THE BEGINNERS GUIDE TO SEO

PDF

SEOmoz

2nd Edition

CHAPTER ONE — HOW SEARCH ENGINES OPERATE

Search engines have four functions – crawling, building an index, calculating relevancy & rankings and 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.

I. CRAWLING AND INDEXING
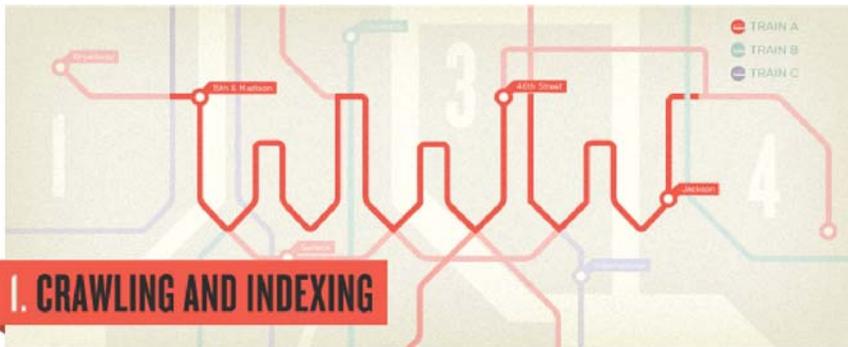
Imagine the World Wide Web as a network of stops in a big city subway system.

Each stop is its own unique document (usually a web page, but sometimes a PDF, 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 together all of the pages in existence."

(Or, at least, all those that the engines can access.) Through links, search engines' automated robots, called "crawlers," or "spiders" can reach the many billions of interconnected documents.

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

These monstrous storage facilities hold thousands of machines processing unimaginably large quantities of information. After all, when a person performs a search at any of the major engines, they demand results instantaneously – even a 3 or 4 second delay can cause dissatisfaction, so the engines work hard to provide answers as fast as possible.

2. PROVIDING ANSWERS
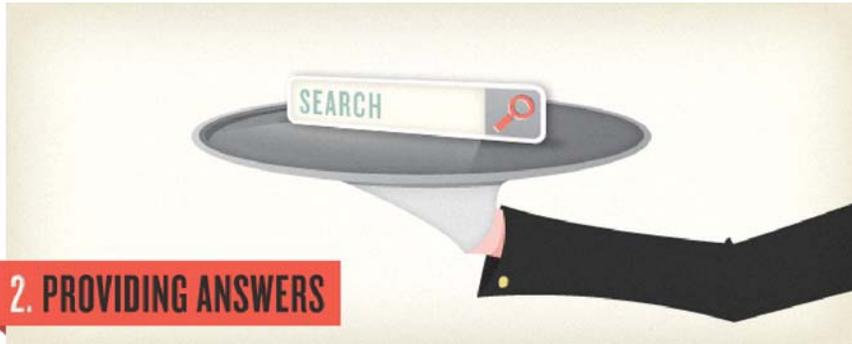
When a person searches for something online, it requires the search engines to scour their corpus of billions of documents and do two things – first, return only those results that are relevant or useful to the searcher's query, and second, rank those results in order of perceived value (or importance). It is both "relevance" and "importance" that the process of search engine optimization is meant to influence.

To the search engines, relevance means more than simply having a page with the words you searched for prominently displayed. In the early days of the web, search engines didn't go much further than this simplistic step, and found that their results suffered as a consequence. Thus, through iterative evolution, smart engineers at the various engines devised better ways to find valuable results that searchers would appreciate and enjoy. Today, hundreds of factors influence relevance, many of which we'll discuss throughout this guide.

Importance is an equally tough concept to quantify, but Search engines must do their best.

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 (and thank goodness, because those trillions of man-hours would require Earth's entire population as a workforce). Instead, the engines craft careful, mathematical equations – algorithms – to sort the wheat from the chaff and to then rank the wheat in order of 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." For those who are particularly interested, 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 Study & Learn How to Succeed in the Engines

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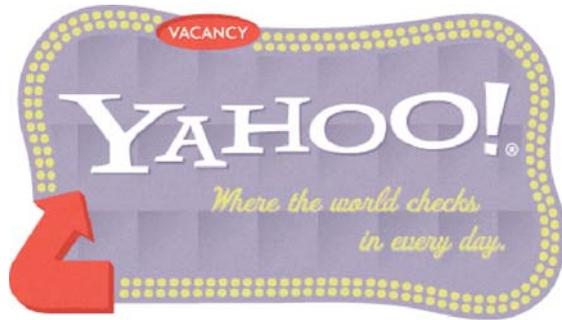
SEO INFORMATION FROM YAHOO! WEBMASTER GUIDELINES

Many factors influence whether a particular web site appears in Web Search results and where it falls in the ranking.

These factors can include:

* *The number of other sites linking to it*
* *The content of the pages*
* *The updates made to indicies*
* *The testing of new product versions*
* *The discovery of additional sites*
* *Changes to the search algorithm – and other factors*

SEO INFORMATION FROM BING WEBMASTER GUIDELINES

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

* *In the visible page text, include words users might choose as search query terms to find the information on your site.*
* *Limit all pages to a reasonable size. We recommend one topic per page. An HTML page with no pictures should be under 150 KB.*
* *Make sure that each page is accessible by at least one static text link.*
* *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.*

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.*
* *Keep the links on a given page to a reasonable number (fewer than 100).*

yes.

But Have No Fear Fellow Search Marketer!

Over the 12 plus years that web search has existed online, search marketers have found methodologies to extract information about how the search engines rank pages and use that data to help their sites and their clients achieve better positioning.

Surprisingly, the engines do support many of these efforts, though the public visibility is frequently low. Conferences on search marketing, such as the Search Marketing Expo, WebMasterWorld, Search Engine Strategies, & SEOmoz's SEO Training Seminars 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 would carry 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 the pages, we found that the page linked to from the highest link on the home page ranked first and continued our iterations of testing.
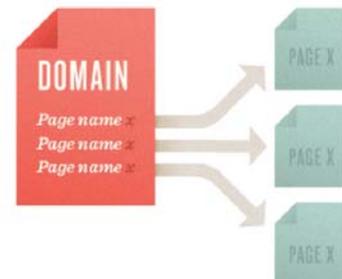
This process is certainly not alone in helping to educate search marketers.

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 and edification. 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 and successful search marketers – remedial calculus isn't required to practice search engine optimization.
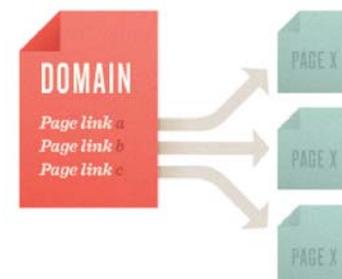
Through methods like patent analysis, experiments, and live testing and tweaking, search marketers as a community have come to understand many of the basic operations of search engines and the critical components of creating websites and pages that garner high rankings and significant traffic.

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

"Step" to "Domain"

DOMAIN
Page name x
Page name x
Page name x

PAGE X
PAGE X
PAGE X

"Step" to "Domain"

DOMAIN
Page link a
Page link b
Page link c

PAGE X
PAGE X
PAGE X

CHAPTER TWO — HOW PEOPLE INTERACT WITH SEARCH ENGINES

DOWNLOAD PDF

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

Search engine usage has evolved over the years but the primary principles of conducting a search remain largely unchanged. Listed here is 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. *Execute the query at 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.*

When this process results in the satisfactory completion of a task, a positive experience is created, both with the search engine and the site providing the information or result. Since the inception of web search, the activity has grown to heights of great popularity, such that in December of 2005, the Pew Internet & American Life Project (PDF Study in Conjunction with ComScore) found that 90% of online men and 91% of online women used search engines. Of these, 42% of the men and 39% of the women reported using search engines every day and more than 85% of both groups say they "found the information they were looking for."

A Broad Picture With Fascinating Data

When looking at the broad picture of search engine usage, fascinating data is available from a multitude of sources. We've extracted those that are recent, relevant, and valuable, not only for understanding how users search, but in presenting a compelling argument about the power of search (which we suspect many readers of this guide may need to do for their managers):
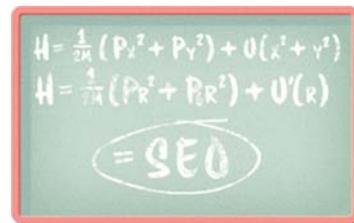
An April 2010 study by comScore found:

* *Google Sites led the U.S. core search market in April with 64.4 percent of the searches conducted, followed by Yahoo! Sites (up 0.8 percentage points to 17.7 percent), and Microsoft Sites (up 0.1 percentage points to 11.8 percent).*

* *Americans conducted 15.5 billion searches in April, up slightly from March. Google Sites accounted for 10 billion searches, followed by Yahoo! Sites (2.8 billion), Microsoft Sites (1.8 billion), Ask Network (574 million) and AOL LLC (371 million).*

* *In the April analysis of the top properties where search activity is observed, Google Sites led the search market with 14.0 billion search queries, followed by Yahoo! Sites with 2.8 billion queries and Microsoft Sites with 1.9 billion. Amazon Sites experienced sizeable growth during the month with an 8-percent increase to 245 million searches, rounding off the top 10 ranking.*

VIEW ONLINE

An August 2008 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 just under one-half (49 percent).*

* *With this increase, the number of those using a search engine on a typical day is pulling ever closer to the 60 percent of Internet users who use e-mail, arguably the Internet's all-time killer app, on a typical day.*

VIEW ONLINE

A EightFoldLogic (formally Enquisite) report from 2009 on click-through traffic in the US showed:

* *Google sends 78.43% of traffic.*

* *Yahoo! sends 9.73% of traffic.*

* *Bing sends 7.86% of traffic.*

VIEW ONLINE

A July 2009 Forrester report remarked:

* Interactive marketing will near $55 billion in 2014.

