THE BEGINNERS GUIDE TO SEO

PDF

SEOmoz

2nd Edition

HOW SEARCH ENGINES OPERATE

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

I. CRAWLING AND INDEXING

1. *Crawling and Indexing*
Crawling and indexing the billions of documents, pages, files, news videos and media on the world wide web.

2. *Providing Answers*
Providing answers to user queries, most frequently through lists of relevant pages through retrieval and rankings.
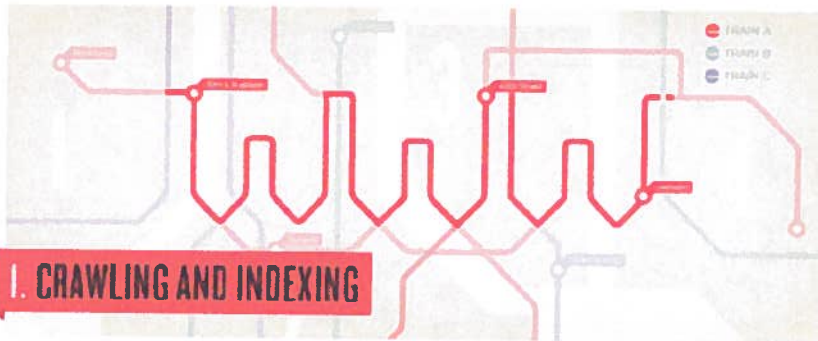
Imagine the World Wide Web as a network of stops in a big city subway system.

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

"The link structure of the web serves to bind 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.
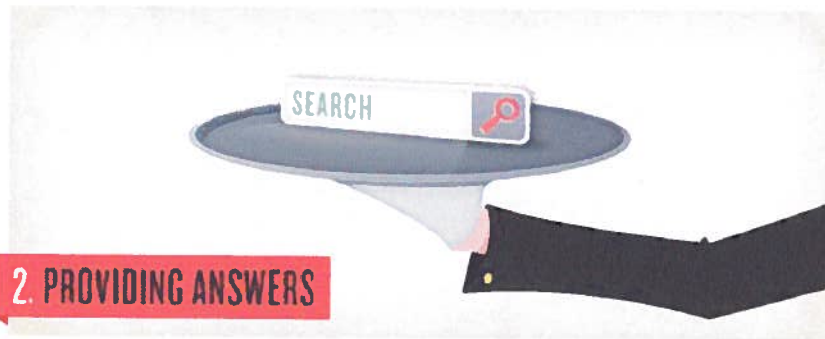
2. PROVIDING ANSWERS

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.
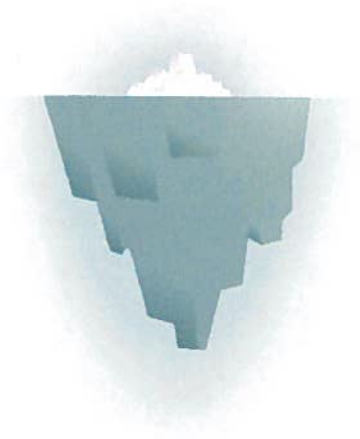
« TIME FOR AN EXPERIMENT »

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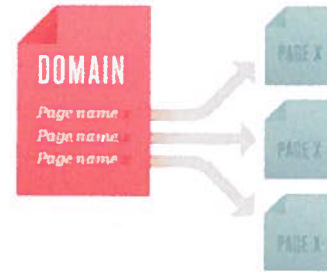
This process is certainly not alone in helping to educate search marketers.

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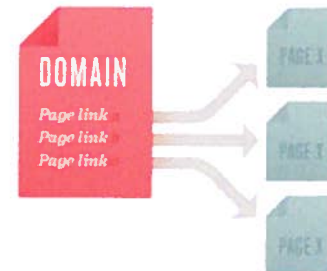
Through methods like patent analysis, experiments, and live testing and tweaking, search marketers as a community have come to understand many of the basic operations of search engines and the critical components of creating websites and pages that garner high rankings and significant traffic.

