SEO Guidelines

SEO Best Practices

Michael Borodin
7/21/2014

Contents

Google Webmaster Guidelines

Following these guidelines will help Google find, index, and rank your site. Even if you choose not to implement any of these suggestions, we strongly encourage you to pay very close attention to the "Quality Guidelines," which outline some of the illicit practices that may lead to a site being removed entirely from the Google index or otherwise impacted by an algorithmic or manual spam action. If a site has been affected by a spam action, it may no longer show up in results on Google.com or on any of Google's partner sites.

When your Site is ready

- Submit it to Google at http://www.google.com/submityourcontent/.
- Submit a Sitemap using Google Webmaster Tools. Google uses your Sitemap to learn about the structure of your site and increases the coverage of your website.
- Make sure all sites that should know about your pages are aware that you site is online

Design and Content Guidelines

- Make a site with a clear hierarchy and text links. Every page should be reachable from at least one static text link.
- Offer a site map to your users with links that point to the important parts of your site. If the site map has an extremely large number of links, you may want to break the site map into multiple pages.
- Keep the links on a given page to a reasonable number.
- Create a useful, information-rich site, and write pages that clearly and accurately describe your content.
- Think about the words users would type to find your pages, and make sure that your site actually includes those words within it.
- Try to use text instead of images to display important names, content, or links. The Google crawler doesn't recognize text contained in images. If you must use images for textual content, consider using the "ALT" attribute to include a few words of descriptive text.
- Make sure that your <title> elements and ALT attributes are descriptive and accurate.
- Check for broken links and correct HTML.
- If you decide to use dynamic pages (i.e., the URL contains a "?" character), be aware that not every search engine spider crawls dynamic pages as well as static pages. It helps to keep the parameters short and the number of them few.
- Review our recommended best practices for images, video and rich snippets.

Technical Guidelines

- Use a text browser such as Lynx to examine your site, because most search engine spiders see your site much as Lynx would. If fancy features such as JavaScript, cookies, session IDs, frames, DHTML, or Flash keep you from seeing your entire site in a text browser, then search engine spiders may have trouble crawling your site.
- Allow search bots to crawl your sites without session IDs or arguments that track their path through the site. These techniques are useful for tracking individual user behavior, but the access pattern of bots is entirely different. Using these techniques may result in incomplete indexing of your site, as bots may not be able to eliminate URLs that look different but actually point to the same page.
- Make sure your web server supports the If-Modified-Since HTTP header. This feature allows your web server to tell Google whether your content has changed since we last crawled your site. Supporting this feature saves you bandwidth and overhead.
- Make use of the robots.txt file on your web server. This file tells crawlers which directories can or cannot be crawled. Make sure it's current for your site so that you don't accidentally block the Googlebot crawler. Visithttp://code.google.com/web/controlcrawlindex/docs/faq.html to learn how to instruct robots when they visit your site. You can test your robots.txt file to make sure you're using it correctly with the robots.txt analysis tool available in Google Webmaster Tools.
- Make reasonable efforts to ensure that advertisements do not affect search engine rankings. For example, Google's AdSense ads and DoubleClick links are blocked from being crawled by a robots.txt file.
- If your company buys a content management system, make sure that the system creates pages and links that search engines can crawl.
- Use robots.txt to prevent crawling of search results pages or other auto-generated pages that don't add much value for users coming from search engines.
- Test your site to make sure that it appears correctly in different browsers.
- Monitor your site's performance and optimize load times. Google's goal is to provide users with the most relevant results and a great user experience. Fast sites increase user satisfaction and improve the overall quality of the web (especially for those users with slow Internet connections), and we hope that as webmasters improve their sites, the overall speed of the web will improve.
- Google strongly recommends that all webmasters regularly monitor site performance using Page Speed, YSlow, WebPagetest, or other tools. For more information, tools, and resources, see Let's Make the Web Faster. In addition, the Site Performance tool in Webmaster Tools shows the speed of your website as experienced by users around the world.

Quality Guidelines

These quality guidelines cover the most common forms of deceptive or manipulative behavior, but Google may respond negatively to other misleading practices not listed here. It's not safe to assume that just because a specific deceptive technique isn't included on this page, Google approves of it. Webmasters who spend their energies upholding the spirit of the basic principles will provide a much better user experience and subsequently enjoy better ranking than those who spend their time looking for loopholes they can exploit.

