

How to Conduct a Technical SEO Audit

HubSpot OnCrawl



Table of Contents



1	Introduction
4	Goals of a technical audit
10	Essential tools for technical auditing
14	Indexing
21	Server performance
29	Site structure
39	Keyword performance
45	On-page optimization
52	Duplicate content
56	Mobile issues
60	International issues
65	Schema.org
71	Drawing conclusions from your audit
73	Conclusion

Introduction



Introduction

Understanding the technical aspects of your website's SEO is crucial to the success of your efforts online. However, keeping up with Google's numerous algorithm changes proves challenging to even the most tech-savvy of marketers. To help you set your website up for success, HubSpot and OnCrawl have teamed up to provide you with this Guide to Conducting a Technical SEO audit. This ebook will provide you with all the information, strategies, and tools necessary to help you understand how to use technical SEO to improve the performance of your website.

SEO (search engine optimisation) refers to techniques that help your website rank higher in search engine results pages (SERPs). One aspect of the overarching SEO topic is technical SEO. Technical SEO includes key ranking factors that are embedded in the structure of your website, including crawling, indexing and more! To understand how technical SEO affects the overall performance of your website, it is necessary to conduct a technical SEO audit.

What is a technical audit?

A technical SEO audit is a process of analyzing and tracking key website performance metrics that impact all areas of your website. The goal of a technical SEO audit is to identify the shortcomings of your website's structure that negatively affect its performance and thereby hurt its SEO and rankings in search.

Who should audit?

The technical nature of this type of SEO audit can scare many people away. But with the help of this guide, anyone can do it. Whether you're a content marketer, web developer, analyst, or someone else entirely, this guide will give you all the information and tools necessary to conduct a technical SEO audit successfully. Once your audit has revealed areas for improvement, it may be necessary to work with your web team create solutions.

When to audit

Many people think that a technical SEO audit is a one-and-done project. However, it is critical to get started with your audit today and continue to evaluate your website's performance regularly. As your business grows and your website changes and expands, it will become increasingly important to understand how these changes affect your SEO performance.

Here are a few excellent times to conduct a technical SEO audit:

- Your website is recently built
- Your website had been redesigned
- Your website's structure has changed
- It has been over a year since your last audit

Why are HubSpot and OnCrawl talking about technical SEO audits?

HubSpot offers a full stack of software for marketing, sales, and customer service, with a completely free CRM at its core. As thought leaders in a digital world, they are experts at helping businesses grow through with the help of their Inbound Methodology.

OnCrawl is the leading technical SEO platform for enterprise SEO audits and daily monitoring. Their solution helps businesses around the world open Google's blackbox and understand the technical aspects of their website to identify areas for improvement.

Together, HubSpot and OnCrawl are experts at empowering businesses to perform their best.

Goals of a technical audit



Goals of a technical audit

Find technical errors

Technical errors are essential red-flags that reveal key issues on your website that negatively impact its performance. Some of these errors could include slow website speed, broken links or duplicate content. Fixing these issues is crucial to improving your website's SEO as well as to improving the overall user experience.

The purpose of SEO is to optimize organic traffic, that is, visitors from search engines. In today's competitive world, that includes providing maximum value to your website visitors. Issues such as long load times, for example, could result in less traffic to your website and take away from the experience you want your website to provide.

Find opportunities to earn better SERP rankings

Understand and adapt to how Google answers search queries

RankBrain is Google's artificial intelligence system that processes and evaluates search results. RankBrain is essentially Google's way to improve understanding of search queries and of websites that deliver value to the users performing searches in Google. This is done through the use of [machine learning](#). As RankBrain analyses more search queries every day, it becomes better at delivering web pages in the search results that are relevant to the query.

Since Rankbrain matures with every search, optimizing your SEO for Rankbrain is like trying to hit a moving target. Instead, focusing on technical SEO and on medium and long tail keywords (i.e. keywords that are in the middle of the pack, not too broad but not too narrow) will be your best bet. On one hand, technical SEO improves Google's ability to discover and understand your site. On the other hand, keyword selection is key to appropriate content creation. By being too vague in your keyword selection, it will be difficult to rank content with a high level of competition. Selecting medium and long tail keywords, however, will give you less competition and it will be easier to rank your content for these keywords.