* This spend will represent 21% of all marketing budgets.

`VIEW ONLINE`

Webvisible & Nielsen produced a 2007 report on local search that noted:

* 74% of respondents used search engines to find local business information vs. 65% who turned to print yellow pages, 50% who used Internet yellow pages, and 44% who used traditional newspapers.

* 86% surveyed said they have used the Internet to find a local business, a rise from the 70% figure reported last year (2006.)

* 80% reported researching a product or service online, then making that purchase offline from a local business.

`VIEW ONLINE`

A Yahoo! study from 2007 showed:

* Online advertising drives in-store sales at a 6:1 ratio to online sales.

* Consumers in the study spent $16 offline (in stores) to every $1 spent online.

`VIEW ONLINE`

`VIEW PDF`

A study on data leaked from AOL's search query logs reveals:

* The first ranking position in the search results receives 42.25% of all click-through traffic

* The second position receives 11.94%, the third 8.47%, the fourth 6.05%, and all others are under 5%

* The first ten results received 89.71% of all click-through traffic, the next 10 results (normally listed on the second page of results) received 4.37%, the third page – 2.42%, and the fifth – 1.07%. All other pages of results received less than 1% of total search traffic clicks.

`VIEW ONLINE`

`VIEW ONLINE`

Thats Some Spicy Data You Got There!

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

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

* Search is very, very popular. It reaches nearly every online American, and billions of people around the world.

* Being listed in the first few results is 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.

* An incredible amount of offline economic activity is driven by searches on the web

"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."

3

Search Engine Optimization is the process of taking a page built by humans and making it easily consumable for both other humans and for search engine robots. This sections details some of the compromises you will need to make in order to satisfy these two very important kinds of user.

One of the most common issues we hear from folks on both the business and technology sides of a company goes something 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."

Sounds Brutal...

Initially, this argument can seem like a tough obstacle to overcome, but the more you're able to explain details and examine the inner-workings of the engines, the less powerful this argument becomes.

Limitations x3

Limitations 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, following links and indexing content in massive databases. But, modern search technology is not all-powerful. There are technical limitations of all kinds that can cause immense problems in both inclusion and rankings. We've enumerated some of the most common of these below:

1. SPIDERING AND INDEXING PROBLEMS

* *Search engines cannot fill out online forms, and thus any content contained behind them will remain hidden.*

* *Poor link structures can lead to search engines failing to reach all of the content contained on a website, or allow them to spider it, but leave it so minimally exposed that it's deemed "unimportant" by the engines' index.*

* *Web pages that use Flash, frames, Java applets, plug-in content, audio files & video have content that search engines cannot access.*

Interpreting Non-Text Content

* *Text that is not in HTML format in the parse-able code of a web page is inherently invisible to search engines.*

* *This can include text in Flash files, images, photos, video, audio & plug-in content.*

2. CONTENT TO QUERY MATCHING

* *Text that is not written in terms that users use to search in the major search engines. For example, writing about refrigerators when people actually search for "fridges". We had a client once who used the phrase "Climate Connections" to refer to Global Warming.*

* *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.*

* *Language. For example, writing content in Polish when the majority of the people who would visit your website are from Japan.*

3. THE "TREE FALLS IN A FOREST" EFFECT

This is perhaps the most important concept to grasp about the functionality of search engines & the importance of search marketers. Even when the technical details of search-engine friendly web development are correct, content can remain virtually invisible to search engines. This is due to the inherent nature of modern search technology, which rely on the aforementioned 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. The major engines **have no inherent gauge of quality or notability and no potential way to discover and make visible** fantastic pieces of writing, art or multimedia on the web. Only humans have this power – to **discover, react, comment and (most important for search engines) link. Thus, it is only natural** that great content cannot simply be created – it must be marketed. Search engines already do a **great job of promoting high quality content on popular websites or on individual web pages that have become popular, but they cannot** *generate* **this popularity – this is a task that demands talented Internet marketers.**

What was that?

THE COMPETITIVE NATURE OF SEARCH ENGINES

Take a look at any search results page and you'll find the answer to why search marketing, as a practice, has a long, healthy life ahead.
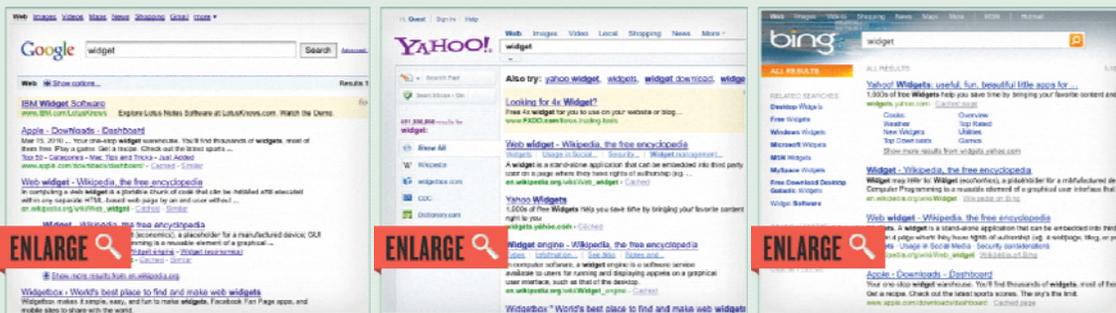
10 positions, ordered by rank, with click-through traffic based on their relative position & ability to attract searchers. The fact that so much traffic goes to so few listings for any given search means that there will always be a financial incentive for search engine rankings. No matter what variables may make up the algorithms of the future, websites and businesses will contend with one another for this traffic, branding, marketing & sales goals 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 2010, social media marketing and vertical search inclusion are mainstream methods for conducting search engine optimization.

The future may be uncertain, but in the world of search, change is a constant. For this reason, along with all the many others listed above, search marketing will remain a steadfast need in the diet of those who wish to remain competitive on the web. Others have mounted an effective defense of search engine optimization in the past, but as we see it, there's no need for a defense other than simple logic – websites and pages compete for attention and placement in the search engines, and those with the best knowledge and experience with these rankings will receive the benefits of increased traffic and visibility.

CHAPTER FOUR

THE BASICS OF SEARCH ENGINE FRIENDLY DESIGN & DEVELOPMENT

DOWNLOAD PDF

Search engines are limited in how they crawl the web and interpret content to retrieve and display in the results. In this section of the guide, we'll focus on the specific technical aspects of building (or modifying) web pages so they're optimally structured for search engines and human visitors. 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 content – the material available to visitors of your site – must be in HTML text format. Images, Flash files, Java applets, and other non-text content is virtually invisible to 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:

"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?"

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. Images can also be shown to visitors as replacement for text by using CSS styles.

3. Flash or Java plug-in contained content can be repeated in 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.
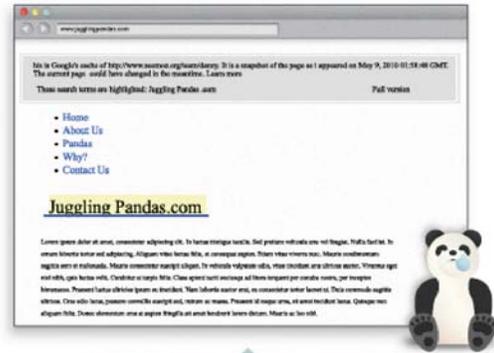
Now let's double-check some stuff

Most sites do not have significant problems with indexable content, but double-checking is worthwhile. By using tools like Google's cache, SEO-browser.com, the mozBar or Yellowpipe you can see what elements of your content are visible and indexable to the engines.

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 is simply a link to another page. This is bad because it makes it difficult to interpret relevancy.

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

What Humans See

What Search Engines See

That's alot 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.

If you're curious about exactly what terms and phrases search engine can see on a webpage, We have a nifty tool called "Term Extractor" that will display words & phrases ordered by frequency. However, it's wise to not only check for text content but to also use a tool like SEO Browser to double-check that the pages you're building are visible to the engines. It's very hard to rank if you don't even appear in the search engine keyword databases.

Crawlable Link Structures

Search engines need to see content in order to list pages in their massive keyword-based indices. They also need to have access to a crawlable link structure – one that lets their spiders browse the pathways of a website – in order to find all of the pages on a website. Hundreds of thousands of sites make the critical mistake of hiding or obsfucating 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 & D might be important pages on the site, the spider has no way to reach them (or even know they exist) 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 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 engine's 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.