If you believe that another site is abusing Google's quality guidelines, please let us know by filing a spam report. Google prefers developing scalable and automated solutions to problems, so we attempt to minimize hand-to-hand spam fighting. While we may not take manual action in response to every report, spam reports are prioritized based on user impact, and in some cases may lead to complete removal of a spammy site from Google's search results. Not all manual actions result in removal, however. Even in cases where we take action on a reported site, the effects of these actions may not be obvious.

Quality guidelines - basic principles

- Make pages primarily for users, not for search engines.
- Don't deceive your users.
- Avoid tricks intended to improve search engine rankings. A good rule of thumb is whether you'd feel comfortable explaining what you've done to a website that competes with you, or to a Google employee. Another useful test is to ask, "Does this help my users? Would I do this if search engines didn't exist?"
- Think about what makes your website unique, valuable, or engaging. Make your website stand out from others in your field.

Quality guidelines - specific guidelines

Avoid the following techniques:

- Automatically generated content
- Participating in link schemes
- Cloaking
- Sneaky redirects
- Hidden text or links
- Doorway pages
- Scraped content
- Participating in affiliate programs without adding sufficient value
- Loading pages with irrelevant keywords
- Creating pages with malicious behavior, such as phishing or installing viruses, Trojans, or other badware
- Abusing rich snippets markup
- Sending automated queries to Google

Engage in good practices like the following:

- Monitoring your site for hacking and removing hacked content as soon as it appears
- Preventing and removing user-generated spam on your site

If your site violates one or more of these guidelines, then Google may take manual action against it. Once you have remedied the problem, you can submit your site for reconsideration.

Robots.txt File

Purpose

It Can Avoid Wastage of Server Resources

Many, if not most websites, have some sort of scripts (computer programs) that run on their website. For example, many websites have some sort of contact form, such as that created using the Free Feedback Form Script Wizard. Some also have a search, such as that which you see in the left column of every page on thesitewizard.com.

When search engine robots or spiders index your site, they actually call your scripts just as a browser would. If your site is like mine, where the scripts are solely meant for the use of humans and serve no practical use for a search engine (why should a search engine need to invoke my feedback form or use my site search engine?) you may want to block spiders from the directories that contain your scripts. For example, I block spiders from my feedback form, search engine and CGI-BIN directory. Hopefully, this will reduce the load on the web server that occurs when scripts are executed by removing unnecessary executions.

Of course there are the occasional ill-behaved robots that hit your server at high speed. Such spiders can actually bring down your server or at the very least slow it down for the real users who are trying to access it. If you know of any such spiders, you might want to exclude them too. You can do this with a robots.txt file. Unfortunately though, ill-behaved spiders often ignore robots.txt files as well.

It Can Save Your Bandwidth

If you look at your website's web statistics, you will undoubtedly find many requests for the robots.txt file by various search engine spiders. The search engines try to retrieve the robots.txt file before indexing your website, to see if you have any special instructions for them.

If you don't have a robots.txt file, your web server will return a 404 error page to the engine instead. For those who have customized, that customised 404 page will end up being sent to the spider repeatedly throughout the day. Now, if you have customized your 404 page, chances are that it's bigger than the standard server error message "404 File Not Found" (since you will want your error page to say more than the default error message). In other words, failing to create arobots.txt will cause the search engine spider to use up more of your bandwidth as a result of its repeated retrieval of your large 404 error file. (How much more depends, of course, on the size of your 404 error page.)

Some spiders may also request for files which you feel they should not. For example, some search engines also index graphic files (like ".gif", ".jpg" and ".png" files"). If you don't want them to do so, you can ban it from your graphic files directory using your robots.txt file.

It Removes Clutter from your Web Statistics

I don't know about you, but one of the things I check from my web statistics is the list of URLs that visitors tried to access, but met with a 404 File Not Found Error. Often this tells me if I made a spelling error in one of the internal links on one of my sites (yes, I know — I should have checked all links in the first place, but mistakes do happen).

If you don't have a robots.txt file, you can be sure that /robots.txt is going to feature in your web statistics 404 report, adding clutter and perhaps unnecessarily distracting your attention from the real bad URLs that need your attention.

