

CHAPTER FOUR

THE BASICS OF SEARCH ENGINE FRIENDLY
DESIGN & DEVELOPMENT

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



Indexable Content

To perform better in search engine listings, your most important content should be in HTML text format. Images, Flash files, Java applets, and other non-text content are often ignored or devalued by search engine crawlers, 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 them in the HTML text on the page. However, more advanced methods are available for those who demand greater formatting or visual display styles:

1. **Provide alt text for images.** Assign images in gif, jpg, or png format "alt attributes" in HTML to give search engines a text description of the visual content.
2. **Supplement search boxes** with navigation and crawlable links.
3. **Supplement Flash or Java plug-ins** with text on the page.
4. **Provide a transcript for video and audio content** if the words and phrases used are meant to be indexed by the engines.

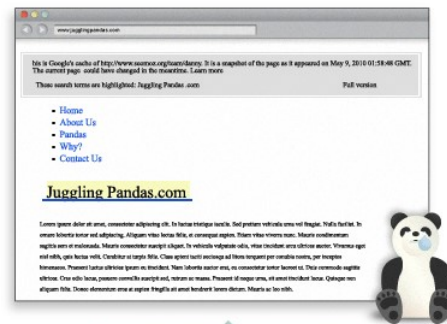
"I have a problem with getting found. I built a huge Flash site for juggling pandas and I'm not showing up anywhere on Google. What's up?"

Seeing your site as the search engines do

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



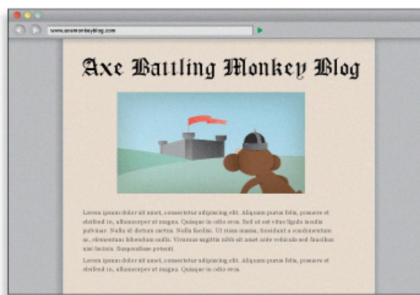
Through Browser



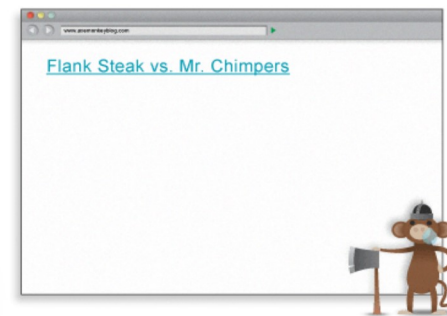
Through Google Cache

Whoa! That's what we look like?

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



What Humans See



What Search Engines See

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 built entirely in Flash, but sadly, this means that search engines cannot index any of the text content, or even the links to the individual games. Without any HTML text, this page would have a very hard time ranking in search results.

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

Crawlable Link Structures

Just as search engines need to see content in order to list pages in their massive keyword-based indexes, they also need to see links in order to find the content in the first place. A crawlable [link structure](#)—one that lets the crawlers browse the pathways of a website—is vital to them finding all of the pages on a website. Hundreds of thousands of sites make the critical mistake of structuring their navigation in ways that search engines cannot access, hindering their ability to get pages listed in the search engines' indexes.

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



In the example above, Google's crawler has reached page A and sees links to pages B and E. However, even though C and D might be important pages on the site, the crawler has no way to reach them (or even know they exist). This is because no direct, crawlable links point pages C and D. As far as Google can see, they don't exist! Great content, good keyword targeting, and smart marketing won't make any difference if the crawlers can't reach



your pages in the first place.



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. These links are the original navigational elements of the Internet – known as hyperlinks. In the above illustration, the "<a" tag indicates the start of a link. The link referral location tells the browser (and the search engines) where the link points. 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 linked-to page is about custom belts made by Jon Wye, thus the anchor text "Jon Wye's Custom Designed Belts." The "" tag closes the link to constrain the linked text between the tags and prevent the link from encompassing other elements on the page.

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

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

Submission-required forms

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

Links in unparseable 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 you'd like crawlers to crawl.