Links in submission-required forms

Forms can include something as basic as a drop down menu or as complex as a full-blown survey. In either case, search spiders 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 (full description here) 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.

Links in 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.

Links only accessible through search

Although this relates directly to the above warning on forms, it's such a common problem that it bears mentioning. Spiders will not attempt to perform searches to find content, and thus, it's estimated that millions of pages are hidden behind completely 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 padnas 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 tend to only crawl about 100 links on any given page. This loose restriction is necessary to keep down on spam and conserve rankings.

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:

Lousy Punks!

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 came about as a method to help stop automated blog comment, guestbook, 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 nofollow are interpreted slightly differently by each of the engines. You can read more about the affect of this and PageRank sculpting on this blog post.

Google

nofollowed links carry no weight or impact and are interpreted as HTML text (as though the link did not exist). Google's representatives have said that they will not count those links in their link graph of the web at all.

Yahoo! & Bing

Both of these engines say that nofollowed links do not impact search results or rankings, but may be used by their crawlers as a way to discover new pages. That is to say that while they "may" follow the links, they will not count them as a method for positively impacting rankings.

Ask.com

Ask is unique in its position, claiming that nofollowed links will not be treated any differently than any other kind of link. It is Ask's public position that their algorithms (based on local, rather than global popularity) are already immune to most of the problems that nofollow is intended to solve.

Old PageRank Algorithm & Effect of Nofollow Sculpting

LINK A — link juice X/5
LINK B — link juice X/5
LINK C — link juice X/5
LINK D — link juice X/5
LINK E — link juice X/5

Page with "X" amount
of passable link juice

LINK A — link juice X/5
LINK B — No Follow
LINK C — link juice X/5
LINK D — link juice X/5
LINK E — link juice X/5

Page with "X" amount
of passable link juice

NEW PageRank Algorithm & Effect of Nofollow Sculpting

LINK A — link juice X/5
LINK B — No Follow
LINK C — link juice X/5
LINK D — link juice X/5
LINK E — link juice X/5
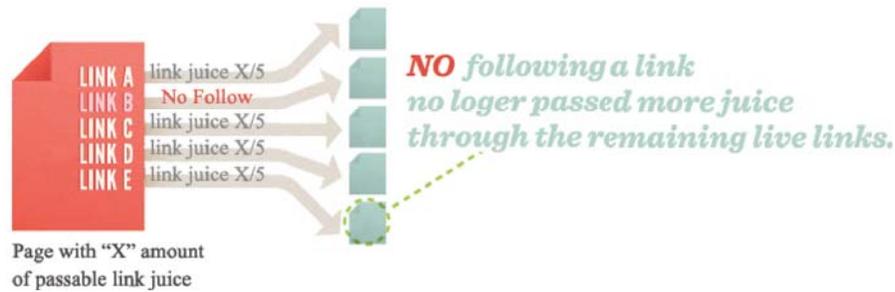
Page with "X" amount
of passable link juice

NO following a link
no loger passed more juice
through the remaining live links.

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 (which would get pretty big), 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 being listed in the search results for "dog," it's extremely wise to make sure the word "dog" is part of the indexable content of your document.

STEP 1: RECEIVE QUERY

Google

DOG

STEP 2: FIND RELEVANT DATABASE

"DOG"

http://www.thedog.com	PR 9
http://www.theplog.com	PR 7
http://www.goddog.com	PR 5
http://www.dog.com	PR 3
http://www.dogyear.com	PR 5
http://www.baddog.com	PR 2

this one

STEP 3: RANK DOCUMENTS

"DOG"

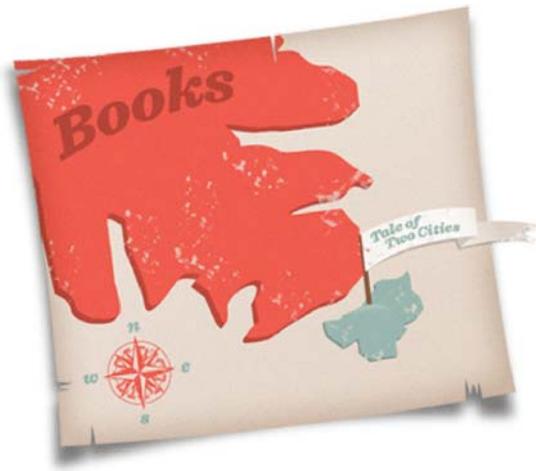|	
http://www.thedog.com	PR 9
http://www.theplog.com	PR 7
http://www.goddog.com	PR 5
http://www.dog.com	PR 3
http://www.dogyear.com	PR 5
http://www.baddog.com	PR 2

STEP 4: RETURN SEARCH RESULTS PAGE

Google

Keyword Domination

Keywords also 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 knows which pages to retrieve based on the words entered into the search box. Other data, such as the order of the words ("tanks shooting" vs. "shootingtanks"), spelling, punctuation, and capitalization of those terms provide additional information that the engines can use to help retrieve the right pages and rank them.

For obvious reasons, 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, therefore, to ensure that keywords are prominently used in titles, text, and meta data.

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 is a lot less results and thus competition for the specific result.

Keyword Density Myth

Whenever the topic of keyword usage and search engines come together, a natural tendency to use the phrase "keyword density". This is tragic. Keyword density is, without question, NOT a part of modern web search engine ranking algorithms for the simple reason that it provides far worse results than many other, more advanced methods of keyword analysis. Rather than cover this logical fallacy in depth in this guide, we'll simply reference Dr. Edel Garcia's seminal work on the topic – [The Keyword Density of Non-Sense](#).

The notion of keyword density value predates all commercial search engines and the Internet and can hardly be considered an information retrieval concept. What is worse, keyword density plays no role on how commercial search engines process text, index documents, or assign weights to terms. Why then do many optimizers still believe in keyword density values? The answer is simple: misinformation.

Dr. Garcia's background in information retrieval and his mathematical proofs should debunk any notion that keyword density can be used to help "optimize" a page for better rankings. However, this same document illustrates the unfortunate truth about keyword optimization – without access to a global index of web pages (to calculate term weight) and a representative corpus of the Internet's collected documents (to help build a semantic library), we have little chance to create formulas that would be helpful for true optimization.

On-Page Optimization

That said, keyword usage and targeting are only a small part of the search engines' ranking algorithms, and we can still leverage some effective "best practices" for keyword usage to help make pages that are very 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, and possibly twice (or as a variation) if it makes sense and sounds good (this is subjective, but necessary). 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 in the H1 header tag of the page.

Keyword Density Myth Example

If two documents, D1 and D2, 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-occurance)*
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:

- *At least 3X in the body copy on the page (sometimes a few more times if there's a lot of text content). You may find additional value in adding the keyword more than 3X, 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 bold. You can use either the <strong> or <b> tag, as search engines consider them equivalent.*

- *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 (sometimes 2X when it makes sense) 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).*

You can read more information about On-Page Optimization at this post.

The title tag of any page appears at the top of Internet browsing software, but this location has been noted to receive a relatively small amount of attention from users, making it the least important of the three.
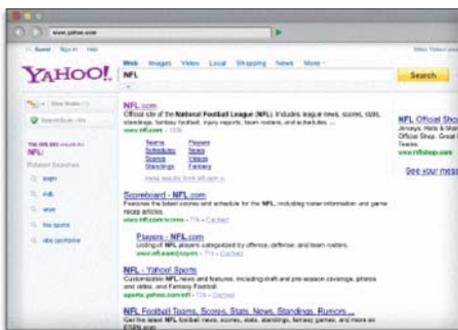
Using keywords in the title tag means that search engines will "bold" (or highlight) those terms in the search results when a user has performed a query with those terms. This helps garner a greater visibility and a higher click-through rate.

Title Tags

The title element of a page is meant to be an accurate, concise description of a page's content. It creates value in three specific areas (covered to the left) and is critical to both user experience and search engine optimization.

As title tags are such an important part of search engine optimization, 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

70 characters is the maximum amount that will display in the search results (the engines will show an ellipsis – "…" to indicate when a title tag has been cut off), and 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 (at least, according to SEOmoz's testing and experience).

Leverage branding

At SEOmoz, we love to start 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. Many SEO firms recommend using the brand name at the end of a title tag instead, and there are times when this can be a better approach – think about what matters to your site (or your client's site) and how strong the brand is.

Consider readability and emotional impact

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 final important reason to create descriptive, keyword-laden title tags is for ranking at the search engines. The above screenshot comes from SEOmoz's survey of 37 influential thought leaders and practitioners in the SEO industry on the search engine ranking factors. In that survey, 35 of the 37 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. Each 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.*

✴ *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 is a specialized tag telling the engines not to grab a descriptive snippet about a page from the Open Directory Project (DMOZ) for display in the search results.*

✴ *NoYDir, like NoODP, is specific to Yahoo!, informing that engine not to use the Yahoo! Directory description of a page/site in the search results*

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 (as in the Balboa Park example to the right), so it's generally wise to stay in these limits.