Refusing a Robot

Sometimes you don't want a particular spider to index your site for some reason or other. Perhaps the robot is ill-behaved and spiders your site at such a high speed that it takes down your entire server. Or perhaps you prefer that you don't want the images on your site indexed in an image search engine. With a robots.txt file, you can exclude certain spiders from indexing your site with a robots.txt directive, provided the spider obeys the rules in that file.

Meta Tags that Google understands

Purpose

Meta Tags are a great way for webmasters to provide search engines with information about their sites. Meta Tags can be used to provide information in all sorts of clients, and each system processes only the Meta Tags they understand and ignores the rest. Meta Tags are added to the <Head> Section of your HTML page.

Google understands the following Meta Tags (and related items)

Tag	Description
<meta name="description" content="A description of the page" />	This tag provides a short description of the page. In some situations this description is used as a part of the snippet shown in the search results. More information
<title>The Title of the Page</title>	While technically not a meta tag, this tag is often used together with the "description". The contents of this tag are generally shown as the title in search results (and of course in the user's browser). More information
<meta name="robots" content="...," /> **<meta name="Googlebot" content="...," />**	These meta tags can control the behavior of search engine crawling and indexing. The robots meta tag applies to all search engines, while the "Googlebot" meta tag is specific to Google. The default values are "index, follow" (the same as "all") and do not need to be specified. We understand the following values (when specifying multiple values, separate them with a comma): - noindex: prevents the page from being indexed - nofollow: prevents the Googlebot from following links from this page - nosnippet: prevents a snippet from being shown in the search results - noodp: prevents the alternative description from the ODP/DMOZ from being used - Noarchive: prevents Google from showing the Cached link for a page. - unavailable_after:[date]: lets you specify the exact time and date you want to stop crawling and indexing of this page - Noimageindex: lets you specify that you do not want your page to appear as the referring page for an image that appears in Google search results. - None: is equivalent to noindex, nofollow. You can now also specify this information in the header of your pages using the "X-Robots-Tag" HTTP header directive. This is particularly useful if you wish to limit indexing of non-HTML files like graphics or other kinds of documents. More information about robots meta tags
<meta name="Google" content="notranslate" />	When we recognize that the contents of a page are not in the language that the user is likely to want to read, we often provide a link to a translation in the search results. In general, this gives you the chance to provide your unique and compelling content to a much larger group of users. However, there may be situations where this is not desired. This

	meta tag tells Google that you don't want us to provide a translation for this page.
<meta name="Google-site-verification" content="..." />	You can use this tag on the top-level page of your site to verify ownership for Webmaster Tools. Please note that while the values of the "name" and "content" attributes must match exactly what is provided to you (including upper and lower case), it doesn't matter if you change the tag from XHTML to HTML or if the format of the tag matches the format of your page. More information
<meta http-equiv="Content-Type" content="...; charset=..." /> <meta charset="..." >	This defines the page's content type and character set. Make sure that you surround the value of the content attribute with quotes - otherwise the charset attribute may be interpreted incorrectly. We recommend using Unicode/UTF-8 where possible. More information
<meta http-equiv="refresh" content="...;url=..." />	This meta tag sends the user to a new URL after a certain amount of time, and is sometimes used as a simple form of redirection. However, it is not supported by all browsers and can be confusing to the user. The W3C recommends that this tag not be used. We recommend using a server-side 301 redirect instead.

Other points to note:

- Google can read both HTML and XHTML-style Meta Tags, regardless of the code used on the page.
- With the exception of verify, case is generally not important in Meta Tags.

This is not an exclusive list of available Meta Tags, and you should feel free to use unlisted Meta Tags if they are important to your site. Just remember that Google will ignore Meta Tags it doesn't know.

How Google Search works

The three key processes in delivering search results to you are:

- Crawling: Does Google know about your site? Can we find it?
- Indexing: Can Google index your site?
- Serving: Does the site have good and useful content that is relevant to the user's search?

Crawling

Crawling is the process by which Googlebot discovers new and updated pages to be added to the Google index.

We use a huge set of computers to fetch (or "crawl") billions of pages on the web. The program that does the fetching is called Googlebot (also known as a robot, bot, or spider). Googlebot uses an algorithmic process: computer programs determine which sites to crawl, how often, and how many pages to fetch from each site.

