow/nofollow** tells the engines whether links on the page should be crawled. If you elect to employ "nofollow," the engines will disregard the links on the page both for discovery and ranking purposes. By default, all pages are assumed to have the "follow" attribute.  
Example: `<META NAME="ROBOTS" CONTENT="NOINDEX, NOFOLLOW">`
- \* **noarchive** is used to restrict search engines from saving a cached copy of the page. By default, the engines will maintain visible copies of all pages they indexed, accessible to searchers through the "cached" link in the search results.
- \* **nosnippet** informs the engines that they should refrain from displaying a descriptive block of text next to the page's title and URL in the search results.
- \* **noodp/noydir** are specialized tags telling the engines not to grab a descriptive snippet about a page from the Open Directory Project (DMOZ) or the Yahoo! Directory for display in the search results.

The [X-Robots-Tag](#) HTTP header directive also accomplishes these same objectives. This technique works especially well for content within non-HTML files, like images.

## Meta Description

The [meta description](#) tag exists as a short description of a page's content. Search engines do not use the keywords or phrases in this tag for rankings, but meta descriptions are the primary source for the snippet of text displayed beneath a listing in the results.

The meta description tag serves the function of advertising copy, drawing readers to your site from the results and thus, is an extremely important part of search marketing. Crafting a readable, compelling description using important keywords (notice how Google "bolds" the searched keywords in the description) can draw a much higher click-through rate of searchers to your page.

Meta descriptions can be any length, but search engines generally will cut snippets longer than 160 characters, so it's generally wise to stay in these limits.

In the absence of meta descriptions, search engines will create the search snippet from other elements of the page. For pages that target multiple keywords and topics, this is a perfectly valid tactic.

## Not as Important Meta Tags

### Meta Keywords

The meta keywords tag had value at one time, but is no longer valuable or important to search engine optimization. For more on the history and a full account of why meta keywords has fallen into disuse, read [Meta Keywords Tag 101](#) from SearchEngineLand.

### Meta refresh, meta revisit-after, meta content type, etc.



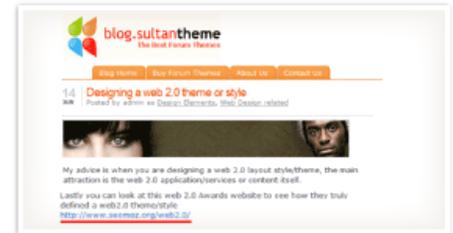
META DESCRIPTION TAG

Although these tags can have uses for search engine optimization, they are less critical to the process, and so we'll leave it to Google's Webmaster Tools Help to answer in greater detail - [Meta Tags](#).

---

## URL Structures

URLs, the web address for a particular document, are of great value from a search perspective. They appear in multiple important locations.



Since search engines display URLs in the results, they can impact click-through and visibility. URLs are also used in ranking documents, and those pages whose names include the queried search terms receive some benefit from proper, descriptive use of keywords.

URLs make an appearance in the web browser's address bar, and while this generally has little impact on search engines, poor URL structure and design can result in negative user experiences.

The URL above is used as the link anchor text pointing to the referenced page in this blog post.

---

## URL Construction Guidelines

### *Employ Empathy*

Place yourself in the mind of a user and look at your URL. If you can easily and accurately predict the content you'd expect to find on the page, your URLs are appropriately descriptive. You don't need to spell out every last detail in the URL, but a rough idea is a good starting point.

### *Shorter is better*

While a descriptive URL is important, minimizing length and trailing slashes will make your URLs easier to copy and paste (into emails, blog posts, text messages, etc) and will be fully visible in the search results.

### *Keyword use is important (but overuse is dangerous)*

If your page is targeting a specific term or phrase, make sure to include it in the URL. However, don't go overboard by trying to stuff in multiple keywords for SEO purposes - overuse will result in less usable URLs and can trip spam filters.

### *Go static*

The best URLs are human readable without lots of parameters, numbers and symbols. Using technologies like mod\_rewrite for Apache and ISAPI\_rewrite for Microsoft, you can easily transform



dynamic URLs like this [www.seomoz.org/blog?id=123](http://www.seomoz.org/blog?id=123) into a more readable static version like this:

<http://www.seomoz.org/blog/google-fresh-factor>. Even single dynamic parameters in a URL can result in lower overall ranking and indexing.

### *Use hyphens to separate words*

Not all web applications accurately interpret separators like underscore "\_," plus "+," or space "%20," so use the hyphen "-" character to separate words in a URL, as in [google-fresh-factor](http://www.seomoz.org/blog/google-fresh-factor) for URLs example above.

---

## Canonical and Duplicate Versions of Content

**Duplicate content** is one of the most vexing and troublesome problems any website can face. Over the past few years, search engines have cracked down on "thin" and duplicate content through penalties and lower rankings.

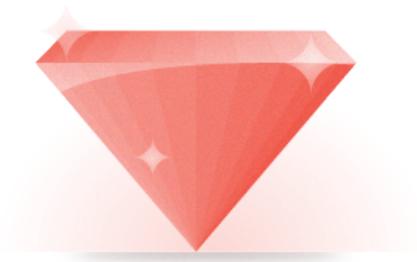
**Canonicalization** happens when two or more duplicate versions of a webpage appear on different URLs. This is very common with modern Content Management Systems. For example, you offer a regular version of a page and a "print optimized" version of the same content. Duplicate content can even appear on multiple websites. For search engines, this presents a big problem - which version of this content should they show to searchers? In SEO circles, this issue is often referred to as **duplicate content** - described in greater detail [here](#).



*The engines are picky about duplicate versions of a single piece of material. To provide the best searcher experience, they will rarely show multiple, duplicate pieces of content and thus, are forced to choose which version is most likely to be the original. The end result is ALL of your duplicate content could rank lower than it should.*

*Canonicalization is the practice of organizing your content in such a way that **every unique piece has one and only one URL**. If you leave multiple versions of content on a website (or websites), you might end up with a scenario like that to the right. Which diamond is the right one?*





Instead, if the site owner took those three pages and **301-redirected** them, the search engines would have only one, **stronger** page to show in the listings from that site.

***When multiple pages with the potential to rank well are combined into a single page, they not only no longer compete with one another, but create a stronger relevancy and popularity single overall. This will positively impact their ability to rank well in the search engines.***

### ***The Canonical Tag to the Rescue!***

A different option from the search engines, called the "**Canonical URL Tag**" is another way to reduce instances of duplicate content on a single site and canonicalize to an individual URL. This can also be used **across different websites**, from one URL on one domain to a different URL on a different domain.

Use the canonical tag within the page that contains duplicate content. The "target" of the canonical tag points to the "master" URL that you want to rank for.

#### « THE INNER WORKINGS »

```
<link rel="canonical" href="http://www.seomoz.org/blog"/>
```

*This tells search engines that the page in question should be treated as though it were a copy of the URL [www.seomoz.org/blog](http://www.seomoz.org/blog) and that all of the link & content metrics the engines apply should flow back to that URL.*



The Canonical URL tag attribute is similar in many ways to a 301 redirect from an SEO perspective. In essence, you're telling the engines that multiple pages should be considered as one (which a 301 does), without actually redirecting visitors to the new URL - often saving your development staff considerable heartache.

For more about different types of duplicate content, [this post by Dr. Pete](#) deserves special mention.

# Easy as pie!



## Rich Snippets

Ever see a 5 star rating in a search result? Chances are, the search engine received that information from rich snippets embedded on the webpage. Rich snippets are a type of structured data that allow webmasters to mark up content in ways that provide information to the search engines.

While the use of rich snippets and structured data is not a required element of search engine friendly design, it's growing adoption means that webmasters who take advantage may enjoy an advantage in some circumstances.

Structured data means adding markup to your content so that search engines can easily identify what type of content it is. Schema.org provides several types of examples of data that can benefit from structured markup. These include people, products, reviews, businesses, recipes and events.

Often the search engines include structured data in search results, such as in the case of user reviews (stars) and author profiles (pictures.) There are several good resources for learning more about rich snippets online, including [information at Schema.org](#) and Google's [Rich Snippet Testing Tool](#).

## Rich Snippets in the Wild

Let's say you announce an SEO Conference on your blog. In regular HTML, your code might look like this:

```
<div>
SEO Conference<br/>
Learn about SEO from experts in the field.<br/>
Event date:<br/>
May 8, 7:30pm
</div>
```

Now, by structuring the data, we can tell the search engines more specific information about the type of data. The end result might look like this:

```
<div itemscope
itemtype="http://schema.org/Event">
<div itemprop="name">SEO Conference</div>
<span itemprop="description">Learn about SEO
from experts in the field.</span>
Event date:
<time itemprop="startDate" datetime="2012-05-
08T19:30">May 8, 7:30pm</time>
</div>
```

## Defending Your Sites Honor

### *How scrapers steal your rankings*

Unfortunately, the web is filled with hundreds of thousands (if not millions) of unscrupulous websites whose business and traffic models depend on plucking the content off other sites and re-using them (sometimes in strangely modified ways) on their own domains. This practice of fetching your content and re-publishing is called "scraping," and the scrapers make remarkably good earnings by outranking sites for their own content and displaying ads (ironically, often Google's own AdSense program).

When you publish content in any type of feed format - RSS/XML/etc - make sure to ping the major blogging/tracking services (like Google, Technorati, Yahoo!, etc.). You can find instructions for how to ping services like Google and Technorati directly from their sites, or use a

service like [Pingomatic](#) to automate the process. If your publishing software is custom-built, it's typically wise for the developer(s) to include auto-pinging upon publishing.

Next, you can use the scrapers' laziness against them. Most of the scrapers on the web will re-publish content without editing, and thus, by including links back to your site, and the specific post you've authored, you can ensure that the search engines see most of the copies linking back to you (indicating that your source is probably the originator). To do this, you'll need to use absolute, rather than relative links in your internal linking structure. Thus, rather than linking to your home page using:

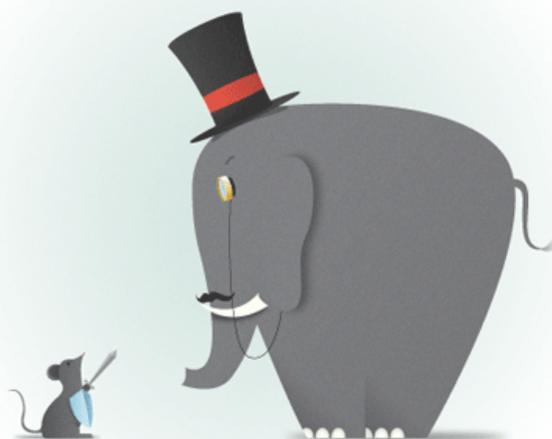
```
<a href="."/>Home</a>
```

You would instead use:

```
<a href="http://www.seomoz.org">Home</a>
```

This way, when a scraper picks up and copies the content, the link remains pointing to your site.

There are more advanced ways to protect against scraping, but none of them are entirely foolproof. You should expect that the more popular and visible your site gets, the more often you'll find your content scraped and re-published. Many times, you can ignore this problem, but if it gets very severe, and you find the scrapers taking away your rankings and traffic, you may consider using a legal process called a DMCA takedown. Luckily, SEOMoz's own in-house counsel, Sarah Bird, has authored a brilliant piece to help solve just this problem - [Four Ways to Enforce Your Copyright: What to Do When Your Online Content is Being Stolen](#).



# CHAPTER FIVE

# KEYWORD RESEARCH

It all begins with words typed into a search box.

Keyword research is one of the most important, valuable, and high return activities in the search marketing field. Ranking for the "right" keywords can make or break your website. Through the detective work of puzzling out your market's keyword demand, you not only learn which terms and phrases to target with SEO, but also learn more about your customers as a whole.

It's not always about getting visitors to your site, **but about getting the right kind of visitors.** The usefulness of this intelligence cannot be overstated - with keyword research you can predict shifts in demand, respond to changing market conditions, and produce the products, services, and content that web searchers are already actively seeking. In the history of marketing, there has never been such a low barrier to entry in understanding the motivations of consumers in virtually every niche.

## How to Judge the Value of a Keyword

How much is a keyword worth to your website? If you own an online shoe store, do you make more sales from visitors searching for "brown shoes" or "black boots?" The keywords visitors type into search engines are often available to webmasters, and keyword research tools allow us to find this information. However, those tools cannot show us directly how valuable it is to receive traffic from those searches. To understand the value of a keyword, we need to understand our own websites, make some hypotheses, test, and repeat - the classic web marketing formula.