Links pointing to pages blocked by the Meta Robots tag or robots.txt

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

Frames or iframes

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

Robots don't use search forms

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

Links in Flash, Java, and other plug-ins

The links embedded inside the Juggling Panda site (from our above example) are perfect illustrations of this phenomenon. Although dozens of pandas are listed and linked to on the page, no crawler can reach them through the site's link structure, rendering them invisible to the engines and hidden from users' search queries.

Links on pages with many hundreds or thousands of links

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

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

rel="nofollow"

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

```
<a href="https://moz.com" rel="nofollow">Lousy Punks!</a>
```

Links can have lots of attributes. The engines ignore nearly all of them, with the important exception of the rel="nofollow" attribute. In the example above, adding the rel="nofollow" attribute to the link tag tells the search engines that the site owners do not want this link to be interpreted as an endorsement of the target page.

Nofollow, taken literally, instructs search engines to not follow a link (although some do). The nofollow tag came about as a method to help stop automated blog comment, guest book, and link injection spam, 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, but it is clear they do not pass as much weight as normal links.

Are nofollow links bad?

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

Google

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

Bing & Yahoo!

Bing, which powers Yahoo search results, has also stated that they do not include nofollow links in the link graph, though their crawlers may still use nofollow links as a way to discover new pages. So while they may follow the links, they don't use them in rankings calculations.

Keyword Usage and 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 indexes rather than storing 25 billion web pages all in one database. Millions and millions of smaller databases, each centered on a particular keyword term or phrase, allow the engines to retrieve the data they need in a mere fraction of a second.

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



Keyword Domination

Keywords dominate how we communicate our search intent and interact with the engines. When we enter words to search for, the engine matches pages to retrieve based on the words we entered. The order of the words ("pandas juggling" vs. "juggling pandas"), spelling, punctuation, and capitalization provide additional information that the engines use to help retrieve the right pages and rank them.

Search engines measure how keywords are used on pages to help determine the relevance of a particular document to a query. One of the best ways to optimize a page's rankings is to ensure that the keywords you want to rank for are prominently used in titles, text, and metadata.

Generally speaking, as you make your keywords more specific, you narrow the competition for search results, and improve your chances of achieving a higher ranking. The map graphic to the left compares the relevance of the broad term "books" to the specific title *Tale of Two Cities*. Notice that while there are a lot of results for the broad term, there are considerably fewer results (and thus, less competition) for the specific result.

Keyword Abuse

Since the dawn of online search, folks have abused keywords in a

misguided effort to manipulate the engines. This involves "stuffing" keywords into text, URLs, meta tags, and links. Unfortunately, this tactic almost always does more harm than good for your site.

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

The best practice is to use your keywords naturally and strategically (more on this below). If your page targets the keyword phrase "Eiffel Tower" then you might naturally include content about the Eiffel Tower itself, the history of the tower, or even recommended Paris hotels. On the other hand, if you simply sprinkle the words "Eiffel Tower" onto a page with irrelevant content, such as a page about dog breeding, then your efforts to rank for "Eiffel Tower" will be a long, uphill battle. **The point of using keywords is not to rank highly for all keywords, but to rank highly for the keywords that people are searching for when they want what your site provides.**

On-Page Optimization

Keyword usage and targeting are still a part of the search engines' ranking algorithms, and we can apply some effective techniques for keyword usage to help create pages that are well-optimized. Here at Moz, 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 phrase:

In the title tag at least once. Try to keep the keyword phrase as close to the beginning of the title tag as possible. More detail on title tags follows later in this section.

Once prominently near the top of the page.