META DESCRIPTION TAG

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 in disuse, read Meta Keywords Tag 101 from SearchEngineLand.

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

Although these tags can have uses for search engine optimization, they are less critical to the process, and so I'll leave them to John Mueller of Google's Webmaster Central division to answer in greater detail – Meta Tags & Web Search.

URL Structures

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

Above, the green text shows the url for SEOmoz's Web 2.0 awards. Since search engines display URLs in the results, they can impact clickthrough 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(by 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 (from email clients, search engines, and even people!).

Go static

Go static

With technologies like mod_rewrite for Apache and ISAPI_rewrite for Microsoft, there's no excuse not to create simple, static URLs. Even single dynamic parameters in a URL can result in lower overall ranking and indexing (SEOmoz itself switched from dynamic URLs – e.g. www.seomoz-.org/blog?id=123, to static URLS – e.g. www.seomoz.org/blog/11-best-practices-for-urls, in 2007 and saw a 15% rise in search traffic over the following 6 weeks).

Choose descriptives whenever possible

Rather than selecting numbers or meaningless figures to categorize information, use real words. For example, a URL like www.thestore.com/hardware/screwdrivers is far more usable and valuable than www.thestore.com/cat33/item4326.

Use hyphens to separate words

Not all of the search engines accurately interpret separators like underscore "_," plus "+," or space "%20," so use the hyphen "-" character to separate words in a URL, as in the SEOmoz 11 Best Practices for URLs example above.

Canonical and Duplicate Versions of Content

META DESCRIPTION

Canonicalization can be a challenging concept to understand (and hard to pronounce – "ca-non-ick-cull-eye-zay-shun"), but it's essential to creating an optimized website. The fundamental problems stem from multiple uses for a single piece of writing – a paragraph or, more often, an **entire page of content will appear in multiple locations on a website, or even on multiple websites.** For search engines, this presents a conundrum – which version of this content should they show **to searchers? In SEO circles, this issue 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 (or best).

Canonicalization is the practice of organizing your content in such a way that every unique piece has one and only one URL. By following this process, you can ensure that the search engines will find a singular version of your content and assign it the highest achievable rankings based on your domain strength, trust, relevance, and other factors. If you leave multiple versions of content on a website (or websites), you might end up with a scenario like that to the right.

If, instead, 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 sigan overall. This will positively impact their ability to rank well in the search engines.

You say you want another option though?

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 from one URL on one domain to a different URL on a different domain.)**

The tag is part of the HTML header on a web page, the same section you'd find the Title element and Meta Description **tag. This simply uses a new rel parameter.**

THE INNER WORKINGS

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

This would tell Yahoo!, Bing & Google that the page in question should be treated as though it were a copy of the URL www.seomoz.org/blog and that all of the link & content metrics the engines apply should technically flow back to that URL.

DISPLAY ONLY

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).

How we do it

SEOmoz has worked on several campaigns where two versions of every content page existed in both a standard, web version and a print-friendly version. In one instance, the publisher's own site linked to both versions, and many external links pointed to both as well (this is a common phenomenon, as bloggers & social media types like to link to print-friendly versions to avoid advertising). We worked to individually 301 re-direct all of the print-friendly versions of the content back to the originals and created a CSS option to show the page in printer-friendly format (on the same URL). This resulted in a boost of more than 20% in search engine traffic within 60 days. Not bad for a project that only required an hour to identify and a few clever rules in the htaccess file to fix.

Easy as pie!

Defending Your Sites Honor

How scrapers like 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 of 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 that relative
links in your internal linking structure. Thus, rather than linking to
your home page using:

Home

You would instead use:

Home

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, and for
WordPress users Joost de Valk has a useful plugin, 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

CHAPTER FIVE KEYWORD RESEARCH

**Keyword research is one of the most important, valuable, and high return
activities in the search marketing field.** 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.** 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 –
not taking advantage is practically criminal.

How to Judge the Value of a Keyword

Every search phrase that's typed into an engine is recorded in one way or another, and keyword research tools like those described below allow us to retrieve this information. However, those tools cannot show us (directly) how valuable or important it might be to rank for and receive traffic from those searches. To understand the value of a keyword, we need to research further, make some hypotheses, test, and iterate – the classic web marketing formula.

The following is a basic, but valuable process for determining a keyword's value:

Ask yourself

Is the keyword relevant to the content your website offers? Will searchers who find your site through this term find the likely answer to their implied question(s)? And will this traffic result in financial rewards (or other organizational goals) directly or indirectly? If the answer to all of these questions is a clear "Yes!," proceed…

Search for the term/phrase in the major engines

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

In Google Adwords, Choose "exact match" and point the traffic to the most relevant page on your website. Measure the traffic to your site, and track impressions and conversion rate over the course of at least 2-300 clicks (this may take only a day or two with highly trafficked terms, or several weeks with keyword in lesser demand).

Using the data you've collected, make an educated guess about the value of a single visitor to your site with the given search term or phrase.

For example, if, in the past 24 hours, your search ad has 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 approx. $3 to your business. Those 5,000 impressions in 24 hours could probably generate a click-through rate of between 30-40% with a #1 ranking (see the leaked AOL data mining for more on potential click-through-rates), which would mean 1500-2000 visits per day, at $3 each, or ~$1.75 million dollars per year. No wonder businesses love search marketing!

Of course, even the best estimates of value fall flat against the hands-on process of optimizing and calculating ROI. Remember that the time and money you invest in a search marketing campaign must be weighed against any returns, and 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

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 overall searches performed on the web. The remaining 70% lie in what's commonly 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 (or even only once, ever!), but, when taken together, they comprise the majority of the world's demand for information through search engines.

Understanding the search demand curve is critical, because it stresses the importance of "long-tail" targeted content – pages with information not directed at any particular single, popular query, but rather simply exposing the myriad of human thought, research, and opinion to the spiders of the search engines. As an example, below we've included a sample keyword demand curve, illustrating the small number of queries sending larger amounts of traffic alongside the plethora of rarely-searched terms and phrases that bring the bulk of our search referrals:
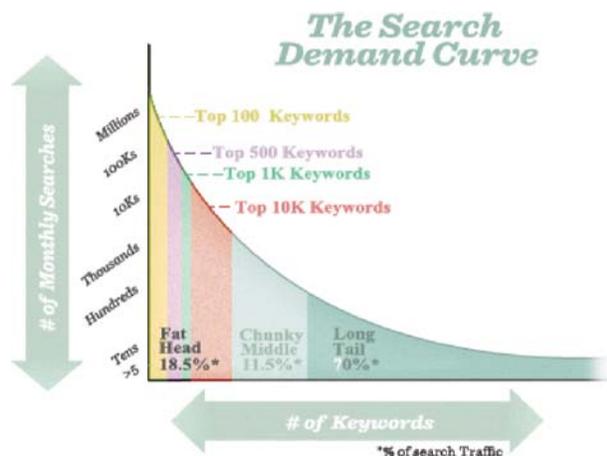
"Ignore the long tail at your peril - search marketing and web site content strategies must allow for this "impossible to predict" form of visits or risk losing out ot a more expository and prolific competitor."

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 Estimator*

✳ *Google Insights for Search*

✳ *Google Trends Keyword Demand Prediction*

✳ *Microsoft AdCenter Keyword Forecasting*

✳ *Wordtracker's Free Basic Keyword Demand*
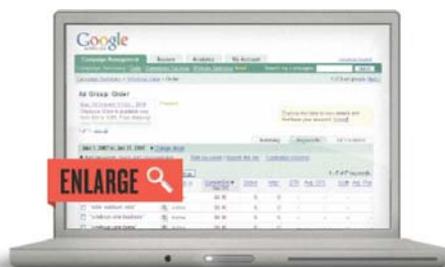
We can see that Google is predicting both the cost of running campaigns for these terms as well as estimates of the number of clicks a campaign might receive. You can use these latter numbers (under the "estimated clicks/day" column) to get a rough idea of how popular a particular keyword or phrase is in comparison to another.

Here's an example of mining keyword research data from Google's adWords Estimator Tool.

Keyword Difficulty

I've got a lock on a bogey!

In order to know which keywords to target now (and which to pursue later), it's essential to not only understand the demand for a given term or phrase, but the work required to achieve those rankings. If mighty competitors block the top 10 results and you're just starting out on the web, the uphill battle for rankings can take months or years of effort bearing little to no fruit. This is why it's essential to understand keyword difficulty.

ENLARGE
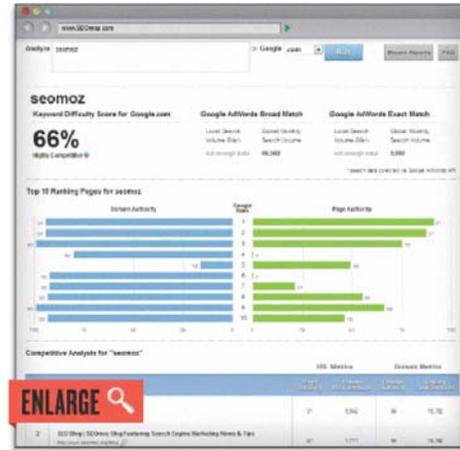
Of course, if you'd like to save time, SEOmoz's own Keyword Difficulty Tool does a good job collecting all of these metrics and providing a comparative score for any given search term or phrase.