Google's crawl process begins with a list of web page URLs, generated from previous crawl processes, and augmented with Sitemap data provided by webmasters. As Googlebot visits each of these websites it detects links on each page and adds them to its list of pages to crawl. New sites, changes to existing sites, and dead links are noted and used to update the Google index.

Google doesn't accept payment to crawl a site more frequently, and we keep the search side of our business separate from our revenue-generating AdWords service.

Indexing

Googlebot processes each of the pages it crawls in order to compile a massive index of all the words it sees and their location on each page. In addition, we process information included in key content tags and attributes, such as Title tags and ALT attributes. Googlebot can process many, but not all, content types. For example, we cannot process the content of some rich media files or dynamic pages.

Serving results

When a user enters a query, our machines search the index for matching pages and return the results we believe are the most relevant to the user. Relevancy is determined by over 200 factors, one of which is the PageRank for a given page. PageRank is the measure of the importance of a page based on the incoming links from other pages. In simple terms, each link to a page on your site from another site adds to your site's PageRank. Not all links are equal: Google works hard to improve the user experience by identifying spam links and other practices that negatively impact search results. The best types of links are those that are given based on the quality of your content.

Google Site Ranking

In order for your site to rank well in search results pages, it's important to make sure that Google can crawl and index your site correctly.

Google Webmaster Tools

https://support.google.com/webmasters/?hl=en#topic=3309300

Bing Webmasters Guide

These guidelines cover a broad range of topics and are intended to help your content be found and indexed within Bing. These guidelines will not cover every instance, nor provide prescriptive actions specific to every website. For more information, you should read our self-help documents and follow the Bing Webmaster Blog. In your Bing Webmaster Tools account, you will find SEO Reportsand the SEO Analyzer tool for on-demand scanning of individual pages. Both resources will offer basic guidance and recommendations in regards to site optimizations that you can apply to your site.

CONTENT

Content is what Bing seeks. By providing clear, deep, easy to find content on your website, we are more likely to index and show your content in search results. Websites that are thin on content, showing mostly ads or affiliate links, or those otherwise redirect visitors away to other sites quickly tend not to rank well. Your content should be easy to navigate, rich and engaging to the visitor, and provide them the information they seek. In many cases, content produced today will still be relevant years from now. In some cases, however, content produced today will go out of date quickly.

LINKS

Links pointing to your site help Bing discover new pages on your site. It also, traditionally, is regarded as a signal of popularity. The site linking to your content is essentially telling Bing they trust your content. As a result, Bing rewards links that have grown organically, that is, that have been added over time by content creators on other trusted, relevant websites made to drive real users from their site to your site. Abusive tactics that aim to inflate the number and nature of inbound links such as links buying, participating in link schemes (link farms, link spamming and excessive link manipulation) can lead to your site being delisted from the Bing index.

SOCIAL

Social media plays a role in today's effort to rank well in search results. The most obvious part it plays is via influence. If you are influential socially, this leads to your followers sharing your information widely, which in turn results in Bing seeing these positive signals. These positive signals can have an impact on how you rank organically in the long run.

INDEXATION
Being indexed is the first step to developing traffic from Bing. The main pathways to being indexed are:

- Links to your content help Bing find it, which can lead us to index your content
- Use of features within Bing Webmaster Tools such as Submit URL and Sitemap Upload are also ways to ensure we are aware of your content

Managing how Bingbot crawls your content can be done using the Crawl Control feature inside Bing Webmaster Tools. This feature allows you to control when, and at what pace, Bingbot crawls your

website. Webmasters are encouraged to allow Bingbot to crawl quickly and deeply to ensure we find and index as much content as possible.

TECHNICAL

PAGE LOAD TIME (PLT)

This element has a direct impact on the satisfaction a user has when they visit your website. Slow load times can lead to a visitor simply leaving your website, seeking their information elsewhere. If they came from our search results that may appear to us to be an unsatisfactory result that we showed. Faster is better, but take care to balance absolute page load speed with a positive, useful user experience.

ROBOTS.TXT

This file is a touch point for Bingbot to understand how to interact with your website and its content. You can tell Bingbot where to go, where not to go and by doing so guide its efforts to crawl your content. The best practice is to have this file placed at the root of your domain (www.yourwebsite.com/robots.txt) and maintain it to ensure it remains accurate.