A basic process for assessing a keyword's value:

### Ask yourself...

*Is the keyword relevant to your website's content? Will searchers find what they are looking on your site when they search using these keywords? Will they be happy with what*



they find? Will this traffic result in financial rewards or other organizational goals? If the answer to all of these questions is a clear "Yes!", proceed...

### Search for the term/phrase in the major engines

Understanding which websites already rank for your keyword gives you valuable insight into the competition, and also how hard it will be to rank for the given term. Are there search advertisements running along the top and right-hand side of the organic results? Typically, many search ads means a high value keyword, and multiple search ads above the organic results often means a highly lucrative and directly conversion-prone keyword.

### Buy a sample campaign for the keyword at Google AdWords and/or Bing Adcenter

If your website doesn't rank for the keyword, you can nonetheless buy "test" traffic to see how well it converts. In [Google Adwords](#), choose "exact match" and point the traffic to the relevant page on your website. Track impressions and conversion rate over the course of at least 2-300 clicks.

### Using the data you've collected, determine the exact value of each keyword.

For example, if your search ad generated 5,000 impressions, of which 100 visitors have come to your site and 3 have converted for total profit (not revenue!) of \$300, then a single visitor for that keyword is worth \$3 to your business. Those 5,000 impressions in 24 hours could generate a click-through rate of between 18-36% with a #1 ranking (see the [Slingshot SEO study](#) for more on potential click-through rates), which would mean 900-1800 visits per day, at \$3 each, or between **1-2 million dollars per year**. No wonder businesses love search marketing!



Even the best estimates of value fall flat against the hands-on process of optimizing and calculating ROI. Search Engine Optimization involves constant testing, experimenting and improvement. Remember, even though SEO is typically one of the highest return marketing investments, measuring success is still critical to the process.

---

## Understanding the Long Tail of Keyword Demand

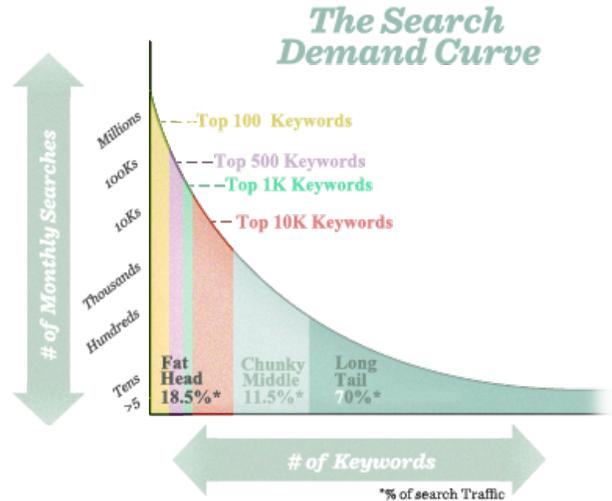
Going back to our online shoe store example, it would be great to rank #1 for the keyword "shoes" - or would it?

It's wonderful to deal with keywords that have 5,000 searches a day, or even 500 searches a day, but in reality, these "popular" search terms actually make up less than 30% of the searches performed on the web. The remaining 70% lie in what's called the "long tail" of search. The long tail contains hundreds of millions of unique searches that might be conducted a few times in any given day, but, when taken together, they comprise the majority of the world's demand for information through search engines.

Another lesson search marketers have learned is that long tail keywords often convert better, because they catch people later in the buying/conversion cycle. A person searching for "shoes" is probably browsing, and not ready to buy. On the other hand, someone searching for "best price on Air Jordan size 12" practically has their

wallet out!

Understanding the search demand curve is critical. To the right we've included a sample keyword demand curve, illustrating the small number of queries sending larger amounts of traffic alongside the volume of less-searched terms and phrases that bring the bulk of our search referrals.



***“Ignore the long tail at your peril - search marketing and web site content strategies must allow for this “impossible to predict” form of visits or risk losing out to a more expository and prolific competitor.”***

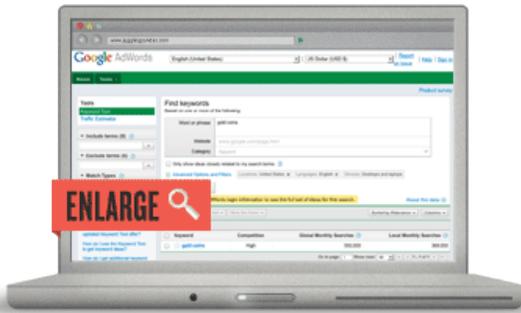


# Keyword Research

## Resources

Where do we get all of this knowledge about keyword demand and keyword referrals? From research sources like these listed here:

- \* [Google Adwords' Keyword Tool](#)
- \* [Google Insights for Search](#)
- \* [Google Trends Keyword Demand Prediction](#)
- \* [Microsoft Advertising Intelligence](#)
- \* [Wordtracker's Free Basic Keyword Demand](#)



*Google's AdWords Keyword Tool provides suggested keyword and volume data.*

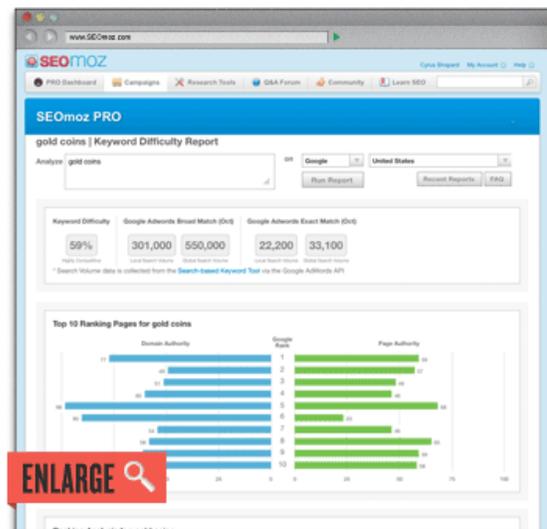
Google's AdWords Keyword tool is a common starting point for SEO keyword research. It not only suggests keywords and provides estimated search volume, but also predicts the cost of running paid campaigns for these terms. To determine volume for a particular keyword, be sure to set the Match Type to [Exact] and look under Local Monthly Searches. Remember that these represent total searches. Depending on your ranking and click-through rate, the actual number of visitors you achieve for these keywords will usually be much lower.

Other sources for keyword information exist, as do tools with more advanced data. The SEOmoz blog category on [Keyword Research](#) is an excellent place to start.

## Keyword Difficulty

### *What are my chances of success?*

In order to know which keywords to target, it's essential to not only understand the **demand** for a given term or phrase, but also the **work required** to achieve those rankings. If big brands take the top 10 results and you're just starting out on the web, the uphill battle for rankings can take years of effort. This is why it's essential to understand keyword difficulty.



*Different tools around the web help provide this information. One of these, SEOmoz's own [Keyword Analysis Tool](#) does a good job collecting all of these metrics and providing a comparative score for any given search term or phrase.*

## CHAPTER SIX

# HOW USABILITY, USER EXPERIENCE & CONTENT AFFECT SEARCH ENGINE RANKINGS

The search engines constantly strive to improve their performance by providing the best possible results. While "best" is subjective, the engines have a very good idea of the kinds of pages and sites that satisfy their searchers. Generally, these sites have several traits in common:

- 1 *Easy to use, navigate, and understand*
- 2 *Provide direct, actionable information relevant to the query*
- 3 *Professionally designed and accessible to modern browsers*
- 4 *Deliver high quality, legitimate, credible content*

Despite amazing technological advances, search engines can't yet understand text, view images, or watch video the same way a human can. Thus, in order to understand content they rely on meta information (not necessarily meta tags) about sites and pages in order to rank content. Web pages do not exist in a vacuum - real human beings interact with them. Search engines use data to "observe" how people engage with web pages, and this gives them incredible insight as to the quality of the pages themselves.

## The Impact of Usability and User Experience

### *On Search Engine Rankings*

There are a limited number of variables that search engines can take into account directly, including keywords, links, and site structure. However, through linking patterns, user engagement metrics and machine learning, the engines make a considerable number of intuitions about a given site. Usability and user experience are "second order" influences on search engine ranking success. They provide an indirect, but measurable benefit to a site's external popularity, which the engines can then interpret as a signal of higher quality. This is called the "**no one likes to link to a crummy site**" phenomenon.



*“Yeah, I dabble...”*

*“Dooood, I just got’s me some  
Adobe CS4 with extra filters!”*



Crafting a thoughtful, empathetic user experience can ensure that your site is perceived positively by those who visit, encouraging sharing, bookmarking, return visits and links - signals that trickle down to the search engines and contribute to high rankings.

---

## Signals of Quality Content

### **1. Engagement Metrics**

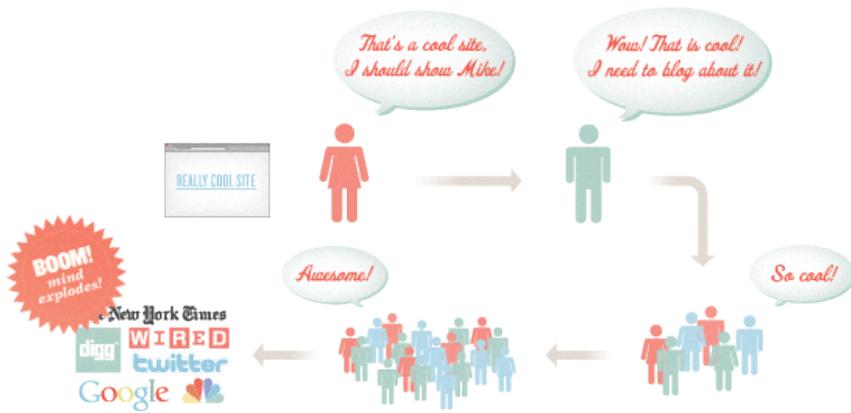
When a search engine delivers a page of results to you, they can measure their success by observing how you engage with those results. If you hit the first link, then immediately hit the "back" button to try the second link, this indicates that you were not satisfied with the first result. Since the beginning, search engines have sought the "long click" - where users click a result without immediately returning to the search page to try again. Taken in aggregate over millions and millions of queries a day, the engines build up a good pool of data to judge the quality of their results.

### **2. Machine Learning**

In 2011 Google introduced the [Panda Update](#) to its ranking algorithm, significantly changing the way it judged websites for quality. Google started by using human evaluators to manually rate 1000s of sites, searching for "low quality" content. Google then incorporated machine learning to mimic the human evaluators. Once its computers could accurately predict what the humans would judge a low quality site, the algorithm was introduced across millions of sites spanning the Internet. The end result was a seismic shift which rearranged over 20% of all of Google's search results. For more on the Panda update, some good resources can be found [here](#) and [here](#).

### **3. Linking Patterns**

The engines discovered early on that the link structure of the web could serve as a proxy for votes and popularity - higher quality sites and information earned more links than their less useful, lower quality peers. Today, link analysis algorithms have advanced considerably, but these principles hold true.



All of that positive attention and excitement around the content offered by the new site translates into a machine parse-able (and algorithmically valuable) collection of links. The timing, source, anchor text, and number of links to the new site are all factored into its potential performance (i.e., ranking) for relevant queries at the engines.

Now imagine that site wasn't so great - let's say it's just an ordinary site without anything unique or impressive.

## Crafting Content

### For Search Engine Success

Developing "great content" may be the most repeated suggestion in the SEO world. Yet, despite its clichéd status, appealing, useful content is critical to search engine optimization. Every search performed at the engines comes with an intent - to find, learn, solve, buy, fix, treat, or understand. Search engines place web pages in their results in order to satisfy that intent in the best possible way, and crafting the most fulfilling, thorough content that addresses a searcher's needs provides an excellent chance to earn top rankings.

## Search Intent Flavors

Search intent comes in a variety of flavors...



### Transactional Searches

Identifying a local business, making a purchase online and completing a task.

Transactional searches don't necessarily involve a credit card or wire transfer. Signing up for a free trial account at Cook's Illustrated, creating a Gmail account, or finding the best local Mexican cuisine (in Seattle it's Carta de Oaxaca) are all transactional queries.

### **Navigational Searches**

Visiting a pre-determined destination and sourcing the “correct” website URL.