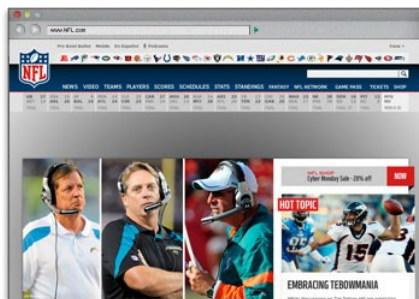
At least two or three times, including variations, in the body copy on the page. Perhaps a few more times if there's a lot of text content. You may find additional value in using the keyword or variations more than this, but in our experience adding more instances of a term or phrase tends to have little or no impact on rankings.

At least once in the alt attribute of an image on the page. This not only helps with web search, but also image search, which can occasionally bring valuable traffic.

Once in the URL. Additional rules for URLs and keywords are discussed later on in this section.

At least once in the meta description tag. Note that the meta description tag does not get used by the engines for rankings, but rather helps to attract clicks by searchers reading the results page, as the meta description becomes the snippet of text used by the search engines.

And you should generally *not* use keywords in link anchor text pointing to other pages on your site; this is known as [Keyword Cannibalization](#).



The title tag of any page appears at the top of Internet browsing

Keyword Density Myth

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

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-occurrence)
4. The main theme, topic, and sub-topics (on-topic issues) of the documents

The Conclusion:

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

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



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

Title Tags

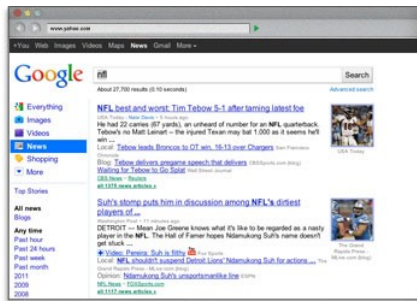
The [title element of a page](#) is meant to be an accurate, concise description of a page's content. It is critical to both user experience and search engine optimization.

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

Be mindful of length

Search engines display only the first 65-75 characters of a title tag

Title tags should be descriptive and readable. The title tag is a new visitor's first interaction with your brand and should convey the most positive impression possible. Creating a compelling title tag will help grab attention on the search results page, and attract more visitors to your site. This underscores that SEO is about not only optimization and strategic keyword usage, but the entire user experience.



09-Mar-17 9:54 PM

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. It is an extremely important part of search marketing. Crafting a readable, compelling description using important keywords (notice how Google bolds the searched keywords in the description) can draw a much higher click-through rate of searchers to your page.

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

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

Not as important meta tags

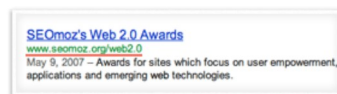
Meta Keywords: The meta keywords tag had value at one time, but is no longer valuable or important to search engine optimization. For more on the history and a full account of why meta keywords has fallen into disuse, read [Meta Keywords Tag 101](#) from SearchEngineLand.

Meta Refresh, Meta Revisit-after, Meta Content-type, and others: Although these tags can have uses for search engine optimization, they are less critical to the process, and so we'll leave it to [Google's Search Console Help](#) to discuss in greater detail.



URL Structures

URLs—the addresses for documents on the web—are of great value from a search perspective. They appear in multiple important locations.



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

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

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

URL Construction Guidelines

Employ empathy

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

Shorter is better

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

Keyword use is important (but overuse is dangerous)

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

Go static



The best [URLs](#) are human-readable and without lots of parameters, numbers, and symbols. Using technologies like `mod_rewrite` for Apache and `ISAPI_rewrite` for Microsoft, you can easily transform dynamic URLs like this <https://moz.com/blog?id=123> into a more readable static version like this: <https://moz.com/blog/google-fresh-factor>. Even single dynamic parameters in a URL can result in lower overall ranking and indexing.

Use hyphens to separate words

Not all web applications accurately interpret separators like underscores (`_`), plus signs (`+`), or spaces (`%20`), so instead use the hyphen character (`-`) to separate words in a URL, as in the "google-fresh-factor" URL example above.

Canonical and Duplicate Versions of Content

Duplicate content is one of the most vexing and troublesome problems any website can face. Over the past few years, search engines have cracked down on pages with thin or duplicate content by assigning them lower rankings.