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

"Step" to "Domain"

DOMAIN
Page name
Page name
Page name

PAGE X
PAGE X
PAGE X

"Step" to "Domain"

DOMAIN
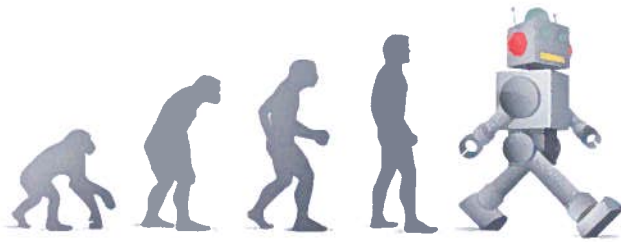Page link
Page link
Page link

PAGE X
PAGE Y
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."

CHAPTER THREE

WHY SEARCH ENGINE MARKETING IS NECESSARY

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.

**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.
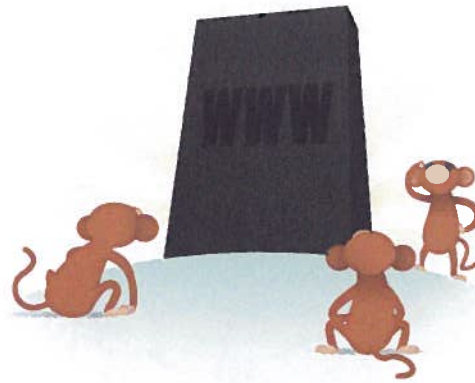

CHAPTER FOUR   THE BASICS OF SEARCH ENGINE FRIENDLY DESIGN & DEVELOPMENT

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.
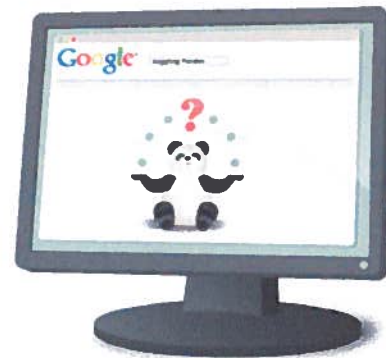
Indexable Content

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:

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.

"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?"
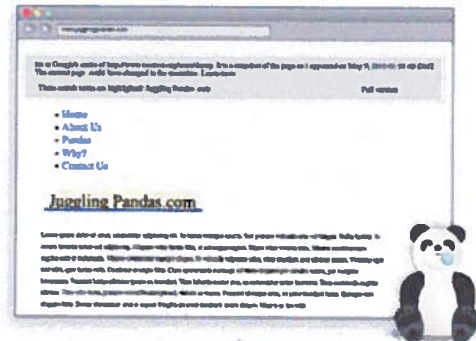
Now let's double-check some stuff

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

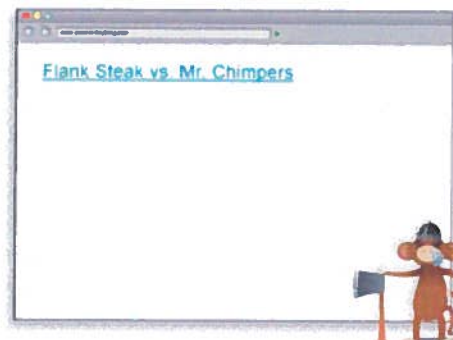## Whoa! That's what we look like?

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:
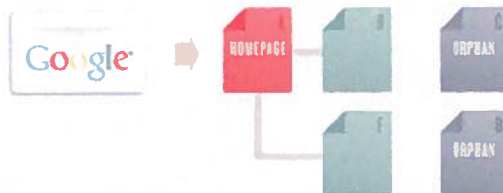
In the example above, Google's spider has reached page "A" and sees links to pages "B" and "E." However, even though C & D might be important pages on the site, the spider has no way to reach them (or even know they exist) because no direct, crawlable links point to those pages. As far as Google is concerned, 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