Navigational searches are performed with the intent of surfing directly to a specific website. In some cases, the user may not know the exact URL, and the search engine serves as the “White Pages”, passing along the (hopefully) correct location.

*I want to check in for my Alaska Airlines flight:*

**YAHOO! SEARCH**

Web Images Video Local Shopping More

ALASKAN AIRLINES



*What was the name of that actor from “Firefly?”*

**Ask**.com

Web Images Video Local Shopping More

ACTOR WHO PLAYED WASH FROM FIREFLY



### **Informational Searches**

Researching non-transactional information, getting quick answers and ego-searching.

Informational searches involve a huge range of queries from finding out the local weather, getting a map and directions, to finding the name of Tony Starks' military buddy from the Iron Man movie or checking on just how long that trip to Mars really takes. The common thread here is that the searches are primarily non-commercial and non-transaction-oriented in nature; the information itself is the goal, and no interaction beyond clicking and reading is required.

---

Fulfilling these intents is up to you - Creativity, high quality writing, use of examples, images, and multimedia all help in crafting content that perfectly fits with a searcher's goals. Your reward is satisfied searchers who demonstrate positive experience through engaged activity on your site or with links to it.



# CHAPTER SEVEN

## GROWING POPULARITY & LINKS

For search engines that crawl the web, links are the streets between pages. Using sophisticated link analysis, the engines can discover how pages are related to each other and in what ways.

Since the late 1990's search engines have used links as votes - representing the democracy of the web's opinion about what pages are important and popular. The engines themselves have refined the use of link data to a fine art, and complex algorithms create nuance evaluations of sites and pages based on this information.

Links aren't everything in SEO, but search professionals attribute a large portion of the engines' algorithms to link-based factors (see [Search Engine Ranking Factors](#)). Through links, engines can not only analyze the popularity of a website & page based on the number and popularity of pages linking to them, but also metrics like trust, spam, and authority. Trustworthy sites tend to link to other trusted sites, while spammy sites receive very few links from trusted sources (see [mozTrust](#)). Authority models, like those postulated in the [Hilltop Algorithm](#), suggest that links are a very good way of identifying expert documents on a given subject.



*Thanks to this focus on algorithmic use and analysis of links, growing the link profile of a website is critical to gaining traction, attention, and traffic from the engines. As an SEO, link building is among the top tasks required for search ranking and traffic success.*

---

## Link Signals

### *used by search engines*

Before embarking on a link building effort, it's critical to understand the [elements of a link used by the search engines](#) as well as how those elements factor into the weighting of links in the algorithms. Search engines use links in many different ways. While we don't know all the link attributes measured by the engines, through analysis of patent applications, years of experience and hands-on testing, we can draw some intelligent assumptions that hold up in the real world. Below is a list of notable factors worthy of consideration. These signals, and many more, are considered by professional SEOs when measuring link value and a site's link profile.

### *Global Popularity*



The more popular and important a site is, the more links from that site matter. A site like [Wikipedia](#) has literally 1000's of diverse sites linking to it, which means it's probably a popular and important site. To earn trust and authority with the engines, you'll need the help of other link partners. The more popular, the better.

### Local/Topic-Specific Popularity

The concept of "local" popularity, first pioneered by the Teoma search engine, suggests that links from sites within a topic-specific community matter more than links from general or off-topic sites. For example, if your website sells dog houses, earning links from the Society of Dog Breeders matters much more than earning links from an off-topic, roller skating site.

### Anchor Text

One of the strongest signals the engines use in rankings is [anchor text](#). If dozens of links point to a page with the right keywords, that page has a very good probability of ranking well for the targeted phrase in that anchor text. You can see examples of this in action with searches like "[click here](#)", where many results rank solely due to the anchor text of inbound links.

### TrustRank

It's no surprise that the Internet contains massive amounts of spam. Some estimate as much as 60% of the web's pages are spam. In order to weed out this irrelevant content, search engines use systems for measuring trust, many of which are based on the link graph. Earning links from highly trusted domains can result in a significant boost to this scoring metric. Universities, government websites and non-profit organizations represent examples of high-trust domains.

### Link Neighborhood

Spam links often go both ways. A website that links to spam is likely spam itself, and in turn often has many spam sites linking back to it. By looking at the totality of these links in aggregate, search engines can understand the "link neighborhood" your website exists in. Thus, it's wise to choose those sites you link to carefully and be equally selective with the sites you attempt to earn links from.

### Freshness

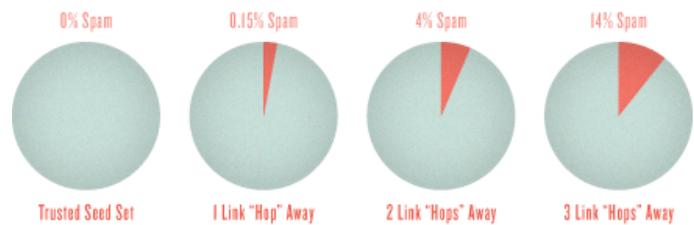
Link signals tend to decay over time. Sites that were once popular often go stale, and eventually fail to earn new links. Thus, it's important not only to earn links to your website, but also to continue to earn additional links over time. Commonly referred to as "[FreshRank](#)," search engines use the freshness signals of links to judge current popularity and relevance.

### Social Sharing

The last few years has seen an explosion in the amount of content shared through social services such as Facebook, Twitter and Google+. Although search engines treat socially shared links differently than other types of links, they notice them nonetheless. There is much debate among search professionals as to how exactly search engines factor social link signals into their algorithm, but there is no denying the rising importance of social channels.



### THE CONCEPT OF TRUSTRANK



*Search researchers found that the further a site/page's distance from a trusted seed set, the more likely they were to contain spam results.*

# The Power of Social Sharing

## **How Google+, Twitter and Facebook Change the Game**

The years 2011=2012 saw a huge rise in social sharing and its effects on search. Google, in particular, began to incorporate a huge number of social signals into its search results. This involves serving personalized results to logged-in users that include content shared by the searcher's social circle (Facebook, Twitter and others). Normally, these results might not appear in the top ten, but are promoted because of this social influence.

The potential power of this shift towards social for search marketers is huge. Those with large social circles, who share a lot of material, are more likely to see that material (and their face) promoted in search results. For publishers, this means it's beneficial to have your content shared by these same highly influential folks with large social followings. For Google in particular, this is especially true of content shared on Google+.

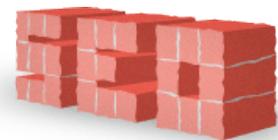
## **Are Social Shares the Same as Links**

In a word, no. Although [there is evidence](#) that social shares such as Tweets, Likes, and Plusses effect rankings, at this time links are considered a far superior and more lasting way to promote the popularity of your content than any other method.

---

## Link Building Basics

Link building is an art. It's almost always the most challenging part of an SEO's job, but also the one most critical to success. Link building requires creativity, hustle, and often, a budget. No two link building campaigns are the same, and the way you choose to build links depends as much upon your website as it does your personality. Below are three basic types of link acquisition.



### **1 "Natural" Editorial Links**

Links that are given naturally by sites and pages that want to link to your content or company. These links require no specific action from the SEO, other than the creation of worthy material (great content) and the ability to create awareness about it.

### **2 Manual "Outreach" Link Building**

The SEO creates these links by emailing bloggers for links, submitting sites to directories, or paying for listings of any kind. The SEO often creates a value proposition by explaining to the link target why creating the link is in their best interest. Examples include filling out forms for submissions to a website award program or convincing a professor that your resource is worthy of inclusion on the public syllabus.

### **3 Self-Created, Non-Editorial**

Hundreds of thousands of websites offer any visitor the opportunity to create links through guest book signings, forum signatures, blog comments, or user profiles. These links offer the lowest value, but can, in aggregate, still have an impact for some sites. In general, search engines continue to devalue most of these types of links, and have been known to penalize sites that pursue these links aggressively. Today, these types of links are often considered spammy and should be pursued with caution.

*It's up to you, as an SEO, to select which of these will have the highest return on the effort invested. As a general rule, it's wise to build as vast and varied a link profile as possible, as this brings the best search engine results. Any link building pattern that appears non-standard, unnatural, or manipulative will eventually become a target for advancing search algorithms to discount.*

# STARTING a Link Building Campaign



As with any marketing activity, the first step in any link building campaign is the creation of goals and strategies. Unfortunately, link building is one of the most difficult activities to measure. Although the engines internally weigh each link with precise, mathematical metrics, it's impossible for those on the outside to know this data.

SEOs rely on a number of signals to help build a rating scale of link value. Along with the data from the link signals mentioned above, these metrics include the following:

### **Ranking for Relevant Search Terms**

One of the best ways to determine how well a search engine values a given page is to search for some of the keywords and phrases that page targets (particularly those in the [title tag](#) and headline). For example, if you are trying to rank for the phrase "dog kennel", earning links from pages that already rank for this phrase would help significantly.

### **SEOmox mozRank**

[mozRank](#) (mR) shows how popular a given web page is on the web. Pages with high mozRank (popular) scores tend to rank better. The more links to a given page, the more popular it becomes. Links from important pages (like [www.cnn.com](#) or [www.irs.gov](#)) increase a page's popularity, and subsequently its mozRank, more than unpopular websites.

### **Competitor's Backlinks**

By examining the backlinks of a website that already ranks well for your targeted keyword phrase, you gain valuable intelligence about the links that help them achieve this ranking. Using tools like [Open Site Explorer](#), SEOs can discover these links and target these domains in their own link building campaigns.

### **Number of Links on a Page**

According to the [original PageRank formula](#), the value that a link passes is diluted by the presence of other links on a page. Thus, getting linked-to by a page with few links is better than being linked-to by the same page with many links on it (all other things being equal). The degree to which this is relevant is unknowable (and in our testing, it appears to be important, but not overwhelmingly so), but it's certainly something to be aware of as you conduct link

A web page's mozRank can be improved by getting lots of links from semi-popular pages or a few links from very popular pages.

### **Domain Authority**

SEOMoz [Domain Authority](#) (or DA) is a query independent measure of how likely a domain is to rank for any given query. It is calculated by analyzing the Internet's domain graph and comparing it to tens of thousands of queries in Google.

*It takes time, practice, and experience to build comfort with these variables as they relate to search engine traffic. However, using your website's analytics, you should be able to determine whether your campaign is successful.*

*Success comes when you see increases in search traffic, higher rankings, more frequent search engine crawling, and increases in referring link traffic. If these metrics do not rise after a successful link building campaign, it's possible you either need to seek better quality link targets, or improve your [on-page optimization](#).*

acquisition.

### **Potential Referral Traffic**

Link building should never be solely about search engines. Links that send high amounts of direct click-through traffic not only tend to provide better search engine value for rankings, but also send targeted, valuable visitors to your site (the basic goal of all Internet marketing). This is something you can estimate based on the numbers of visits/page views according to site analytics. If you can't get access to these, services like [Google Trends for Websites](#) can give you a rough idea of at least domain-wide traffic, although these estimates are known to be wildly inaccurate at times.

---

## 5 Samples of Link Building Strategies

### **Get your customers to link to you.**

\* If you have partners you work with regularly or loyal customers that love your brand, you can use this to your advantage by sending out partnership badges - graphic icons that link back to your site (like Google often does with their Adwords certification program). Just as you'd get customers wearing your t-shirts or sporting your bumper stickers, links are the best way to accomplish the same feat on the web. Check out [this post on E-commerce links](#) for more.

### **Build a company blog. Make it a valuable, informative and entertaining resource.**

\* This content and link building strategy is so popular and valuable that it's one of the few recommended personally by the engineers at Google (source: [USA Today](#) & [Stone Temple](#)). Blogs have the unique ability to contribute fresh material on a consistent basis, participate in conversations across the web, and earn listings and links from other blogs, including blogrolls and blog directories.