This file is very powerful and has the capacity to block Bingbot from crawling your content. Should you block Bingbot, we will not crawl your content and your site or content from your site may not appear in our search results.

SITEMAP

This file often resides at the root of your host, say, www.yourdomain.com/sitemap.xml, and contains a list of all of the URLs from your website. Large sites may wish to create an index file containing links to multiple sitemap.xml documents, each containing URLs from the website. Care should be taken to keep these files as clean as possible, so remove old URLs if you take that content off your website.

Most websites have their sitemap files crawled daily to locate any fresh content. It's important to keep your sitemap files clean and current to help us find your latest content.

SITE TECHNOLOGY

The technology used on your website can sometimes prevent Bingbot from being able to find your content. Rich media (Flash, JavaScript, etc.) can lead to Bing not being able to crawl through navigation, or not see content embedded in a webpage. To avoid any issue, you should consider implementing a down-level experience which includes the same content elements and links as your rich version does. This will allow anyone (Bingbot) without rich media enabled to see and interact with your website.

REDIRECTS

If you move content on your website from one location to another, using a redirect makes sense. It can help preserve value the search engine has assigned to the older URL, helps ensure any bookmarks people have remain useful and keeps visitors to your website engaged with your content. Bing prefers you use a 301 permanent redirect when moving content, should the move be permanent. If the move is temporary, then a 302 temporary redirect will work fine. Do not use the rel=canonical tag in place of a proper redirect.

CANONICAL TAGS

The rel=canonical element helps us determine which version of a URL is the original, when multiple version of a URL return the same content. This can happen when, for example, you append a tracking notation to a URL. Two discrete URLs then exist, yet both have identical content. By implementing a rel=canonical, you can tell us the original one, giving us a hint as to where we should place our trust. Do not use this element in place of a proper redirect when moving content.

SEARCH ENGINE OPTIMIZATION (SEO)

Search Engine Optimization is a valid practice which seeks to improve technical and content aspects of a website, making the content easier to find, relevant, and more accessible to the search engine crawlers. Taken to extremes, some practices can be abused. The vast majority of instances render a website more appealing to Bing, though performing SEO-related work is no guarantee of improving rankings or receive more traffic from Bing. The main area of focus when optimizing a website should include:

- **<title>** tags – keep these clear and relevant
- **<meta description>** tags – keep these clear and relevant, though use the added space to expand on the <title> tag in a meaningful way
- **alt** attributes – use this attribute on <img> tags to describe the image, so that we can understand the content of the image
- **<h1>** tag – helps users understand the content of a page more clearly when properly used
- **Internal links** – helps create a view of how content inside your website is related. Also helps users navigate easily to related content.
- **Links to external sources** – be careful who you link to as it's a signal you trust them. The number of links pointing from your page to external locations should be reasonable.
- **Social sharing** – enabling social sharing encourages visitors to share your content with their networks
- **Crawlability**
- **XML Sitemaps** – make sure you have these set up and that you keep them fresh and current
- **Navigational structure** – keep it clean, simple and easy to crawl
- **Rich media cautions** – don't bury links to content inside JavaScript
- **Graceful degradation** – enable a clean down-level experience so crawlers can see your content
- URL structure – avoid using session IDs, &, # and other characters when possible
- Robots.txt – often placed at root of domain, be careful as its powerful; reference sitemap.xml (or your sitemap-index file) in this document
- Verify that Bingbot is not disallowed or throttled in robots.txt: reference
- Define high crawl rate hours in the Bing Webmaster Tools via the Crawl Control feature.
- Verify that Bingbot is not blocked accidentally at the server level by doing a "Fetch as Bingbot": reference
- Webmasters are encouraged to use the **Ignore URL Parameters (found under Configure My Site)** tool inside Bing Webmaster Tools to help Bingbot understand which URLs are to be indexed and which URLs from a site may be ignored
- **Site Structure**
- **Links** – cross link liberally inside your site between relevant, related content; link to external sites as well
- **URL structure and keyword usage** - keep it clean and keyword rich when possible