CHAPTER SIX

HOW USABILITY, USER EXPERIENCE & CONTENT AFFECT SEARCH ENGINE RANKINGS

DOWNLOAD PDF

The search engines are in a constant quest to improve their results 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:

HOW & WHY GREAT SITES RISE TO THE TOP OF SEARCH ENGINE RANKINGS

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.**

Search engines can't understand text, view images, or watch video the same way a human being can, and thus they rely on meta information (not necessary meta tags) about sites and pages in order to rank content. 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.
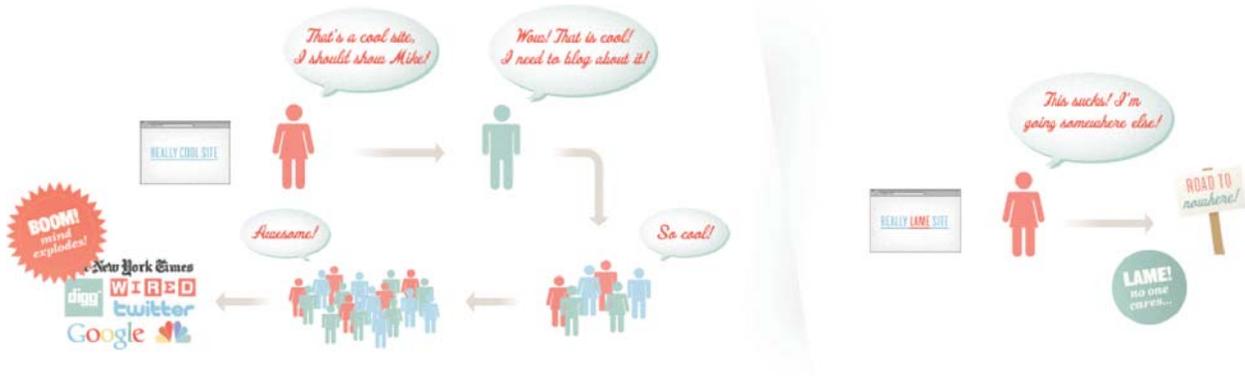
All of that positive attention and excitement around the content offered by the new site translates into a machine parseable (and algorithmically valuable) collection of links. The timing, source, anchor text, and number of links to the new site are all factored into its potential performance (i.e., 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.

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 keyword placement, links, and site structure. However, through linking patterns, 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.

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.

Crafting Content

for Search Engine Success

Developing "great content" may be the most oft-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...

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.

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.

COMMERCIAL INVESTIGATION

Researching sources for a story, uncovering potential clients/partners, acquiring competitive intelligence, discovering options for future transactions.

A commercial investigation search straddles the line between pure research and commercial intent. For example, sourcing potential partners for distribution of your new t-shirts in Albuquerque, determining what companies make laptop bags for sale in the United Kingdom, or researching the best brand of digital cameras for an upcoming purchase all qualify. They're not directly transactional, and may never result in an exchange of goods, services, or monies, but they're not purely informational either.

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.

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 find their queries fulfilled and reward that positive experience through 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 link analysis, the engines can discover how pages are related to other pages and in what ways. Since the late 1990's links have also served as a stand-in for votes – representing the democracy of the web's opinion about what pages are important and popular. (Some refer to this as the reasonable surfer model). The engines themselves have refined the use of link data to a fine art, and incredibly sophisticated algorithms create nuanced evaluations of site and pages based on this information.

Professional SEOs attribute a considerable portion of the search engines' algorithms to link-based factors (see Search Engine Ranking Factors). Through links, engines analyze the popularity of a site & page based on the number and popularity of pages linking to them, as well as metrics like trust, spam, & 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 in a given space.

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. We don't know all the attributes measured by the engines, but through analysis of patent applications, papers submitted to information retrieval conferences, and hands-on experience & testing, we can draw some intelligent assumptions. Below is a list of notable factors worthy of consideration. All of these issues, 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 to the search engines. Getting lots of local, topic-specific links is great, too, but to earn trust and authority with the engines, you'll need the help of some powerful link partners.

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 matters more than links from general or off-topic sites.

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" and "leave," where many results rank solely due to the anchor text of inbound links.

TrustRank

In order to weed out massive amounts of spam (some estimate as much as 60% of the web's pages are spam), search engines use systems for measuring trust, many of which are based on the link graph. Earning links from highly trusted domains can, thus, result in a significant boost to this scoring metric.

Link Neighborhood

In many papers on spam detection and information retrieval, using the sites that link to a domain as well as the sites that domain links to has an impressive knack for spam detection and filtering. 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.

THE CONCEPT OF TRUSTRANK

0% Spam	0.15% Spam	4% Spam	14% Spam
Trusted Seed Set	1 Link "Hop" Away	2 Link "Hops" Away	3 Link "Hops" Away
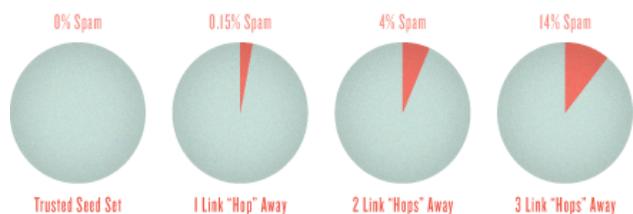
Search researchers found that the further a site/page's distance from a trusted seed set, the more likely they were to contain spam results.

Link Building Basics

Building links is an art. It's almost certainly the most challenging part of an SEO's job, and, for many sites, the one most critical to achieving long term success. Many companies can afford to hire SEOs to help make their websites search friendly and search optimized, but a robust backlink profile is an extremely high barrier to competition.
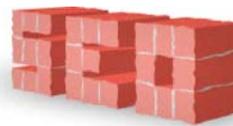
1 *Editorial Accumulation*

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

2 *Manual Suggestion & Approval*

Emailing bloggers with links, submitting sites to directories, or paying for listings of any kind fit into this group. The SEO must create a value proposition with the link target and complete that transaction manually (whether it be filling our 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 guestbook signings, forum signatures, blog comments, or user profiles. These links are typically quite low in value, but can, in aggregate, have a significant impact. However, automatic methods of generating these links is certainly spamming, and even the manual creation of such links is frowned upon by many site owners and search engines. Exceptions abound, and for those sites that offer these options and don't use the rel="nofollow" attribute on outbound links, there can be opportunity.

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 process undertaken in a link building campaign must be the creation of goals and strategies. Sadly, link building is one of the most difficult activities to measure, particularly from a search engine optimization perspective. Although the engines internally weight each link with precise, mathematical metrics, it's impossible for those outside of the engineering teams at these companies to extract this data. Thus, as SEOs, we rely on a number of signals to help build a rating scale of link value. Along with the less-measurable data from the link signals mentioned above, these metrics include the following:

Page 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). Pages that rank well for relevant queries tend to be more valuable than those that don't.

Google PageRank

Despite much maligning over the years for accuracy and freshness problems (Google only updates their toolbar PageRank data every 3-5 months and sometimes manipulates the values intentionally to discourage spam and over-analysis), there is still value to looking at the number reported. This is discussed more in this blog post on PageRank Correlation. Pages with high PageRank do tend to pass on more link value than those with little or none. Be careful with those that have PageRank "unranked" (a gray bar) as these may be highly valuable pages that simply haven't received visible PageRank since the last update.

SEOmoz 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.

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.

SEOmoz Domain Authority

Domain authority (or DA) is a query independent measure of how likely a domain is likely 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.

Google blogsearch

Google Blog Search is the only property controlled by the search giant that offers accurate backlink information. While this only shows links from blogs and feeds, there's still great value in seeing which sites/pages have earned authority and attention in the blogosphere, as this can be a useful predictor of the link value they'll pass.

Yahoo! Site Explorer Reported Inlinks

Yahoo! Site Explorer is a valuable tool for seeing the links that point to a given site or page. Using this tool, you can make estimates about the relative link popularity and importance a page has based on who links to it. Typically those pages/sites with more powerful and important links will pass on greater value through their links.

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 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 stats, but if you can't get access to these, services like Google Trends for Websites, Compete, Quantcast, & Alexa can give you a rough idea of at least domain-wide traffic, from which you can estimate page-specific popularity.

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. Increases in search traffic when accompanied by more frequent search engine crawling and increases in referring link traffic correlates with a well-managed, intelligently structured campaign. If you see traffic from engines like Bing and Yahoo! are rising while Google stays constant, it's possible that you need to seek more authoritative, better trusted links (as Google is the most "picky" of the engines when it comes to link evaluation).