### **Create content that inspires viral sharing and natural linking**

\* In the SEO world, we often call this "linkbait." Good examples might include David Mihm's [Local Search Ranking Factors](#), [Compare the Meerkat](#), or the funny [How Not To Clean a Window](#). Each leverages aspects of usefulness, information dissemination, or humor to create a viral effect - users who see it once want to share it with friends, and bloggers/tech-savvy webmasters who see it will often do so through links. This high quality, editorially earned votes are invaluable to building trust, authority, and rankings potential

### **Be newsworthy.**

\* Earning the attention of the press, bloggers and news media is an effective, time honored way to earn links. Sometimes this is as simple as [giving away something for free](#), releasing a

*The link building activities you engage in depend largely on the type of site you're working with.*

*For smaller sites, manual link building, including directories, link requests, and link exchanges may be a part of the equation. With larger sites, these tactics tend to fall flat and more scalable solutions are required. Sample strategies are listed here, though this is by no means an exhaustive list (see [SEOMoz's Blog Posts on Link Building](#) for more.)*

*Search for sites like yours by using keywords and phrases directly relevant to your business. When you locate sites that aren't directly competitive, email them, use their online forms, call them on the phone, or even send them a letter by mail to start a conversation about getting a link. Check out [this blog post on link requests](#) for more detail.*

great new product, or stating something controversial.

### **Find directories or listings of relevant resources.**

\* You can find many of these on [SEOMoz's Directory List](#) or use the search engines themselves to find lists of pages that offer outbound links in this fashion. For example, try searching for [allintitle: resources directory](#) at Google and notice the millions of results. Be careful: links that are easy to get often carry risks when pursued aggressively and in volume. A diverse, well rounded link profile is always best.



---

## **Show Me the Money**

### **An Aside on Buying Links**

Google and Bing seek to discount the influence of paid links on their search results. While it is impossible for them to detect and discredit all paid links, the search engines put a lot of time and resources into finding ways to detect these. Websites caught buying links or participating in link schemes risk severe penalties that will drop their rankings into oblivion. Conversely, many search professionals wish the search engines would do more to discourage link buying, which in many cases still works.

As such, we at SEOMoz recommend spending your time on long term link building strategies that focus on building links naturally. You can read more about this in [this blog post](#).





SEOs tend to use a lot of tools. Some of the most useful are provided by the search engines themselves. Search engines want webmasters to create sites and content in accessible ways, so they provide a variety of tools, analytics and guidance. These free resources provide data points and opportunities for exchanging information with the engines that are not provided anywhere else.

Below we explain the common elements that each of the major search engines support and identify why they are useful.

## Common Search Engine Protocols

### 1. Sitemaps

Think of a sitemap as a list of files that give hints to the search engines on how they can crawl your website. Sitemaps help search engines find and classify content on your site that they may not have found on their own. Sitemaps also come in a variety of formats and can highlight many different types of content, including video, images, news and mobile.

You can read the full details of the protocols at [Sitemaps.org](http://Sitemaps.org). In addition, you can build your own sitemaps at [XML-Sitemaps.com](http://XML-Sitemaps.com). Sitemaps come in three varieties:

#### XML

Extensible Markup Language (Recommended Format)

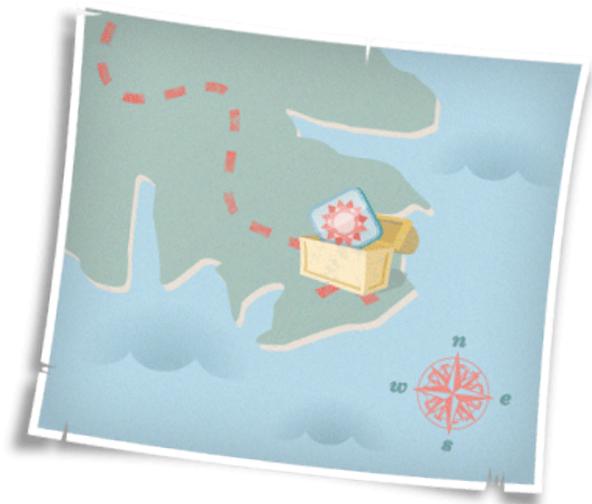
- pro** *This is the most widely accepted format for sitemaps. It is extremely easy for search engines to parse and can be produced by a plethora of sitemap generators. Additionally, it allows for the most granular control of page parameters.*
- con** *Relatively large file sizes. Since XML requires an open tag and a close tag around each element, file sizes can get very large.*

#### RSS

Really Simple Syndication or Rich Site Summary

#### Txt

Text File



**pro** Easy to maintain. RSS sitemaps can easily be coded to automatically update when new content is added.

**con** Harder to manage. Although RSS is a dialect of XML, it is actually much harder to manage due to its updating properties.

**pro** Extremely easy. The text sitemap format is one URL per line up to 50,000 lines.

**con** Does not provide the ability to add meta data to pages.

## 2 Robots.txt

The robots.txt file, a product of the [Robots Exclusion Protocol](#), is a file stored on a website's root directory (e.g., [www.google.com/robots.txt](http://www.google.com/robots.txt)). The robots.txt file gives instructions to automated web crawlers visiting your site, including search spiders.

By using robots.txt, webmasters can indicate to search engines which areas of a site they would like to disallow bots from crawling as well as indicate the locations of sitemap files and crawl-delay parameters. You can read more details about this at the [robots.txt](#) Knowledge Center page.

The following commands are available:

### Disallow

Prevents compliant robots from accessing specific pages or folders.

### Sitemap

Indicates the location of a website's sitemap or sitemaps.

### Crawl Delay

Indicates the speed (in milliseconds) at which a robot can crawl a server.

An Example of Robots.txt

```
#Robots.txt www.example.com/robots.txt
User-agent: *
Disallow:

# Don't allow spambot to crawl any pages
User-agent: spambot
Disallow: /

sitemap:www.example.com/sitemap.xml
```

*Warning: Not all web robots follow robots.txt. People with bad intentions (i.e. e-mail address scrapers) build bots that don't follow this protocol and in extreme cases can use it to identify the location of private information. For this reason, it is recommended that the location of administration sections and other private sections of publicly accessible websites not be included in the robots.txt. Instead, these pages can utilize the meta robots tag (discussed next) to keep the major search engines from indexing their high risk*



content.

### 3. Meta Robots

The meta robots tag creates page-level instructions for search engine bots.

The meta robots tag should be included in the head section of the HTML document.

```
An Example of Meta Robots

<html>
  <head>
<title>The Best Webpage on the Internet</title>
  <meta name="ROBOTS" content="NOINDEX,
    NOFOLLOW">
  </head>
  <body>
<h1>Hello World</h1>
  </body>
</html>
```

In the example above, "NOINDEX, NOFOLLOW" tells robots not to include the given page in their indexes, and also not to follow any of the links on the page.

ROBOTS META TAG		
<code>&lt;meta name = "ROBOT NAME" content = "ARGUMENTS" /&gt;</code>		
ROBOT NAME can be either "robots" for all robots or the user-agent of a specific robot. See robot use-agent list to the left.		
ARGUMENTS		
<b>noindex</b>	Google, Yahoo, Live, Ask	Page not indexed
<b>nofollow</b>	Google, Yahoo, Live, Ask	All links on page become no followed
<b>noarchive</b>	Google, Yahoo, Live, Ask	Page not cached
<b>noodp</b>	Google, Yahoo, Live	Stops description and title tag overwrite by DMOZ (only for Homepage)
<b>noydir</b>	Yahoo	Stops description and title tag overwrite by Yahoo Directory
<b>nosnippet</b>	Google	Stops Google from generating description based on on-page text

### 4. Rel="Nofollow"

Remember how [links act as votes](#)? The rel=nofollow attribute allows you to link to a resource, while removing your "vote" for search engine purposes. Literally, "nofollow" tells search engines not to follow the link, but some engines still follow them for discovering new pages. These links certainly pass less value (and in most cases no juice) than their followed counterparts, but are [useful in various situations](#) where you link to an untrusted source.

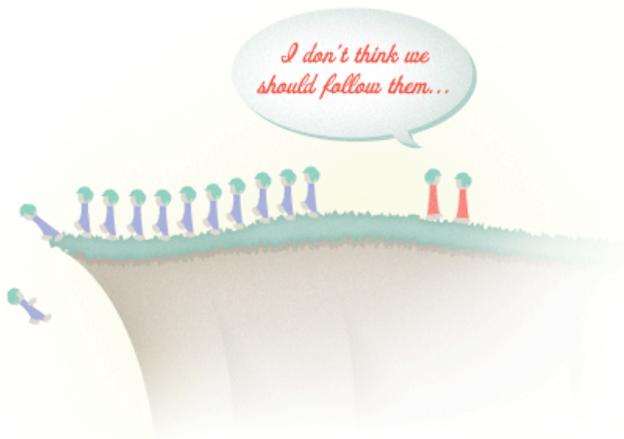
```
An Example of nofollow

<a href="http://www.example.com" title="Example"
  rel="nofollow">Example Link</a>
```

In the example above, the value of the link would not be passed to example.com as the rel=nofollow attribute has been added.

```
An Example of rel="canonical" for the URL
```

### 5. Rel="canonical"



Often, two or more copies of the exact same content appear on your website under different URLs. For example, the following URLs can all refer to a single homepage:

- \* <http://www.example.com/>
- \* <http://www.example.com/default.asp>
- \* <http://example.com/>
- \* <http://example.com/default.asp>
- \* <http://Example.com/Default.asp>

To search engines, these appear as 5 separate pages. Because the content is identical on each page, this can cause the search engines to devalue the content and its potential rankings.

The canonical tag solves this problem by telling search robots which page is the singular "authoritative" version which should count in web results.

```
http://example.com/default.asp

<html>
<head>
<title>The Best Webpage on the Internet</title>
<link rel="canonical" href="http://www.example.com">
</head>
<body>
<h1>Hello World</h1>
</body>
</html>
```

*In the example above, rel=canonical tells robots that this page is a copy of <http://www.example.com>, and should consider the latter URL as the canonical.*

---

## Search Engine Tools

### Google Webmaster Tools - Popular Features

#### Google Webmaster Tools

##### Settings

**Geographic Target** - If a given site targets users in a particular location, webmasters can provide Google with information that will help determine how that site appears in its country-specific search results, and also improve Google search results for geographic queries.

**Preferred Domain** - The preferred domain is the one that a webmaster would like used to index their site's pages. If a webmaster specifies a preferred domain as <http://www.example.com> and Google finds a link to that site that is formatted as <http://example.com>, Google will treat that link as if it were pointing at <http://www.example.com>.

**URL Parameters** - You can indicate to Google information about each parameter on your site, such as "**sort=price**" and "**sessionid=2**". This helps Google crawl your site more efficiently, ignoring those parameters that produce duplicate content and increasing the number of unique pages Google can crawl on your site.

**Crawl Rate** - The crawl rate affects the speed of Googlebot's requests during the crawl process. It has no effect on how often Googlebot crawls a given site. Google determines the recommended rate based on the number of pages on a website.

##### Diagnosics

**Malware** - Google will inform you if it has found any malware on your site. Malware is not only bad for users, but will have a severely negative effect on your rankings.

**Crawl Errors** - If Googlebot encounters significant errors while crawling your site, such as 404s, it will report these and identify



### Your Site on the Web

These statistics offer unique insight to SEOs in particular, as they report keyword impressions, click-through rates, top pages delivered in search results, and linking statistics. Beware, many SEOs complain that the data in Webmaster tools is often incomplete and offers rough estimates at best.

### Site Configuration

This important section allows you to submit sitemaps, test robots.txt files, adjust [sitelinks](#), and submit change of address requests when you move your website from one domain to another. This area also contains the "Settings" and "URL parameters" sections discussed in the previous column.

### +1 Metrics

When users share your content on Google+ with the +1 button, this activity is often annotated in search results. [Watch this illuminating video on Google+](#) to understand why this is important. In this section, Google Webmaster Tools reports the effect of +1 sharing on your site performance in search results.

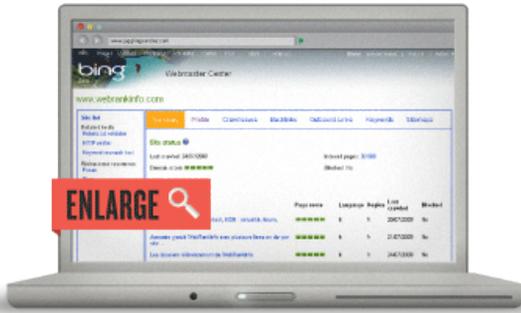
### Labs

The Labs section of Webmaster Tools contains reports that Google considers still in the experimental stage, but important to

where Googlebot found the link to the inaccessible URL.

**HTML Suggestions** - This analysis identifies search engine unfriendly HTML elements. Specifically, it lists meta description issues, title tag issues and non-indexable content issues.

[Sign Up](#)



webmasters nonetheless. One of the most important of these reports is Site Performance, which indicates how fast or slow your site loads for visitors.