- **Clean URLs** – no extraneous parameters (sessions, tracking, etc.)
- **HTML & XML sitemaps** – enable both so users and crawlers can both find what they need – one does not replace the other
- **Content hierarchy** – structure your content to keep valuable content close to the home page
- **Global navigation** – springs from hierarchy planning + style of nav (breadcrumb, link lists, etc.) – helps ensure users can find all your content
- **Rich media warnings** – don't bury links in JavaScript/flash/Silverlight;keep content out of these as well
- **On-Page**
- **Head copy**
- Titles – unique, relevant, 65 characters or so long
- Descriptions – unique, relevant, grammatically correct, roughly 160 or fewer characters
- **Body Copy**
- H1, H2 and other H* tag usage to show content structure on page
- Only one <H1> tag per page
- ALT tag usage – helps crawlers understand what is in an image
- Keyword usage within the content/text – use the keyword/phrase you are targeting a few times; use variations as well
- **Anchor text** – using targeted keywords as the linked text (anchor text) to support other internal pages
- **Content**
- Build based on keyword research – shows you what users are actually looking for
- Down-level experience enhances discoverability – avoid housing content inside Flash or JavaScript – these block crawlers form finding the content
- Keep out of rich media and images – don't use images to house your content either
- Create enough content to fully meet the visitor's expectations. There are no hard and fast rules on the number of words per page, but providing more relevant content is usually safe.
- Produce new content frequently – crawlers respond to you posting fresh content by visiting more frequently
- Make it unique – don't reuse content from other sources – critical – content must be unique in its final form on your page
- Content management – using 301s to reclaim value from retiring content/pages – a 301 redirect can pass some value from the old URL to the new URL
- <rel canonical> to help engines understand which page should be indexed and have value attributed to it
- 404 error page management can help cleanse old pages from search engine indexes; 404 page should return a 404 code, not a 200 OK code. Reference.
- **Links**
- Plan for incoming & outgoing link generation – create a plan around how to build links internally and externally

- Internal & external link management – execute by building internal links between related content; consider social media to help build external links, or simply ask websites for them; paying for links is risky
- Content selection – planning where to link to – be thoughtful and link to only directly related/relevant items of content internally and externally
- Link promotion via social spaces – these can drive direct traffic to you, and help users discover content to link to for you
- Managing anchor text properly – carefully plan which actual words will be linked – use targeted keywords wherever possible

THINGS TO AVOID

CLOAKING

Cloaking is the practice of showing one version of a webpage to a search crawler like Bingbot, and another to normal visitors. Showing users different content than to the crawlers can be seen as a spam tactic and be detrimental to your website's rankings and can lead to your site being de-listed from our index. It is therefore recommended to be extremely cautious about responding differently to crawlers as opposed to "regular" visitors and to not cloak as a principle.

LINK SCHEMES, LINK BUYING, LINK SPAMMING

While link schemes may succeed in increasing the number of links pointing to your website, they will fail to bring quality links to your site, netting no positive gains. In fact, manipulating inbound links to artificially inflate the number of links pointed at a website can even lead to your site being delisted from our index.

SOCIAL MEDIA SCHEMES

Like farms are similar to link farms in that they seek to artificially exploit a network effect to game the algorithm. The reality is these are easy to see in action and their value is deprecated. Auto follows encourage follower growth on social sites such as Twitter. They work by automatically following anyone who follows you. Over time this creates a scenario where the number of followers you have is more or less the same as the number of people following you. This does not indicate you have a strong

influence. Following relatively few people while having a high follower count would tend to indicate a stronger influential voice.

META REFRESH REDIRECTS

These redirects reside in the code of a website and are programmed for a preset time interval. They automatically redirect a visitor when the time expires, redirecting them to other content. Rather than using Meta refresh redirects, we suggest you use a normal 301 redirect.

DUPLICATE CONTENT

Duplicating content across multiple URLs can lead to Bing losing trust in some of those URLs over time. This issue should be managed by fixing the root cause of the problem. The rel=canonical element can also be used but should be seen as a secondary solution to that of fixing the core problem. If excessive parameterization is causing duplicate content issue, we encourage you to use the Ignore.

KEYWORD STUFFING

When creating content, make sure to create your content for real users and readers, not to entice search engines to rank your content better. Stuffing your content with specific keywords with the sole intent of artificially inflating the probability of ranking for specific search terms is in violation of our guidelines and can lead to demotion or even the delisting of your website from our search results.