

CHAPTER EIGHT

SEARCH ENGINE TOOLS AND SERVICES

SEOs tend to use a lot of tools. Some of the most useful are provided by the search engines themselves. Search engines want webmasters to create sites and content in accessible ways, so they provide a variety of tools, analytics and guidance. These free resources provide data points and unique opportunities for exchanging information with the engines.

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

Common Search Engine Protocols

1. Sitemaps

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

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

XML

Extensible Markup Language (recommended format)

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.

Relatively large file sizes. Since XML requires an open tag and a close tag around each element, file sizes can get very large.

RSS

Really Simple Syndication or Rich Site Summary

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

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

Text

Text File

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

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

2. Robots.txt

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

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

The following commands are available:

Disallow

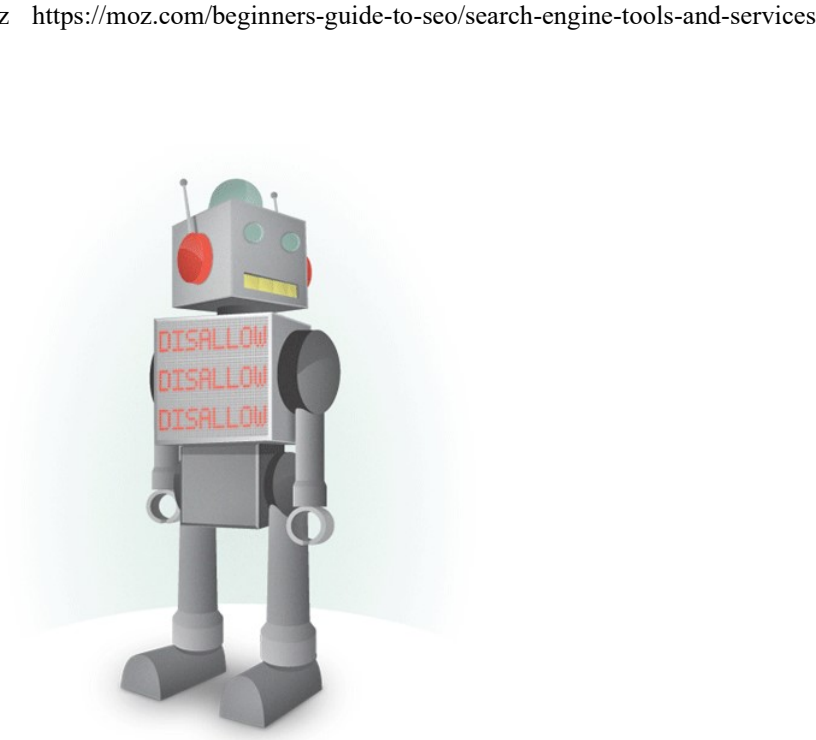
Prevents compliant robots from accessing specific pages or folders.

Sitemap

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

Crawl Delay

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



An Example of Robots.txt

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

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

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

Warning: Not all web robots follow robots.txt. People with bad intentions (e.g., e-mail address scrapers) build bots that don't follow this protocol; and in extreme cases they 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 file. Instead, these pages can utilize the meta robots tag (discussed next) to keep the major search engines from indexing their high-risk content.

3. Meta Robots

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

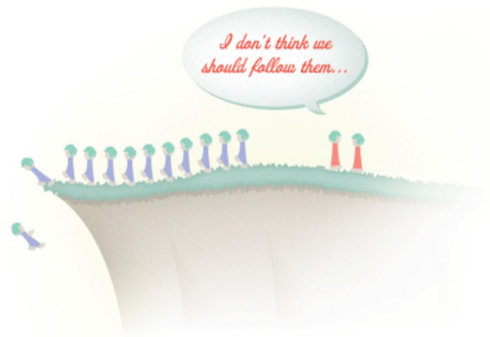
The meta robots tag should be included in the head section of the HTML document.

An Example of Meta Robots

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

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, “NOINDEX, NOFOLLOW” tells robots not to include the given page in their indexes, and also not to follow any of the links on the page.



4. Rel="Nofollow"

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

An Example of nofollow

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

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

5. Rel="canonical"

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

```
http://www.example.com/
http://www.example.com/default.asp
http://example.com/
http://example.com/default.asp
http://Example.com/Default.asp
```

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

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

An Example of rel="canonical" for the URL
http://example.com/default.asp

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

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

Search Engine Tools

Google Search Console

[Google Search Console](#)

Key Features

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

Preferred Domain - The preferred domain is the one that a webmaster would like used to index their site's pages. If a webmaster specifies a preferred domain as http://www.example.com and Google finds a link to that site that is



Your Site on the Web

Statistics provided by search engine tools offer unique insight to SEOs, like keyword impressions, click-through rates, top pages delivered in search results, and linking statistics.

formatted as `http://example.com`, Google will treat that link as if it were pointing at `http://www.example.com`.

URL Parameters - You can indicate to Google information about each parameter on your site, such as `"sort=price"` and `"sessionid=2"`. This helps Google crawl your site more efficiently.

Crawl Rate - The crawl rate affects the speed (but not the frequency) of Googlebot's requests during the crawl process.

Malware - Google will inform you if it has found any malware on your site. Malware creates a bad user experience, and hurts your rankings.

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

HTML Suggestions - Google looks for search engine-unfriendly HTML elements such as issues with meta descriptions and title tags.



Site Configuration

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

+1 Metrics

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

Labs

The Labs section of Search Console contains reports that Google considers still in the experimental stage, but which can nonetheless be useful to webmasters. One of the most important of these reports is Site Performance, which indicates how fast or slow your site loads for visitors.

Bing Webmaster Tools

[Bing Webmaster Tools](#)

Key Features

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

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

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

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

Moz Open Site Explorer

Moz's [Open Site Explorer](#) provides valuable insight into your website and links.

Features

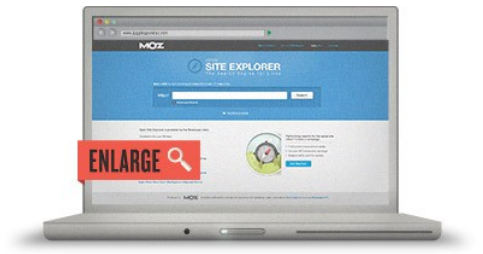
Identify Powerful Links - Open Site Explorer sorts all of your inbound links by their metrics that help you determine which links are most important.

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

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

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

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



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

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

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



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.