## Bing Webmaster Center

### [Bing Webmaster Center](#)

#### Key Features

**Sites Overview**- This interface provides a single overview of all your websites' performance in Bing powered search results. Metrics at a glance include clicks, impressions, pages indexed and number of pages crawled for each site.

**Crawl Stats** - Here you can view reports on how many pages of your site Bing has crawled and discover any errors encountered. Like Google Webmaster, you can also submit sitemaps to help Bing to discover and prioritize your content.

**Index** - This section allows webmasters to view and help control how Bing indexes their web pages. Again, similar to settings in Google Webmaster Tools, here you can explore how your content is organized within Bing, submit URLs, remove URLs from search results, explore inbound links and adjust parameter settings.

**Traffic** - The traffic summary in Bing Webmaster reports impressions and click-through data by combining data from both Bing and Yahoo search results. Reports here show average position as well as cost estimates if you were to buy ads targeting each keyword.

[Sign Up](#)

## SEOMOZ OPEN SITE EXPLORER

While not run by the search engines, SEOMoz's [Open Site Explorer](#) provides similar data.

#### Features

**Identify Powerful Links** - Open Site Explorer sorts all of your inbound links by their metrics that help you determine which links are most important.

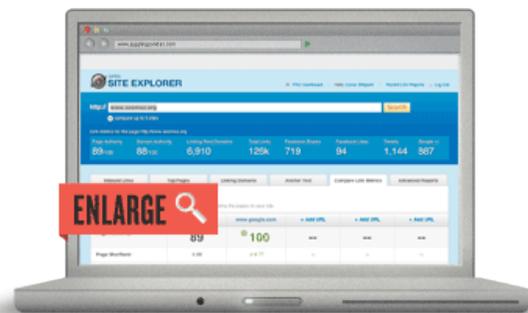
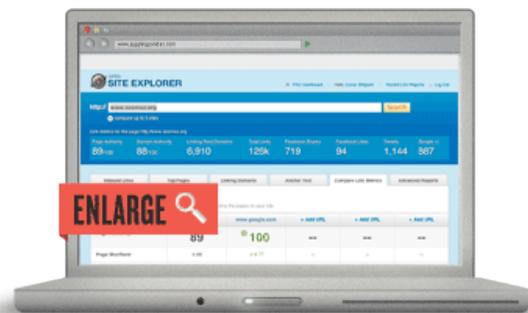
**Find the Strongest Linking Domains** - This tool shows you the strongest domains linking to your domain.

**Analyze Link Anchor Text Distribution** - Open Site Explorer shows you the distribution of the text people used when linking to you.

**Head to Head Comparison View** - This feature allows you to compare two websites to see why one is outranking the other.

**Social Share Metrics** - Measure Facebook Shares, Likes, Tweets, and +1's for any URL.

For more information, click below:



[Learn More](#)

Search engines have only recently started providing better tools to help webmasters improve their search results. This is a big step forward in SEO and the webmaster/search engine relationship. That said, the engines can only go so far with helping webmasters. It is true today, and will likely be true in the future that the ultimate responsibility for SEO is on the marketers and webmasters.

It is for this reason that learning SEO for yourself is so important.

## CHAPTER NINE

# MYTHS & MISCONCEPTIONS ABOUT SEARCH ENGINES

Over the past several years, a number of misconceptions have emerged about how the search engines operate. For the beginner SEO, this causes confusion about what's required to perform effectively. In this section, we'll explain the real story behind the myths.

## SEARCH ENGINE SUBMISSION

In classical SEO times (the late 1990's), search engines had "submission" forms that were part of the optimization process. Webmasters & site owners would tag their sites & pages with keyword information, and "submit" them to the engines. Soon after submission, a bot would crawl and include those resources in their index. Simple SEO!

Unfortunately, this process didn't scale very well, the submissions were often spam, and the practice eventually gave way to purely crawl-based engines. Since 2001, not only has search engine submission not been required, but it is actually virtually useless. The engines all publicly note that they rarely use "submission" URLs, and that the best practice is to earn links from other sites. This will [expose your content to the engines naturally](#).

You can still sometimes find submission pages (here's one for [Bing](#)), but these are remnants of time long past, and are essentially useless to the practice of modern SEO. If you hear a pitch from an SEO offering "search engine submission" services, run, don't walk, to a real SEO. Even if the engines used the submission service to crawl your site, you'd be unlikely to earn enough "link juice" to be included in their indices or rank competitively for search queries.



## META TAGS

Once upon a time, much like search engine submission, meta tags (in particular, the meta keywords tag) were an important part of the SEO process. You would include the keywords you wanted your site to rank for and when users typed in those terms, your page could come up in a query. This process was quickly spammed to death, and eventually dropped by all the major engines as an important ranking signal.

It is true that other tags, namely the [title tag](#) (not strictly a meta tag, but often grouped with them) and [meta description tag](#) (covered [previously in this guide](#)), are of **critical importance to SEO best practices**. Additionally, the meta robots tag is an important tool for controlling spider access. However, SEO is not "all about meta tags", at least, not anymore.

## KEYWORD STUFFING & DENSITY

Ever see a page that just looks spammy? Perhaps something like:

"Bob's cheap Seattle plumber is the best cheap Seattle plumber for all your plumbing needs. Contact a cheap Seattle plumber before it's too late"

Not surprisingly, a persistent myth in SEO revolves around the concept that keyword density - a mathematical formula that divides the number of words on a page by the number of instances of a given keyword - is used by the search engines for relevancy & ranking calculations.

Despite being proven untrue time and again, this myth has legs. Many SEO tools still feed on the concept that keyword density is an important metric. It's not. Ignore it and use keywords intelligently and with usability in mind. The value from an extra 10 instances of your keyword on the page is far less than earning one good editorial link from a source that doesn't think you're a search spammer.

## PAID SEARCH HELPS BOLSTER ORGANIC RESULTS

Put on your tin foil hats, it's time for the most common SEO conspiracy theory: spending on search engine advertising (PPC) improves your organic SEO rankings.

In all of the experiences we've ever witnessed or heard about, this has never been proven nor has it ever been a probable explanation for effects in the organic results. Google, Yahoo! & Bing all have very effective walls in their organizations to prevent precisely this type of crossover.

At Google in particular, advertisers spending tens of millions of dollars each month have noted that even they cannot get special access or consideration from the search quality or web spam teams. So long as the existing barriers are in place and the search engines cultures maintain their separation, we believe that this will remain a myth. That said, we have seen anecdotal evidence that bidding on keywords you already organically rank for can help increase your organic click through rate.



---

## Search Engine Spam

As long as there is search, there will always be spam. The practice of spamming the search engines - creating pages and schemes designed to artificially inflate rankings or abuse the ranking algorithms employed to sort content - has been rising since the mid-1990's.

With payouts so high (at one point, a fellow SEO noted to us that a single day ranking atop Google's search results for the query "buy viagra" could bring upwards of \$20,000 in affiliate revenue), it's little wonder that manipulating the engines is such a popular activity on the web. However, it's become increasingly difficult and, in our opinion, less and less worthwhile for two reasons.

### 1. Not Worth the Effort

*Users hate spam, and the search engines have a financial incentive to fight it. Many believe that Google's greatest product advantage over the last 10 years has been their ability to control and remove spam better than their competitors. It's undoubtedly something all the engines spend a great deal of time, effort and resources on. While spam still works on occasion, it generally takes more effort to succeed than producing "good" content, and the long term payoff is virtually non-existent.*

*Instead of putting all that time and effort into something that the engines will throw away, why not invest in a value added, long term strategy instead?*

### 2. Smarter Engines

*Search engines have done a remarkable job identifying scalable, intelligent methodologies for fighting spam manipulation, making it dramatically more difficult to adversely impact their intended algorithms. Complex concepts like [TrustRank](#) (which SEOMoz's Linkscape index leverages), HITS, statistical analysis, historical data and more have all driven down the value of search spam and made so-called "white hat" tactics (those that don't violate the search engines' guidelines) far more attractive.*

*More recently, [Google's Panda update](#) introduced sophisticated machine learning algorithms to combat spam and low value pages at a scale never before witnessed online. If the search engines' job is to deliver quality results, they have raised the bar year after year.*

*This guide is not intended to show off specific spam tactics, but, due to the large*

number of sites that get penalized, banned or flagged and seek help, we will cover the various factors the engines use to identify spam so as to help SEO practitioners avoid problems. For additional details about spam from the engines, see [Google's Webmaster Guidelines](#) and [Bing's Webmaster FAQs \(pdf\)](#).

The important thing to remember is this: Not only do manipulative techniques not help you in most cases, but often times they cause search engines to impose penalties on your site.

## Page Level Spam Analysis

Search engines perform spam analysis across individual pages and entire websites ([domains](#)). We'll look first at how they evaluate manipulative practices on the [URL](#) level.

### KEYWORD STUFFING

One of the most obvious and unfortunate spamming techniques, keyword stuffing, involves littering repetitions of keyword terms or phrases into a page in order to make it appear more relevant to the search engines. The thought behind this - that increasing the number of times a term is mentioned can considerably boost a page's ranking - is generally false. Studies looking at thousands of the top search results across different queries have found that keyword repetitions play an extremely limited role in boosting rankings, and have a low overall correlation with top placement.

The engines have very obvious and effective ways of fighting this. Scanning a page for stuffed keywords is not massively challenging, and the engines' algorithms are all up to the task. You can read more about this practice, and Google's views on the subject, in a blog post from the head of their web spam team - [SEO Tip: Avoid Keyword Stuffing](#).



### MANIPULATIVE LINKING

One of the most popular forms of web spam, manipulative link acquisition relies on the search engines' use of link popularity in their ranking algorithms to attempt to artificially inflate these metrics and improve visibility. This is one of the most difficult forms of spamming for the search engines to overcome because it can come in so many forms. A few of the many ways manipulative links can appear include:

- \* *Reciprocal link exchange programs, wherein sites create link pages that point back and forth to one another in an attempt to inflate link popularity. The engines are very good at spotting and devaluing these as they fit a very particular pattern.*
- \* *Link schemes, including "link farms" and "link networks" where fake or low value websites are built or maintained purely as link sources to artificially inflate popularity. The engines combat these through numerous methods of detecting connections between site registrations, link overlap or other common factors.*
- \* *Paid links, where those seeking to earn higher rankings buy links from sites and pages willing to place a link in exchange for funds. These sometimes evolve into larger networks of link buyers and sellers, and although the engines work hard to stop them (and Google in particular has taken dramatic actions), they persist in providing value to many buyers & sellers (see [this post on paid links](#) for more on that perspective).*
- \* *Low quality directory links are a frequent source of manipulation for many in the SEO field. A large number of pay-for-placement web directories exist to serve this market and pass*

*themselves off as legitimate with varying degrees of success. Google often takes action against these sites by removing the PageRank score from the toolbar (or reducing it dramatically), but won't do this in all cases.*

There are many more manipulative link building tactics that the search engines have identified and, in most cases, found algorithmic methods for reducing their impact. As new spam systems emerge, engineers will continue to fight them with targeted algorithms, human reviews and the collection of spam reports from webmasters & SEOs.

## CLOAKING

A basic tenet of all the search engine guidelines is to show the same content to the engine's crawlers that you'd show to an ordinary visitor. This means, among other things, not to hide text in the html code of your website that a normal visitor can't see.

When this guideline is broken, the engines call it "cloaking" and take action to prevent these pages from ranking in their results. Cloaking can be accomplished in any number of ways and for a variety of reasons, both positive and negative. **In some cases, the engines may let practices that are technically "cloaking" pass, as they're done for positive user experience reasons.** For more on the subject of cloaking and the levels of risk associated with various tactics and intents, see this post, [White Hat Cloaking](#), from Rand Fishkin.

## "LOW VALUE" PAGES

Although it may not technically be considered "web spam," the engines all have methods to determine if a page provides unique content and "value" to its searchers before including it in their web indices and search results. The most commonly filtered types of pages are "thin" affiliate content, duplicate content, and dynamically generated content pages that provide very little unique text or value. The engines are against including these pages and use a variety of content and link analysis algorithms to filter out "low value" pages from appearing in the results.

Google's 2011 Panda update took the most aggressive steps ever seen in reducing low quality content across the web, and [Google continues to update](#) this process.