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



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

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



*Instead, if the site owner took those three pages and [301-redirected](#) them, the search engines would have only one **strong** 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 stop competing with each other, but also create a stronger relevancy and popularity signal overall. This will positively impact your ability to rank well in the search engines.

Canonical Tag to the rescue!

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

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

« THE INNER WORKINGS »

```
<link rel="canonical" href="https://moz.com/blog"/>
```

This tells search engines that the page in question should be treated as though it were a copy of the URL <https://moz.com/blog> and that all of the link and content metrics the engines apply should flow back to that URL.

From an SEO perspective, the Canonical URL tag attribute is similar to a 301 redirect. In essence, you're telling the engines that multiple pages should be considered as one (which a 301 does), but without actually redirecting visitors to the new URL. This has the added bonus of saving your development staff considerable heartache.

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

Easy as pie!



Rich Snippets

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

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

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

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

Rich Snippets in the Wild

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

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

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

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

Defending Your Site's Honor

How scrapers steal your rankings

Unfortunately, the web is littered with unscrupulous websites whose business and traffic models depend on plucking content from other sites and re-using it (sometimes in strangely modified ways) on their own domains. This practice of fetching your content and re-publishing is called "scraping," and the scrapers perform remarkably well in search engine rankings, often outranking the original sites.

When you publish content in any type of feed format, such as RSS or XML, make sure to ping the major blogging and tracking services (Google, Technorati, Yahoo!, etc.). You can find instructions for pinging 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. So, by including links back to your site, and to the specific post you've authored, you can ensure that the search engines see most of the copies linking back to you (indicating that your source is probably the originator). To do this, you'll need to use absolute, rather than relative links in your internal linking structure. Thus, rather than linking to your home page using:

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

You would instead use:

```
<a href="https://moz.com">Home</a>
```

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

There are more advanced ways to protect against scraping, but none of them are entirely foolproof. You should expect that the more popular and visible your site gets, the more often you'll find your content scraped and re-published. Many times, you can ignore this problem: but if it gets very severe, and you find the scrapers taking away your rankings and traffic, you might consider using a legal process called a DMCA takedown. Moz CEO Sarah Bird offers some quality advice on this topic: [Four Ways to Enforce Your Copyright: What to Do When Your Online Content is Being Stolen](#).



Written by [Rand Fishkin](#) and Moz Staff

CHAPTER 1

How Search Engines Operate

CHAPTER 6

*How Usability, Experience,
& Content Affect Rankings*

CHAPTER 2

*How People Interact With
Search Engines*

CHAPTER 7

Growing Popularity and Links

CHAPTER 3

*Why Search Engine Marketing
is Necessary*

CHAPTER 8

*Search Engine Tools
and Services*

CHAPTER 4

*The Basics of Search Engine
Friendly Design & Dev.*

CHAPTER 9

*Myths & Misconceptions
About Search Engines*

CHAPTER 5

Keyword Research

CHAPTER 10

Measuring and Tracking Success

⬆️ BACK TO TOP



[Contact](#) [Products](#) [API](#) [Terms & Privacy](#) [Jobs](#) [Help](#)

Copyright 2016 © Moz, Inc. All rights reserved.

[Back to Top](#)

SEO Tools

- [Keyword Research](#)
- [SEO Audit & Crawl](#)
- [Backlink Research](#)
- [Rank Tracking](#)
- [SEO Toolbar](#)

Local Marketing

- [Business Listings Audit](#)
- [Citation Cleanup](#)
- [Local Ranking Factors](#)
- [Local For Enterprise](#)

Marketing Resources

- [Moz Blog](#)
- [Beginner's Guide To SEO](#)
- [Community Q&A](#)
- [Workshops & Training](#)
- [Recommended SEO Companies](#)

Get the Moz Top 10

The 10 most valuable pieces of content we can find for SEOs. Every 2 weeks.