5 Samples of Link Building Strategies

Many sites offer directories or listings of relevant resources

✳ *You can find hundreds 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).*

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 Microsoft often does with their partner 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 link requests in order confirmation emails for more.*

Build a company blog and 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 this Peak Season Ingredient Map from Epicurious, this Interactive Graphic Explaining Hand Signals Used on Stock Market Trading Floors from the New York Times, or this Video of an iPod in a Blender from Blendtec. 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.*

Build content that can be shared through a citation-based licensing agreement

✳ *If you have photos, videos, graphics, charts, raw data, or text content that can be licensed out with a system like Creative Commons' Attribution (or Attribution-ShareAlike), you can leverage the power of the web's penchant for information sharing while receiving links back to your originals and your site each time someone uses your material.*

> *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, but 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 Professional's Guide to Link Building for a more comprehensive overview).*
>
> *Search for sites like yours in the search engines by using keywords and phrases directly relevant to your business. When you locate sites that aren't directly competitive, you can 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 email link requests for more detail.*

Show Me the Money

An Aside on Buying Links

Google, Yahoo!, and Bing all 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. This includes sending anonymous representatives to search conferences and joining link networks so they can see who else is involved.

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 at this blog post.

CHAPTER EIGHT 8

SEARCH ENGINE TOOLS AND SERVICES

DOWNLOAD PDF

To encourage webmasters to create sites and content in accessible ways, each of the major search engines have built support and guidance-focused services. Each provides varying levels of value to search marketers, but all of them are worthy of understanding. These tools provide data points and opportunities for exchanging information with the engines that are not provided anywhere else.

The sections below explain the common interactive elements that each of the major search engines support and identify why they are useful. There are enough details on each of these elements to warrant their own articles, but for the purposes of this guide, only the most crucial and valuable components will be discussed.

Common Search Engine Protocols

1. SITEMAPS

Sitemaps are a tool that enable you to give hints to the search engines on how they can crawl your website. You can read the full details of the protocols at Sitemaps.org. In addition, you can build your own sitemaps at 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, files sizes can get very large.*