---

## Domain Level Spam Analysis

In addition to watching individual pages for spam, engines can also identify traits and properties across entire root domains or subdomains that could flag them as spam. Obviously, excluding entire domains is tricky business, but it's also much more practical in cases where greater scalability is required.

## LINKING PRACTICES

Just as with individual pages, the engines can monitor the kinds of links and quality of referrals sent to a website. Sites that are clearly engaging in the manipulative activities described above on a consistent or seriously impacting way may see their search traffic suffer, or even have their sites banned from the index. You can read about some examples of this from past posts - [Widgetbait Gone Wild](#) or the more recent coverage of the [JC Penney Google penalty](#).

## TRUSTWORTHINESS

Websites that earn trusted status are often treated differently from those who have not. In fact, many SEOs have commented on the "double standards" that exist for judging "big brand" and high importance sites vs. newer, independent sites. For the search engines, trust most likely has a lot to do with the links your domain has earned. Thus, if you publish low quality, duplicate content on your personal blog, then buy several links from spammy directories, you're

## CONTENT VALUE

Similar to how a page's value is judged against criteria such as uniqueness and the experience it provides to search visitors, so too does this principle apply to entire domains. Sites that primarily serve non-unique, non-valuable content may find themselves unable to rank, even if classic on and off page factors are performed acceptably. The engines simply don't want thousands of copies of Wikipedia or Amazon affiliate websites filling up their index, and thus use



likely to encounter considerable ranking problems. However, if you were to post that same content to a page on Wikipedia and get those same spammy links to point to that URL, it would likely still rank tremendously well - such is the power of domain trust & authority.

Trust built through links is also a great method for the engines to employ. A little duplicate content and a few suspicious links are far more likely to be overlooked if your site has earned hundreds of links from high quality, editorial sources like CNN.com or Cornell.edu. On the flip side, if you have yet to earn high quality links, judgments may be far stricter from an algorithmic view.

algorithmic and manual review methods to prevent this.

Search engines constantly evaluate the effectiveness of their own results. They measure when users click on a result, quickly hit the "back" button on their browser, and try another result. This indicates that the result they served didn't meet the user's query.

It's not enough just to rank for a query. Once you've earned your ranking, you have to prove it over and over again.

---

## So How Do You Know If You've Been Bad?

It can be tough to know if your site/page actually has a penalty or if things have changed, either in the search engines' algorithms or on your site that negatively impacted rankings or inclusion. Before you assume a penalty, check for the following:



*Once you've ruled out the list below, follow the flowchart beneath for more specific advice.*

### **Errors**

Errors on your site that may have inhibited or prevented crawling. [Google's Webmaster Tools](#) is a good, free place to start.

### **Changes**

Changes to your site or pages that may have changed the way search engines view your content. (on-page changes, internal link structure changes, content moves, etc.)

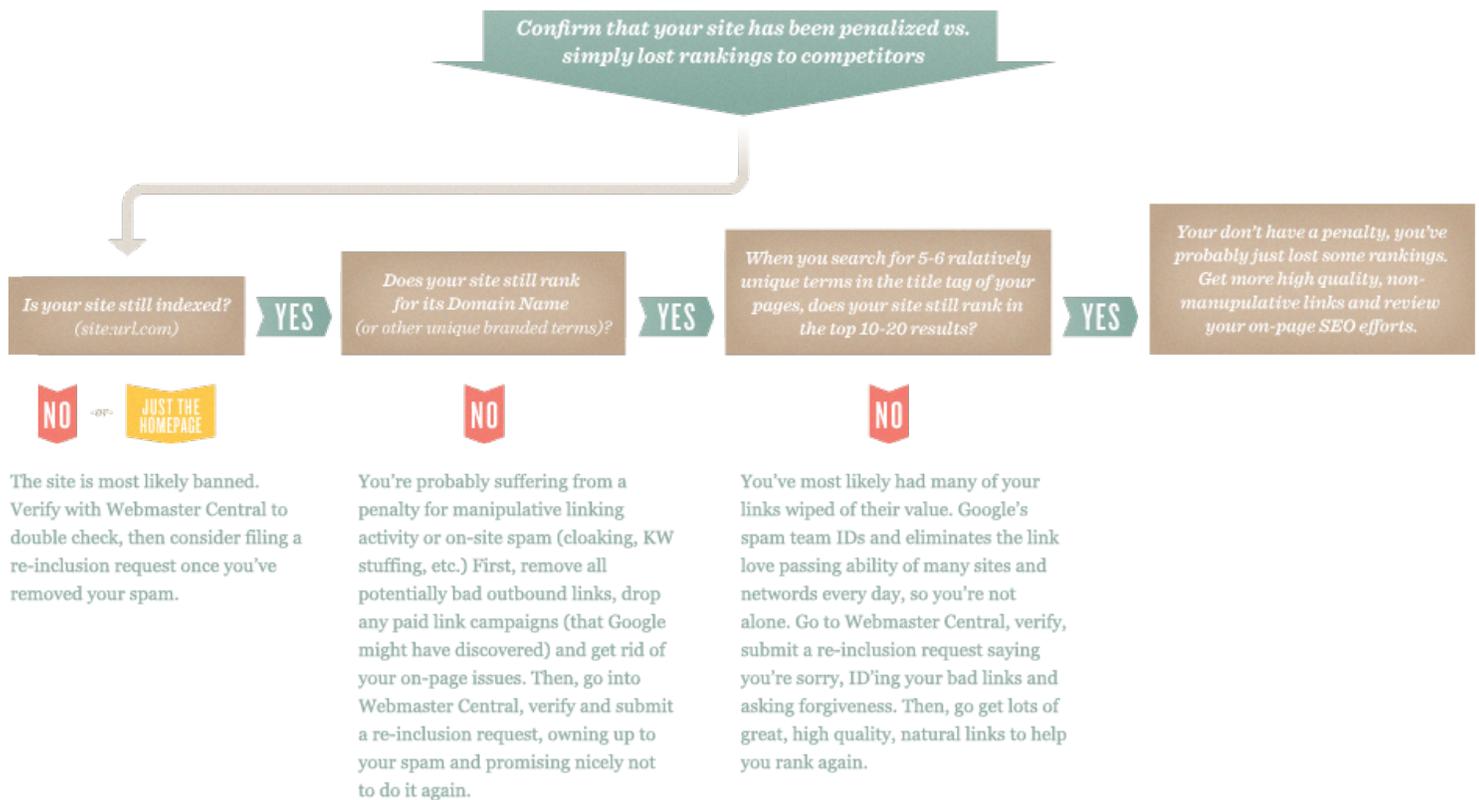
### **Similarity**

Sites that share similar backlink profiles, and whether they've also lost rankings - when the engines update ranking algorithms, link valuation and importance can shift, causing ranking movements.

### **Duplicate Content**

Modern websites are rife with duplicate content problems, especially when they scale to large size. Check out this post on [duplicate content](#) to identify common problems.

## STEP 2 FOLLOW FLOWCHART



While this chart's process won't work for every situation, the logic has been uncanny in helping us identify spam penalties or mistaken flagging for spam by the engines and separating those from basic ranking drops. [This page from Google](#) (and the embedded Youtube video) may also provide value on this topic.

## Getting Penalties Lifted

The task of requesting re-consideration or re-inclusion in the engines is painful and often unsuccessful. It's also rarely accompanied by any feedback to let you know what happened or why. However, it is important to know what to do in the event of a penalty or banning.

Hence, the following recommendations:

**1** If you haven't already, register your site with the engine's Webmaster Tools service ([Google's](#) and [Bing's](#)). This registration creates an additional layer of trust and connection between your site and the webmaster teams.

**5** Remove/fix everything you can. If you've acquired bad links, try to get them taken down. If you've done any manipulation on your own site (over-optimized internal linking, keyword stuffing, etc.), get it off before you submit your request.

- 2 *Make sure to thoroughly review the data in your Webmaster Tools accounts, from broken pages to server or crawl errors to warnings or spam alert messages. Very often, what's initially perceived as a mistaken spam penalty is, in fact, related to accessibility issues.*
- 3 *Send your re-consideration/re-inclusion request through the engine's Webmaster Tools service rather than the public form - again, creating a greater trust layer and a better chance of hearing back.*
- 4 *Full disclosure is critical to getting consideration. If you've been spamming, own up to everything you've done - links you've acquired, how you got them, who sold them to you, etc. The engines, particularly Google, want the details, as they'll apply this information to their algorithms for the future. Hold back, and they're likely to view you as dishonest, corrupt or simply incorrigible (and fail to ever respond).*
- 6 *Get ready to wait - responses can take weeks, even months, and re-inclusion itself, if it happens, is a lengthy process. Hundreds (maybe thousands) of sites are penalized every week, so you can imagine the backlog the webmaster teams encounter.*
- 7 *If you run a large, powerful brand on the web, re-inclusion can be faster by going directly to an individual source at a conference or event. Engineers from all of the engines regularly participate in search industry conferences ([SMX](#), [SES](#), [Pubcon](#), etc.), and the cost of a ticket can easily outweigh the value of being re-included more quickly than a standard request might take.*

Be aware that with the search engines, lifting a penalty is not their obligation or responsibility. Legally, they have the right to include or reject any site/page for any reason. Inclusion is a privilege, not a right, so be cautious and don't apply techniques you're unsure or skeptical of - or you could find yourself in a very rough spot.

# 10

## CHAPTER TEN

# MEASURING & TRACKING SUCCESS

They say that if you can measure it, then you can improve it. In search engine optimization, measurement is critical to success. Professional SEOs track data about rankings, referrals, links and more to help analyze their SEO strategy and create road maps for success.

## Recommended Metrics to Track

Although every business is unique and every website has different metrics that matter, the following list is nearly universal. Note that we're only covering those metrics critical to SEO - optimizing for the search engines. As a result, more general metrics may not be included. For a more comprehensive look at web analytics, check out [Choosing Web Analytics Key Performance Indicators](#) from Avinash Kaushik's excellent [Web Analytics Blog](#).

### 1. SEARCH ENGINE SHARE OF REFERRING VISITS

Every month, it's critical to keep track of the contribution of each traffic source for your site. These include:

- \* **Direct Navigation:** Typed in traffic, bookmarks, email links without tracking codes, etc.
- \* **Referral Traffic:** From links across the web or in trackable email, promotion & branding campaign links
- \* **Search Traffic:** Queries that sent traffic from any major or minor web search engine

Knowing both the percentage and exact numbers will help you identify weaknesses and serve as a comparison over time for trend data. For example, if you see that traffic has spiked dramatically but it comes from referral links with low relevance, it's not time to get excited. On the other hand, if search engine traffic falls dramatically, you may be in trouble. You should use this data to track your marketing efforts and plan your traffic acquisition efforts.

MONTH	VISITS	TOTAL NUMBER OF VISITS TO SEOMOZ (January 1 - March 17, 2009)
January 2009	340,305	
February 2009	391,978	
March 2009	219,900	
<b>Total</b>	<b>952,263</b>	

SEARCH ENGINES	VISITS	NUMBER OF SEARCH ENGINE VISITS (January 1 - March 17, 2009)
Google	218,721	
Yahoo	10,360	
Windows Live	3,257	
MSN Search	1,840	
Alexa	1,098	
Ask Jeeves	925	
AOL Search	745	
AT&T	242	
Dogpile	178	
Altavista	148	
<b>Subtotal</b>	<b>237,514</b>	
<b>Total</b>	<b>238,286</b>	

$238,286 \div 952,263 = 0.2502$

**25% OF SEOMOZ'S TRAFFIC COMES FROM SEARCH ENGINES**

### 2. VISITS REFERRED BY SPECIFIC SEARCH ENGINES

Three major engines make up 95%+ of all search traffic in the US - Google and the Yahoo-Bing alliance. For most countries outside the US 80%+ of search traffic comes solely from Google (with a few notable exceptions including both Russia and China.) Measuring the contribution of your search traffic from each engine is critical for

several reasons:

### **Compare Performance vs. Market Share**

By tracking not only search engines broadly, but by country, you'll be able to see exactly the contribution level of each engine in accordance with its estimated market share. Keep in mind that in sectors like technology and Internet services, demand is likely to be higher on Google (given its younger, more tech-savvy demographic) than in areas like cooking, sports or real estate.