Jon Wye's Custom Designed Belts

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.
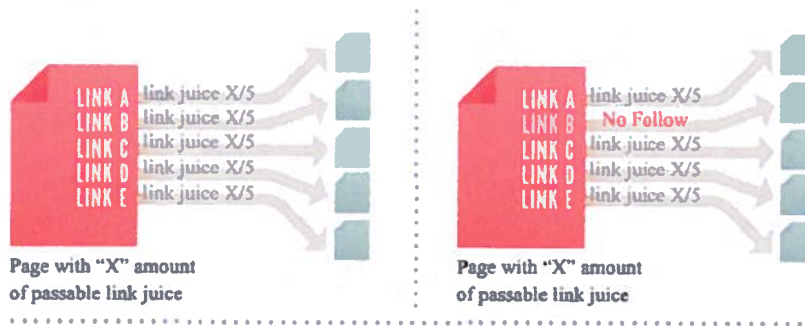
Yahoo! & Bing

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.
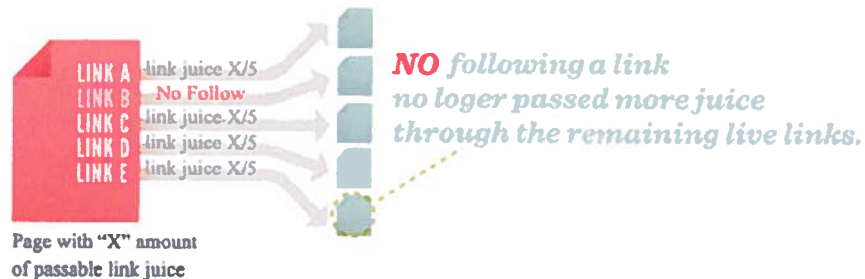
Ask.com

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.

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:


PAGE TARGETING THE PHRASE "RUNNING SHOES"

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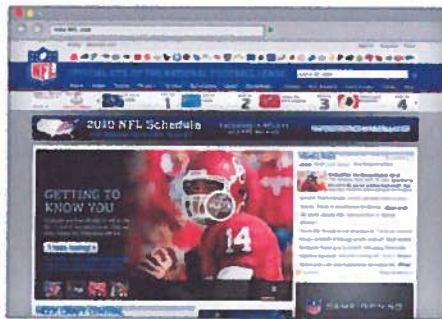
The title tag of any page appears at the top of Internet browsing software, but this location has been noted to receive a relatively small amount of attention from users, making it the least important of the three.
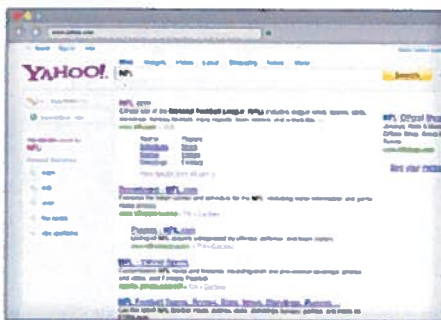
Using keywords in the title tag means that search engines will "bold" (or highlight) those terms in the search results when a user has performed a query with those terms. This helps garner a greater visibility and a higher click-through rate.

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.

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.

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

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.
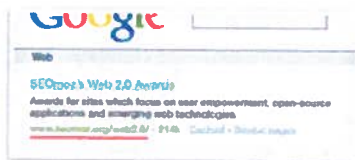
Above, the green text shows the url for SEOmoz's Web 2.0 awards. Since search engines display URLs in the results, they can impact clickthrough and visibility. URLs are also used in ranking documents, and those pages whose names include the queried search terms receive some benefit from proper, descriptive 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

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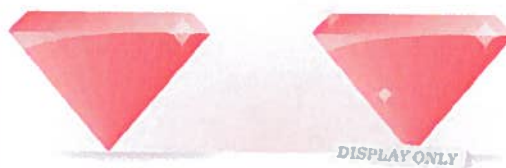
The Canonical URL tag attribute is similar in many ways to a 301 redirect from an SEO perspective. In essence, you're telling the engines that multiple pages should be considered as one (which a 301 does), without actually redirecting visitors to the new URL (often saving your development staff considerable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