RSS

Really Simple Syndication or Rich Site Summary

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.*

Txt

Text File

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) should be stored in a website's root directory (e.g., www.google.com/robots.txt). The file serves as an access guide for automated visitors (web robots). By using robots.txt, webmasters can indicate which areas of a site they would like to disallow bots from crawling as well as indicate the locations of sitemaps files (discussed below) 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: It is very important to realize that not all web robots follow robots.txt. People with bad intentions (e.g., 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="ROBOT NAME" content="ARGUMENTS" />
 </head>
 <body>
  <h1>Hello World</h1>
 </body>
</html>
```

ROBOTS META TAG

<meta name = "ROBOT NAME" content = "ARGUMENTS" />

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

noindex	Google, Yahoo, Live, Ask	*Page not indexed*
nofollow	Google, Yahoo, Live, Ask	*All links on page become no followed*
noarchive	Google, Yahoo, Live, Ask	*Page not cached*
noodp	Google, Yahoo, Live	*Stops description and title tag overwrite by DMOZ (only for Homepage)*
noydir	Yahoo	*Stops description and title tag overwrite by Yahoo Directory*
nosnippet	Google	*Stops Google from generating description based on on-page text*

In the example above, "ROBOT NAME" is the user-agent of a specific web robot (ex. Googlebot) or an asterisk to identify all robots, and "ARGUMENTS" is one arguments listed in the diagram to the right.

4. REL="NOFOLLOW"

The rel=nofollow attribute creates link-level instructions for search engine bots that suggest how the given link should be treated. While the search engines claim to not nofollow links, tests show that they actually do follow them for discovery new pages. These links certainly pass less juice (and in most cases no juice) than their non-nofollowed counterparts and as such are still a recommend for SEO purposes.

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.

Search Engine Tools

GOOGLE WEBMASTERS TOOLS

Google Webmasters 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 our 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

ENLARGE

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.

Image Search – If a webmaster chooses to opt in to enhanced image search, Google may use tools such as [Google Image Labeler](#) to associate the images included in their site with labels that will improve indexing and search quality of those images.

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.

Diagnostics

Web Crawl – Web Crawl identifies problems Googlebot encountered when it crawls a given website. Specifically, it lists Sitemap errors, HTTP errors, nofollowed URLs, URLs restircted by robots.txt and URLs that time out.

Mobile Crawl – Identifies problems with mobile versions of websites.

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

Sign Up

BING WEBMASTER CENTER

[Bing Webmaster Center](#)

Features

Profile – This interface provides a way for webmasters to specify the location of sitemaps and a form to provide contact information so Bing can contact them if it encounters problems while crawling their website.

Crawl Issues – This helpful section identifies HTTP status code errors, Robots.txt problems, long dynamic URLs, unsupported content type and, most importantly, pages infected with malware.

Backlinks – This section allows webmasters to find out which webpages (including their own) are linking to a given website.

Outbound Links – Similarly to the aforementioned section, this interface allows webmasters to view all outbound pages on a given webpage.

Keywords – This section allows webmasters to discover which of their webpages are deemed relevant to specific queries.

Sitemaps – This is the interface for submitting and managing sitemaps directly to Microsoft.

Statistics

These statistics are a window into how Google sees a given website. Specifically, it identifies top search queries, crawl stats, subscriber stats, "What Googlebot sees" and Index stats.

Link Data

This section provides details on links. Specifically, it outlines, external links, internal links and sitelinks. Sitelinks are section links that sometimes appear under websites when they are especially applicable to a given query.

Sitemaps

This is the interface for submitting and managing sitemaps directly with Google

YAHOO! SITE EXPLORER

[Yahoo! Site Explorer](#)

Features

Statistics – These statistics are very basic and include data like the title tag of a homepage and number of indexed pages for the given site.

Feeds – This interface provides a way to directly submit feeds to Yahoo! for inclusion into its index. This is mostly useful for websites with frequently updated blogs.

Actions – This simplistic interface allows webmasters to delete URLs from Yahoo's index and to specify dynamic URLs. The latter is especially important because Yahoo! traditionally has a lot of difficulty differentiating dynamic URLs.

Sign Up

ENLARGE

SEOMOZ OPEN SITE EXPLORER

While not run by the search engines, SEOmoz's Open Site Explorer does provide 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 show 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.

For more information, click below:

Learn More

It is a relatively recent occurrence that search engines have started to provide tools that allow webmasters to interact with 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 webmaster. It is true today, and will likely be true in the future that the ultimate responsibility of SEO is on the marketers and webmasters. It is for this reason that learning SEO is so important.

CHAPTER NINE
MYTHS & MISCONCEPTIONS ABOUT SEARCH ENGINES

DOWNLOAD PDF

Unfortunately, over the past 12 years, a great number of misconceptions have emerged about how the search engines operate and what's required to perform effectively. In this section, we'll cover the most common of these, and 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 information (this would sometimes even include the keywords they wanted to rank for), and "submit" them to the engines, after which a bot would crawl and include those resources in their index. For obvious reasons (manipulation, reliance on submitters, etc.), this practice was unscalable and eventually gave way to purely crawl-based engines. Since 2001, search engine submission has not only not been required, but is actually virtually useless. The engines have all publicly noted that they rarely use the "submission" URL lists, and that the best practice is to earn links from other sites, as this will expose the engines to your content naturally.

You can still see submission pages (for Yahoo!, Google, 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 did use the submission service to crawl your site, you'd be very 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 today, only Yahoo! among the major engines will even index content from the meta keywords tag, and even they claim not to use those terms for ranking, but merely content discovery.

It is true that other meta tags, namely the title tag and meta description tag (which we've covered previously in this guide), are of critical importance to SEO best practices. And, certainly, 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

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 and should therefore be a focus of SEO efforts. Despite being proven untrue time and again, this farce has legs, and indeed, many SEO tools 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 – that upping your PPC spend will improve your organic SEO rankings (or, likewise, that lowering that spend can cause ranking drops). 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 of 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.

PERSONALIZATION

Personalization seems to primarily affect areas in which we devote tons of time, energy and repeated queries. This means for many/most "discovery" and early funnel searches, we're going to get very standardized search results. It's true that it can influence some searches significantly, but it's also true that, 90%+ of queries we perform are unaffected (and that goes for what we hear from other SEOs, too). This post helps to validate this, showing that while rankings changes can be dramatic, they only happen when there's substantive query volume from a user around a specific topic.

RECIPROCAL LINKS

Reciprocal links are of dubious value: they are easy for an algorithm to catch and to discount. Having your own version of the Yahoo directory on your site isn't helping your users, nor is it helping your SEO.

We wouldn't be concerned at all with a technically "reciprocated" link, but we would watch out for schemes and directories that leverage this logic to earn their own links and promise value back to your site in exchange. Also, watch out for those who've evolved to build "three-way" or "four-way" reciprocal directories such that you link to them and they'll link to you from a separate site – it's still attempted manipulation and there's so many relevant directories out there; why bother!?

Search Engine 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.

First,

Search engines have learned that users hate spam. This may seem a trivial and obvious lesson, but in fact, many who study the field of search from a macro perspective believe that along with improved relevancy, Google's greatest product advantage over the last 10 years has been their ability to control and remove spam better than their competitors. While it's hard to say if this directly influenced their dramatic rise to lead in market share worldwide, it's undoubtedly something all the engines spend a great deal of time, effort and resources on – and with hundreds of the world's smartest engineers dedicated to fighting the practice, those of us at SEOmoz loathe to ever recommend search spam as a winnable endeavor in the long term.

Second

Search engines have done a remarkable job identifying scalable, intelligent methodologies for fighting manipulation and making it dramatically more difficult to adversely impact their intended algorithms. Concepts like [TrustRank](#) (which SEOmoz's Linkscape index leverages), HITS, statistical analysis, historical data and more, along with specific implementations like the Google Sandbox, penalties for directories, reduction of value for paid links, combating footer links, etc. 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.

This guide is not intended to show off specific spam tactics (either those that no longer work or are still practiced), 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](#), [Yahoo!'s Search Content Quality Guidelines](#) & [Bing's Guidelines for Successful Indexing](#).

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 numerous 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 (or keyword density) appear to 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.*

* *Incestuous or self-referential links, 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 and this post from Search Engine Land on the official word from Google & other engines).*

* *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 of reducing their impact. As new spam systems (like this new reciprocal link cloaking scheme uncovered by Avvo Marketing Manager Conrad Saam) emerge, engineers will continue to fight them with targeted algorithms, human reviews and the collection of spam reports from webmasters & SEOs.

CLOAKING

A basic tenant 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. 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 risks 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 guidelines and methodologies 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 affiliate content (pages whose material is used on dozens or hundreds of other sites promoting the same product/service), duplicate content (pages whose content is a copy of or extremely similar to other pages already in the index), and dynamically generated content pages that provide very little unique text or value (this frequently occurs on pages where the same products/services are described for many different geographies with little content segmentation). The engines are generally 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.

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 signals. 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, What Makes a Good Directory and Why Google Penalized Dozens of Bad Ones, Google's Sandbox Still Exists: Exemplified by Grader.com, and How to Handle a Google Penalty – And, an Example from the Field of Real Estate.

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 (see these videos on Using Trust Rank to Guide Your Link Building and How the Link Graph Works for more). Thus, if you publish low quality, duplicate content on your personal blog, then buy several links from spammy directories, you're 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 methodology for the engines to employ in considering new domains and analyzing the activities of a site. 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, LII.org, Cornell.edu, and similarly reputable players. On the flip side, if you have yet to earn high quality links, judgments may be far stricter from an algorithmic view.