### **Get Visibility Into Potential Drops**

If your search traffic should drop significantly at any point, knowing the relative and exact contributions from each engine will be essential to diagnosing the issue. If all the engines drop off equally, the problem is almost certainly one of accessibility. If Google drops while the others remain at previous levels, it's more likely to be a penalty or devaluation of your SEO efforts by that singular engine.

### **Uncover Strategic Value**

It's very likely that some efforts you undertake in SEO will have greater positive results on some engines than others. For example, we frequently notice that on-page optimization tactics like better keyword inclusion and targeting has more benefit with Bing & Yahoo! than Google, while gaining specific anchor text links from a large number of domains has a more positive impact on Google than the others. If you can identify the tactics that are having success with one engine, you'll better know how to focus your efforts.

## **3. VISITS REFERRED BY SPECIFIC SEARCH ENGINE TERMS AND PHRASES**

The keywords that send traffic are another important piece of your analytics pie. You'll want to keep track of these on a regular basis to help identify new trends in keyword demand, gauge your performance on key terms and find terms that are bringing significant traffic that you're potentially under optimized for.

You may also find value in tracking search referral counts for terms outside the "top" terms/phrases - those that are important and valuable to your business. If the trend lines are pointing in the wrong direction, you know efforts need to be undertaken to course correct. Search traffic worldwide has consistently risen over the past 15 years, so a decline in quantity of referrals is troubling - check for seasonality issues (keywords that are only in demand certain times of the week/month/year) and rankings (have you dropped, or has search volume ebbed?).

TOP REFERRING PHRASES (January 1 - March 17, 2009)		
SEARCH PHRASES	VISITS	PERCENT
seomoz	10,865	5.30%
seo moz	2,508	1.22%
seo tools	2,267	1.11%
what is seo	1,775	0.87%
seo blog	1,738	0.85%
social media marketing	1,688	0.82%
search engine optimization	1,435	0.70%
seo	1,342	0.65%
yahoo	1,016	0.50%
linkscape	953	0.46%
<b>Subtotal</b>	<b>25,587</b>	<b>12.48%</b>
<b>Total</b>	<b>205,084</b>	<b>100.00%</b>

## **4. CONVERSION RATE BY SEARCH QUERY TERM/PHRASE**

When it comes to the bottom line for your organization, few metrics matter as much as conversion. For example, in the graphic to the right, 5.80% of visitors who reached SEOMoz with the query "SEO Tools" signed up to become members during that visit. This is a much higher conversion rate than most of the 1000s of keywords used to

find our site. With this information, we can now do 2 things.

1. Checking our rankings, we see that we only rank #4 for "SEO Tools". Working to improve this position will undoubtedly lead to more conversion.
2. Because our analytics will also tell us what page these visitors landed on (mostly <http://www.seomoz.org/tools>), we can focus on efforts on that page to improve visitor experience.

The real value from this simplistic tracking comes from the "low-hanging fruit" - seeing keywords that continually send visitors who convert and increasing focus on both rankings and improving the landing pages that visitors reach. While conversion rate tracking from keyword phrase referrals is certainly important, it's never the whole story. Dig deeper and you can often uncover far more interesting and applicable data about how conversion starts and ends on your site.

SEARCH QUERY REFERRAL CONVERSION RATES AT SEOMOZ			
TOOLS & BRANDED TERMS CARRY THE HIGHEST CONVERSION RATES			
SEARCH PHRASES	VISITS	ACTION	CONVERSION
seomoz	10,907	216	1.98%
seo tools	2,274	132	5.80%
seo test	582	95	16.30%
seo moz	2,516	47	1.87%
seomoz.org	867	33	3.81%
free seo tools	400	33	8.25%
rank checker	587	30	5.11%
test my website	131	30	22.90%
seo blog	1,744	28	1.61%
seo	1,346	23	1.71%
linkscope	956	20	2.09%
check backlinks	381	17	4.46%
test my site	56	16	28.57%
search engine optimization	1,441	11	0.76%
crawl my site	42	11	26.19%
crawl test	64	10	15.62%
seo free tools	74	9	12.16%
test seo	86	9	10.47%
seomoz.com	159	8	5.03%
page strength	167	8	4.79%
spider my site	40	7	17.50%
seo analytics	101	6	5.94%

## 5. NUMBER OF PAGES RECEIVING AT LEAST ONE VISIT FROM SEARCH ENGINES

Knowing the number of pages that receive search engine traffic is an essential metric for monitoring overall SEO performance. From this number, we can get a glimpse into indexation - the number of pages the engines are keeping in their indices from our site. For most large websites (50,000+ pages), mere inclusion is essential to earning traffic, and this metric delivers a trackable number that's indicative of success or failure. As you work on issues like site architecture, link acquisition, XML Sitemaps, uniqueness of content and meta data, etc., the trend line should rise, showing that more and more pages are earning their way into the engines' results. Pages receiving search traffic is, quite possibly, the best long tail metric around.

While other analytics data points are of great importance, those mentioned above should be universally applied to get the maximum value from your SEO campaigns.

PAGES RECEIVING SEARCH TRAFFIC FROM GOOGLE	
(January 1 - March 17, 2009)	
GOOGLE SENT AT LEAST ONE REFERRAL TO 8,221 PAGES ON SEOMOZ DURING THIS PERIOD.	
CROSS-REFERENCE FILTERS	
Search Engines (Direct): Google	
ENTRY PAGE URL	VISITS
<a href="http://www.seomoz.org/users/view/100932">http://www.seomoz.org/users/view/100932</a>	1
<a href="http://www.seomoz.org/ga/view/899">http://www.seomoz.org/ga/view/899</a>	1
<a href="http://www.seomoz.org/ga/view/116">http://www.seomoz.org/ga/view/116</a>	1
<a href="http://www.seomoz.org/ga/view/940">http://www.seomoz.org/ga/view/940</a>	1
<a href="http://seomoz.org/marketplace/companies/view/804">http://seomoz.org/marketplace/companies/view/804</a>	1
<a href="http://seomoz.org/marketplace/companies/view/806">http://seomoz.org/marketplace/companies/view/806</a>	1
<a href="http://www.seomoz.org/cat/view/Other-Verticals-in-Search">http://www.seomoz.org/cat/view/Other-Verticals-in-Search</a>	1
<a href="http://www.seomoz.org/cat/view/Blogging-Trends">http://www.seomoz.org/cat/view/Blogging-Trends</a>	1
<a href="http://www.seomoz.org/ga/view/267">http://www.seomoz.org/ga/view/267</a>	1

### Google's (not provided) Keywords

In 2011, Google announced it will no longer pass keyword query data through its referrer string for logged in users. This means that instead of showing organic keyword data in Google Analytics, visits from users logged into Google will show as "not provided." At the time, Google said they expected this to effect less than 10% of all search queries.

Soon after, many webmasters started reporting **up to 20% of their search queries** as keyword (not provided). Google responded by saying that the 10% figure was an average across all worldwide sites and that some differences would exist based on country location and type of website.

With the launch of Google+, webmasters fear that more and more users will create, and log into, Google accounts. This would result in an even greater percentage of "not provided" keywords.

How this will eventually play out is anyone's guess. In the meantime, smart SEOs and web analytics experts have devised workarounds to try and recover some of this missing keyword data, although nothing can substitute for the real thing. Read more about dealing with (not provided) keywords [in this blog post](#).

---

## Analytics Software

### *The Right Tools for the Job*



- \* [Omniture](#)
- \* [Fireclick](#)
- \* [Mint](#)
- \* [Sawmill Analytics](#)
- \* [Clicktale](#)
- \* [Coremetrics](#)
- \* [Unica Affinium NetInsight](#)



- \* [Yahoo! Web Analytics](#)  
(formerly Indextools)
- \* [Google Analytics](#)
- \* [Clicky Web Analytics](#)
- \* [Piwik Open Source Analysis](#)
- \* [Woopra Website Tracking](#)
- \* [AWStats](#)

### **Additional Reading:**

- \* [How to Choose a Web Analytics Tool: A Radical Alternative](#) - From Avinash Kaushik way back in 2006 (but still a relevant and quality piece)

While choosing can be tough, our top recommendation is Google Analytics. Because of its broad adoption you can find many tutorials and guides available online. Google Analytics also has the advantage of cross-integration with other Google products such as Webmaster Tools, Adwords and AdSense.

No matter which analytics software you decide is right for you, we also strongly recommend testing different versions of pages on your site and making conversion rate improvements based on the results. Testing pages on your site can be as simple as using a free tool to test two versions of a page header or as complex as using an expensive multivariate software to simultaneously test hundreds of variants of a page. There are many testing platforms out there, but if you're looking to put a first toe in the testing waters, one free, easy to use solution we recommend is [Google's Website Optimizer](#). It's a great way to get started running tests that can inform powerful conversion rate improvements.

---

## Metrics for Measuring

### *Search Engine Optimization*

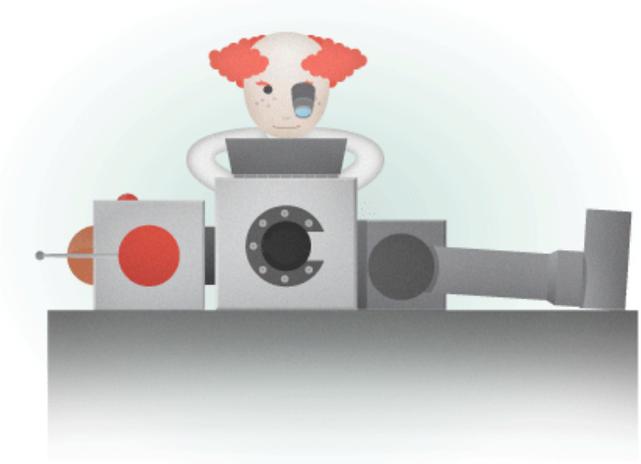
In organic SEO, it can be difficult to track the specific elements of the engines' algorithms effectively given that this data is not public, nor is it even well researched. However, a combination of tactics have

become best practices, and new data is constantly emerging to help track direct ranking elements and positive/negative ranking signals. The data points covered below are ones that we will occasionally recommend to track campaigns and have proven to add value when used in concert with analytics.

## Metrics Provided by Search Engines

We've already discussed many of the data points provided by services such as Google's Webmaster Tools, Yahoo! Site Explorer and Microsoft's Webmaster Tools. In addition to these, the engines provide some insight through publicly available queries and competitive intelligence. Below is a list of queries/tools /metrics from the engines, along with their respective applications.

Employing these queries & tools effectively requires that you have an informational need with an actionable solution. The data itself isn't valuable unless you have a plan of what to change/build/do once you learn what you need to know (this holds true for competitive analysis as well).



# Google™

### Google Site Query

**e.g., [site:seomoz.org](#)** - useful to see the number and list of pages indexed on a particular domain. You can expand the value by adding additional query parameters. For example - [site:seomoz.org/blog inurl:tools](#) - will show only those pages in Google's index that are in the blog and contain the word "tools" in the URL. While this number fluctuates, it's still a good rough measurement. You can read more about this in this [blog post](#).

### Google Trends

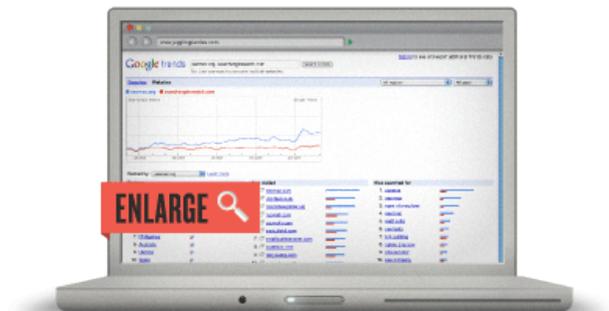
**Available at [Google.com/Trends](#)** - this shows keyword search volume/popularity data over time. If you're logged into your Google account, you can also get specific numbers on the charts, rather than just trend lines.

### Google Trends for Websites

**Available at [Trends.Google.com/websites](#)** - This shows traffic data for websites according to Google's data sources (toolbar, ISP data, analytics and others may be part of this). A logged in user account will show numbers in the chart to indicate estimated traffic levels.

### Google Insights for Search

**Available at [google.com/insights/search](#)** - this tool provides data about regional usage, popularity and related queries for keywords.





### **Bing Site Query**

**e.g., site:seomoz.org** - just like Yahoo! and Google, Bing allows for queries to show the number and list of pages in their index from a given site. Unfortunately, Bing's counts are given to wild fluctuation and massive inaccuracy, often rendering the counts themselves useless.