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 take algorithmic and manual review methods to prevent this.

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:

STEP I — RULE OUT

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.
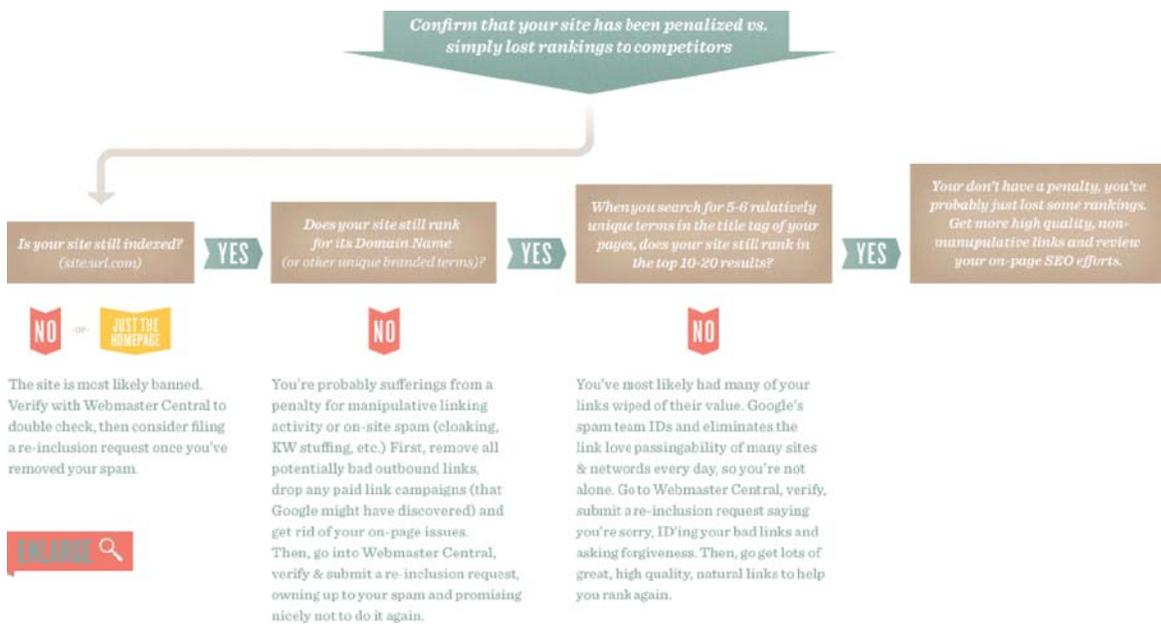
Changes

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.

Confirm that your site has been penalized vs. simply lost rankings to competitors

| Is your site still indexed? (site:url.com) | **YES** | Does your site still rank for its Domain Name (or other unique branded terms)? | **YES** | When you search for 5-6 relatively unique terms in the title tag of your pages, does your site still rank in the top 10-20 results? | **YES** | Your don't have a penalty, you've probably just lost some rankings. Get more high quality, non-manipulative links and review your on-page SEO efforts. |

NO — **JUST THE HOMEPAGE**

The site is most likely banned. Verify with Webmaster Central to double check, then consider filing a re-inclusion request once you've removed your spam.

ENLARGE 🔍

NO

You're probably suffering from a penalty for manipulative linking activity or on-site spam (cloaking, KW stuffing, etc.) First, remove all potentially bad outbound links, drop any paid link campaigns (that Google might have discovered) and get rid of your on-page issues. Then, go into Webmaster Central, verify & submit a re-inclusion request, owning up to your spam and promising nicely not to do it again.

NO

You've most likely had many of your links wiped of their value. Google's spam team IDs and eliminates the link love passingability of many sites & networks every day, so you're not alone. Go to Webmaster Central, verify, submit a re-inclusion request saying you're sorry, ID'ing your bad links and asking forgiveness. Then, go get lots of great, high quality, natural links to help you rank again.

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, Yahoo!'s, Bing's). This registration creates an additional layer of trust and connection between your site and the webmaster teams.*

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).*

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.*

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 (at least, so far), they have the right to include or reject any site/page for any reason (or no reason at all). Inclusion is a privilege, not a right, so be cautious and don't apply techniques you're unsure or skeptical of

CHAPTER TEN — 10 — MEASURING & TRACKING SUCCESS

DOWNLOAD PDF

That which can be measured can be improved, and in search engine optimization, measurement is critical to success. Professional SEOs track data about rankings, referrals, links and more to help analyze their campaigns 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 in appeal. Note that we're only covering those metrics critical to SEO – optimizing for the search engines – and as such, more general but still important metrics may not be included. For a more comprehensive look at web analytics overall, 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. Broadly, these include:

* *Direct Navigation (type 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 Engines (queries that sent traffic from any major or minor web search engine)*

Knowing the percentage and exact numbers will help you identify strengths and weaknesses and serve as a comparison over time for trend data. If, for example, you see that traffic has spiked dramatically but it comes from referral links with low relevance while search engine and direct type-ins fell, you'll know you're actually in much more trouble than the raw numbers would suggest. You should use this data to track your marketing efforts and to serve as a broad measurement for your traffic acquisition efforts.

MONTH	VISITS
January 2009	340,305
February 2009	391,978
March 2009	219,980
Total	**952,263**

TOTAL NUMBER OF VISITS TO SEOMOZ
(January 1 - March 17, 2009)

SEARCH ENGINES	VISITS
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	140
Subtotal	**237,514**
Total	**238,286**

NUMBER OF SEARCH ENGINE VISITS
(January 1 - March 17, 2009)

238,286 ÷ 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 (Yahoo!, Bing & Google), and for most countries outside the US (with the notable exceptions of Russia, China, Japan, Korea & the Czech Republic) 80%+ of search traffic comes solely from Google. 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 arenas like cooking, sports or real estate (where the percentages might be closer to the estimates from firms like Comscore).

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 (or that are failing to succeed with others), you'll better know how to focus your efforts.

If you find your site underperforming at one of the engines (based on broad market share numbers), don't immediately panic. Remember that search engines have demographics and biases just like any other referral source. For example, in the US, Google's market share is supposedly between 65-70%, yet the vast majority of sites we've ever worked with (and those reported by our friends and colleagues in the search marketing industry) show that 80-85% of traffic share from Google is actually far more common.

A number of theories exist to support why this happens:

Don't just rely on Comscore, Hitwise or Compete.com data to tell you what percentage of share an engine should provide – make sure to investigate. You can do this by running PPC ads on the various engines (and comparing impression data), checking rankings across the engines (if your Yahoo! rankings are just as good or better than your Google rankings, it's not missed opportunity, it's lower volume), and making sure you haven't made any dumb mistakes (blocking other engines' spiders, using the meta robots NOODP to control listings at Google, but forgetting to use NOYDIR at Yahoo!, etc.).

SEARCH ENGINES	VISITS
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
Subtotal	237,514
Total	238,286

SEARCH ENGINE SHARE
(January 1 - March 17, 2009)

91.79% GOOGLE
3.45% YAHOO!
2.14% MSN/LIVE
1.72% ALL OTHERS

1. *Yahoo!'s top queries are navigational (their number one query is Google, for example), while Google's queries are more informational*

2. *Many experts believe (and some have private data to suggest that) Yahoo! has a preference for sites participating in their paid inclusion program*

3. *Yahoo! refers a large amount of traffic to Yahoo!'s own properties (Google, meanwhile, seems to have a similar love affair with Wikipedia)*

3. VISITS REFERRED BY SPECIFIC SEARCH ENGINE TERMS AND PHRASES

The terms & phrases 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 you're potentially under-serving (e.g., you rank well and get visits, but don't have content that helps the searcher accomplish their goal).

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%
Subtotal	**25,587**	**12.48%**
Total	**205,084**	**100.00%**

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. However, analytics often misstates the impact of conversion rates from the last referral, clouding the true picture of what brought a visitor who "converted." For example, in the graphic to the right, 4.46% of visitors who reached SEOmoz with the query "check backlinks" signed up to become members during that visit. What we don't know (at least, from this simple analysis), is how many of those visitors had already signed up, how many signed up during a later visit, or even what percentage of those visits were first-time visitors.

The real value from this sort of simplistic tracking comes from the "low-hanging fruit" – seeing terms/phrases that continually send visitors who convert and increasing focus on both rankings and traffic from that keyword referral as well as 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	562	95	16.90%
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%
linkscape	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 (how many pages the engines are keeping in their indices from our site), and, more importantly, watch trends over time. 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 also 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
http://www.seomoz.org/users/view/100932	1
http://www.seomoz.org/ga/view/899	1
http://www.seomoz.org/ga/view/116	1
http://www.seomoz.org/ga/view/940	1
http://seomoz.org/marketplace/companies/view/804	1
http://seomoz.org/marketplace/companies/view/806	1
http://www.seomoz.org/cat/view/Other+Verticals+in+Search	1
http://www.seomoz.org/cat/view/Blogging+Trends	1
http://www.seomoz.org/ga/view/267	1

Additional Sources:

✳ **SEO Metrics that matter** – *(from Stephan Spencer of NetConcepts)*

✳ **Advanced Google Analytics Tips for SEO, Part I, II, and III** *(from Huomah)*

Analytics Software
The Right Tools for the Job

✳ **Omniture**

✳ **Fireclick**

✳ **Mint**

✳ **Sawmill Analytics**

✳ **Clicktale**

✳ **Enquisite**

✳ **Coremetrics**

✳ **Lyris / Clicktracks**

✳ **Unica Affinium NetInsight**

✳ **Yahoo! Web Analytics**
   (formerly Indextools)

✳ **Google Analytics**

✳ **Clicky Web Analytics**

✳ **Piwik Open Source Analysis**

✳ **Woopra Website Tracking**

✳ **AWStates**

While choosing can be tough, at the time of publication, our top recommendation is for Google Analytics (so long as you have few privacy concerns and don't mind the brief data delays), followed closely by Clicky. Yahoo! Web Analytics also has a solution worth considering. If you cannot use tracking code on your web pages and need a log-file based solution, AWStats is our top recommendation, though any log file based tracking will suffer from the inability to track clickstream paths, first time vs. referring and other important metrics as accurately as cookie/session based software.

Additional Sources:

✳ **How to choose a Web Analytics Solution** – *From Bryan Eisenberg way back in 2003 (but still a relevant and quality piece)*

✳ **A Complete Guide to Web Analytics Solutions** – *From ConversionRater.com in 2006 (some data, such as seperation of Omniture vs. WebSideStory, is less relevant today)*

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).
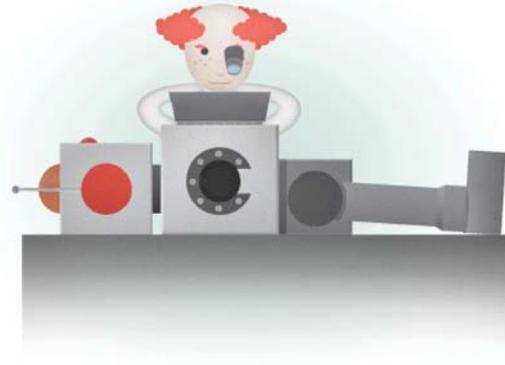
Google Site Query

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, its is still a good rough measurement. You can read more about this on 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.

Yahoo! Site Query

e.g., site:seomoz.org – note that a standard site query will automatically redirect to Yahoo!'s Site Explorer, but advanced queries that include additional parameters such as site:seomoz.org inurl:rand will show Yahoo!'s standard results format. You can use these much in the same way as the Google site query to see the number and list of pages Yahoo! has in their index for a particular site.

Yahoo! Link & Linkdomain Queries

e.g., linkdomain:seomoz.org – as with site queries, these will redirect to Yahoo! Site Explorer unless additional parameters are employed. For example, to see only links to SEOmoz.org that have the word "google" in the title tag, you'd use the query – linkdomain:seomoz.org intitle:google. Yahoo!'s link queries are the most robust and accurate of the major engines, but do include nofollow links (and don't separately mark these, which can cause trouble separate value-passing links).

Bing Site Query

e.g., site:seomoz.org – just like Yahoo! & 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 AdCenter Labs

Available at adlab.microsoft.com/alltools.aspx – 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.

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 on this blog post.

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 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 than 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

Domain Specific Metrics

Domain Authority – Domain Authority predicts how well a web page will rank on a domain. 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 form 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 to a page on 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 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

homepage (http://www.nasa.gov) in both the page content and the footer, this would still be counted as only a single link.

# 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.

the domain's – or example, if http://www.nasa.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 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.
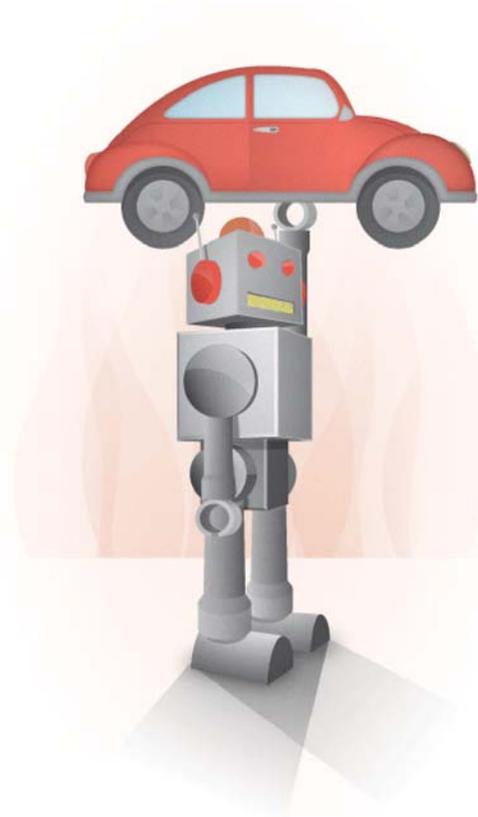
Falling

Search Traffic from a Single Engine

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

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*.

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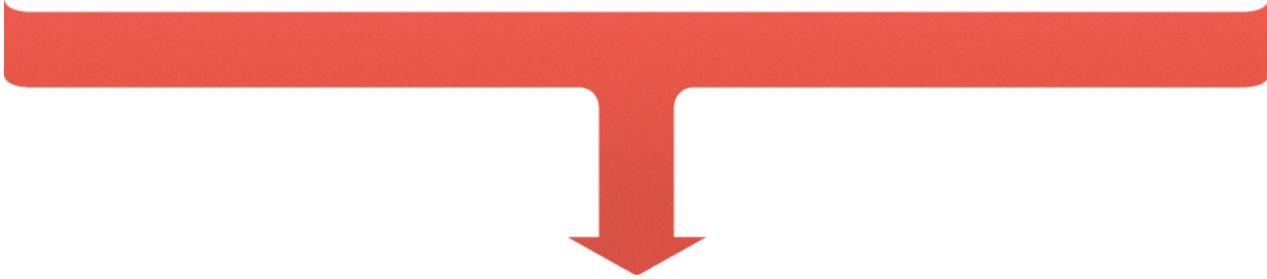
Congratulations! You've made it
through the entire Beginner's Guide to SEO!
Now you are ready to start optimizing your own site,
implement the tricks you have just learned
and improve your search results in the major search engines.

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 (Seattle Brand Studio)
* Linda Jenkinson
* Tom Critchlow
* Will Critchlow
* Dr. Pete
* Hamlet Batista
* chuckallied
* lorisa
* identity
* Pat Sexton
* SeoCatfish
* David LaFerney
* Kimber
* g1smd
* Steph Woods
* robbothan
* RandyP
* bookworm seo
* Sam Niccolls
* Danny Dover
* and Rand Fishkin