### **Bing IP Query**

**e.g., ip:216.176.191.233** - this query will show pages that Microsoft's engine has found on the given IP address. This can be useful in identifying shared hosting and seeing what other sites are hosted on a given IP address.

### **Microsoft Ad Intelligence**

Available at [Microsoft Advertising](#) - a great variety of keyword research and audience intelligence tools are provided by Microsoft, primarily for search and display advertising. This guide won't dive deep into the value of each individual tool, but they are worth investigating and many can be applied to SEO.



### **Page Specific Metrics**

**Page Authority** - Page Authority predicts the likelihood of a single page to rank well, regardless of its content. The higher the Page Authority, the greater the potential for that individual page to rank.

**mozRank** - [mozRank](#) refers to SEOMoz's general, logarithmically scaled 10-point measure of global link authority (or popularity). mozRank is very similar in purpose to the measures of static importance (which means importance independent of a specific query) that are used by the search engines (e.g., Google's PageRank or FAST's StaticRank). Search engines often rank pages with higher global link authority ahead of pages with lower authority. Because measures like mozRank are global and static, this ranking power applies to a broad range of search queries, rather than pages optimized specifically for a particular keyword.

**mozTrust** - Like mozRank, [mozTrust](#) is distributed through links. First, trustworthy "seeds" are identified to feed the calculation of the metric. (These include the homepages of major international university, media and governmental websites.) Websites that earn links from the seed set are then able to cast (lesser) trust-votes through their links. This process continues across the web and the mozTrust of each applicable link decreases as it travels "farther" from the original trusted seed site.

**# of Links** - The total number of pages that contain at least one link to this page. For example, if the Library of Congress homepage (<http://www.loc.gov/index.html>) linked to the White House's homepage (<http://www.whitehouse.gov>) in both the page content and the footer, this would still be counted as only a single link.



### **Ask Site Query**

**e.g., site:seomoz.org inurl:www** - Ask.com is a bit picky in its requirements around use of the site query operator. To function properly, an additional query must be used (although generic queries such as the example above are useful to see what a broad "site" query would normally return).



### **Blog Search Link Query**

**e.g., link:www.seomoz.org/blog** - Although Google's normal web search link command is not always useful, their blog search link query shows generally high quality data and can be sorted by date range and relevance. You can read more about this in this [blog post](#).



### **Domain Specific Metrics**

**Domain Authority** - Domain Authority predicts how well a web page on a specific domain will rank. The higher the Domain Authority, the greater the potential for an individual page on that domain to rank well.

**Domain mozRank** - Domain-level mozRank (DmR) quantifies the popularity of a given domain compared to all other domains on the web. DmR is computed for both [subdomains](#) and [root domains](#). This metric uses the same algorithm as mozRank but applies it to the "domain-level link graph". (A view of the web that only looks at domains as a whole and ignores individual pages) Viewing the web from this perspective offers additional insight about the general authority of a domain. Just as pages can endorse other pages, a link which crosses domain boundaries (e.g., from a page on [searchengineland.com](http://searchengineland.com) to a page on [www.seomoz.org](http://www.seomoz.org)) can be seen as endorsement by one domain for another.

**Domain mozTrust** - Just as mozRank can be applied at the domain level ([Domain-level mozRank](#)), so can mozTrust. Domain-level mozTrust is like mozTrust but instead of being calculated between web pages, it is calculated between entire domains. New or poorly linked-to pages on highly trusted domains may inherit some natural trust by virtue of being hosted on the trusted domain. Domain-Level mozTrust is expressed on a 10-point logarithmic scale.

**# of Links** - the quantity of pages that contain at least one link to the domain. For example, if <http://www.loc.gov/index.html> and <http://www.loc.gov/about> both contained links to <http://www.nasa.gov>, this would count as two links to the domain.

**# of Linking Root Domains** - The total number of unique root domains that contain a link to this page. For example, if topics.nytimes.com and www.nytimes.com both linked to the homepage of SEOmoz (<http://www.seomoz.org>), this would count as only a single linking root domain.

**External mozRank** - Whereas mozRank measures the link juice (ranking power) of both [internal](#) and [external links](#), external mozRank measures only the amount of mozRank flowing through external links (links located on a separate domain). Because external links can play an important role as independent endorsements, external mozRank is an important metric for predicting search engine rankings.

**# of Linking Root Domains** - the quantity of different domains that contain at least one page with a link to any page on this site. For example, if <http://www.loc.gov/index.html> and <http://www.loc.gov/about> both contained links to <http://www.nasa.gov>, this would count as only a single linking root domain to nasa.gov.

---

## Applying that Data

### *To Your Campaign*

Just knowing the numbers won't help unless you can effectively interpret and apply changes to course-correct. Below, we've taken a sample of some of the most common directional signals provided by tracking data points and how to respond with actions to improve or execute on opportunities.

## Fluctuation

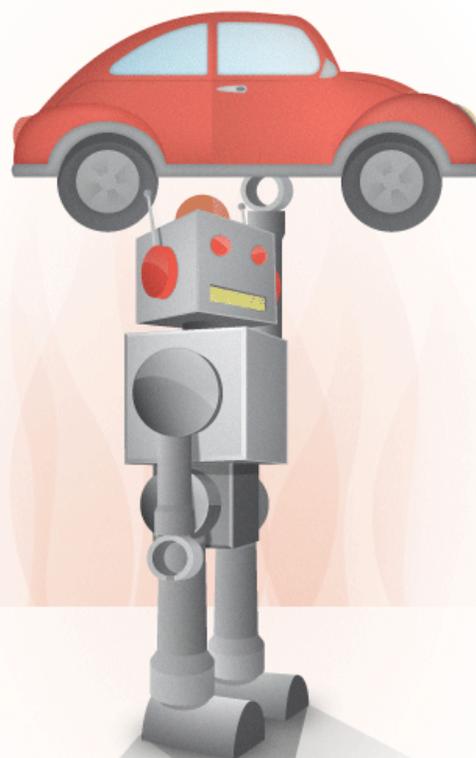
### *In Search Engine Page and Link Count Numbers*

The numbers reported in "site:" and "link:" queries are rarely precise, and thus we strongly recommend not getting too worried about fluctuations showing massive increases or decreases unless they are accompanied by traffic drops. For example, on any given day, Yahoo! reports between 800,000 and 2 million links to the SEOmoz.org domain. Obviously, we don't gain or lose hundreds of thousands of links each day, but the variability of Yahoo!'s indices means that these numbers reports provide little guidance about our actual link growth or shrinkage.

If you do see significant drops in links or pages indexed accompanied by similar traffic referral drops from the search engines, you may be experiencing a real loss of link juice (check to see if important links that were previously sending traffic/rankings boosts still exist) or a loss of indexation due to penalties, hacking, malware, etc. A thorough analysis using your own web analytics and [Google's Webmaster Tools](#) can help to identify potential problems.

## Falling

### *Search Traffic from a Single Engine*



1. You're under a penalty at that engine for violating search quality or terms of service guidelines. Check out [this post on how to identify/handle a search engine penalty](#).

***If a single engine is sending you considerably less traffic for a wide range of search queries, a small number of possibilities exist.***

- 2.** *You've accidentally blocked access to that search engine's crawler. Double-check your robots.txt file and meta robots tags and review the Webmaster Tools for that engine to see if any issues exist.*
- 3.** *That engine has changed their ranking algorithm in a fashion that no longer favors your site. Most frequently, this happens because links pointing to your site have been devalued in some way, and is especially prevalent for sites that engage in manual link building campaigns of low-moderate quality links.*

***“Identify the problem most likely to be the culprit and investigate. Forums like Cre8asit Forums, HighRankings and Google’s Groups for Webmasters who can help.”***

## Falling

### ***Search Traffic from Multiple Engines***

Chances are good that you've done something on your site to block crawlers or stop indexation. This could be something in the robots.txt or meta robots tags, a problem with hosting/uptime, a DNS resolution issue or a number of other technical breakdowns. Talk to your system administrator, developers and/or hosting provider and carefully review your Webmaster Tools accounts and analytics to help determine potential causes.

## Individual

### ***Ranking Fluctuations***

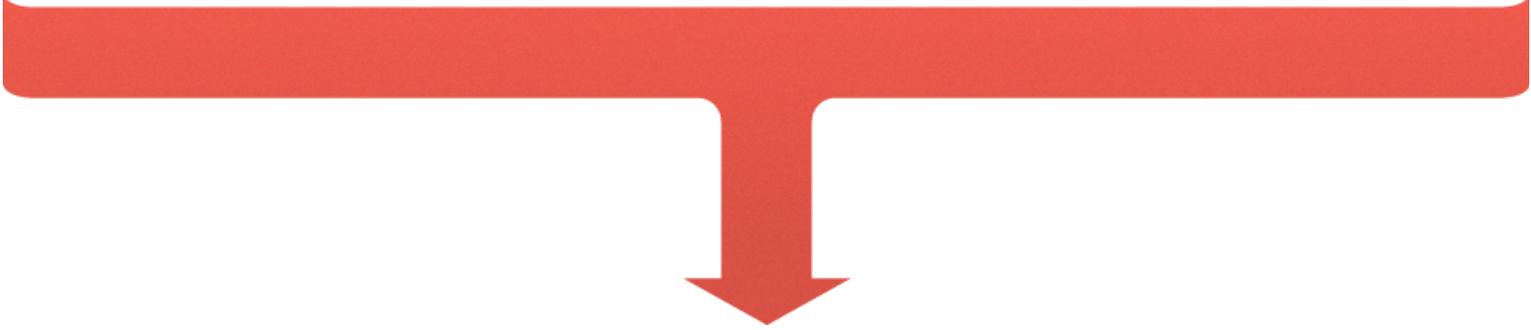
Gaining or losing rankings for a particular term/phrase or even several happens millions of times a day to millions of pages and is generally nothing to be concerned about. Ranking algorithms fluctuate, competitors gain and lose links (and on-page optimization tactics) and search engines even flux between indices (and may sometimes even make mistakes in their crawling, inclusion or ranking processes). When a dramatic rankings decrease occurs, you might want to carefully review on-page elements for any signs of over-optimization or violation of guidelines (cloaking, keyword stuffing, etc.) and check to see if links have recently been gained or lost. Note that with sudden spikes in rankings for new content, a temporary period of high visibility followed by a dramatic drop is common (in the SEO field, we refer to this as the "freshness boost").

*“Don't panic over small fluctuations. With large drops, be wary against making a judgment call until at least a few days have passed. If you run a new site or are in the process of link acquisition and active marketing, these sudden spikes and drops are even more common, so simply be prepared and keep working.”*

## Positive

### ***Increases in Link Metrics Without Rankings Increases***

Many site owners worry that when they've done some "classic" SEO - on-page optimization, link acquisition, etc. they can expect instant results. This, sadly, is not the case. Particularly for new sites, pages and content that's competing in very difficult results, rankings take time and even earning lots of great links is not a sure recipe to instantly reach the top. Remember that the engines need to not only crawl all those pages where you've acquired links, but index and process them - given the almost certain use of delta indices by the engines to help with freshness, the metrics and rankings you're seeking may be days or even weeks behind the progress you've made.



*Congratulations! You've made it  
through the entire Beginner's Guide to SEO!  
Now you are ready to start optimizing your own site,  
implement the tricks you have just learned  
and improve your search results in the major search engines.*



**THE END**

---

## Contributors

We would like to extend a very heartfelt thank you to all of the people who contributed to this guide:

- \* [Urban Influence](#)
- \* [Linda Jenkinson](#)
- \* [Tom Critchlow](#)
- \* [Will Critchlow](#)
- \* [Dr. Pete](#)
- \* [Hamlet Batista](#)
- \* [chuckallied](#)

- \* [lorisa](#)
- \* [Optomo](#)
- \* [identity](#)
- \* [Pat Sexton](#)
- \* [SeoCatfish](#)
- \* [David LaFerney](#)
- \* [Kimber](#)
- \* [g1smd](#)
- \* [Steph Woods](#)
- \* [robbothan](#)
- \* [RandyP](#)
- \* [bookworm seo](#)
- \* [Rafi Kaufman](#)
- \* [Sam Niccolls](#)
- \* [Danny Dover](#)
- \* [Cyrus Shepard](#)
- \* [Sha Menz](#)
- \* [Casey Henry](#)
- \* [and Rand Fishkin](#)