Convince Google of the importance of your key pages

Google classifies URLs during crawls with their algorithm known as Page Importance. By ranking the page on a domain, Google is better able to understand what pages are most important to crawl and which have lower priority. To date, there are simply too many web pages, [nearly 2 billion](#), and crawlers cannot visit every website each time a page is added or updated. Google uses Page Importance to determine which web pages crawlers will prioritize visiting over others.

There are a number of factors that Google considers when determining **Page Importance**:

- Page position
- Page rank
- Type of page or file
- Sitemap.xml
- Internal page rank
- Number and type of internal links
- Number of external links
- Relevance and quality of content
- Update frequency
- Source codes and website quality

By optimizing for all of these factors, you'll be able to increase the Page Importance of your page and help Google understand which pages of your website take first priority in the crawling queue.

Improve how your crawl budget is spent on your site

The crawl budget can be defined as the number of pages that search engines crawl, or visit, on your website within a given period of time. Since search engines only index, or make available for use as search results, pages that they have discovered during a crawl, you can see why improving how your crawl budget is spent can be beneficial.

By increasing the number of key pages on your website that Google crawls and decreasing the crawl budget spent on less important pages, you can improve the success that your website has in organic search.

Your crawl budget is the average number of hits by a googlebot per day. You can calculate this number by analyzing your server logs for Google's bot hits during a given period, or you can find an estimate of "Pages crawled per day" in the old Google Search Console, under Crawl, then Crawl Stats.

How to evaluate your crawl budget:

1. First, find how many pages are on your site. Using a data tool such as OnCrawl to combine crawl data and additional data such as sitemaps and Google Search Console, you can establish how many pages can be found in your site, whether or not Google has indexed them.

If you need a quick estimate, in Google, search "site:yourwebsite.com". The total number of search results will tell you how many indexed pages your website has. The real number of pages on your website will be larger.

2. Second, go to Google Search Console account and go to Crawl, and then to Crawl Stats.
3. Third, take the "Pages crawled per day" average number and the total number of the URLs on your website. Divide the total number of URLs by the Pages crawled per day. A crawl budget that could theoretically allow your site to be fully crawled in less than 5 days is usually a good sign.

You don't have a lot of direct control to increase your crawl budget. Google will increase your crawl budget if Google feels it is necessary based on the ability of your pages to answer search queries. Consequently, if you're concerned about your crawl budget then in addition to improving your SEO in general, it's significantly more effective to optimize the use of your current budget than to attempt to increase it.

Seize key positions in search results to increase traffic

Not all search results are created equal. Some positions draw more clicks; some provide additional information; some are the only result presented for voice searches, or the only content viewed by users who are just looking for a quick, informational answer.

"Looking for a quick, informational answer" is an example of search intent. Search intent can be defined as the ultimate goal of the person conducting the search, the missing piece of information that the user wants to find. Search intent can be informational (What is 14mph in km/h?), navigational (How do I navigate to my airline's webpage for complaints?), or transactional (Find me a winter coat to purchase online).

By addressing the search intent behind your targeted keyword, you can often improve your rankings, and even appear before the main search results begin.

This position, called Position 0 or a **featured snippet**, is saved for search results that so closely align with a search query that they are the first thing that Google displays, above all other search results. In order to optimize content for featured snippets, it is best to provide an answer to a very specific question on a technically optimized page. Here's an example of the type of content you may include in your blogs, landing pages, and other places on your website:



Aiming for the featured snippet for "inbound marketing"/"what is inbound marketing" by including the search term and offering a clearly identifiable definition as the initial content on the page.

Google will pick up on the fact that your content directly answers a question. Google will reward your information with a featured snippet, or Position 0.

Position 0 can also be occupied by carousels, made up of a series of cards for different items such as recipes, news items, images, and more. These are drawn from structured data on your page.

Finally, structured data can also increase the probability that someone will click on your page by providing Google with information to create a rich snippet for your page's listing in the search result. These results include information like starred ratings, links to sections of your site, or even image thumbnails.

Oncrawl SEO page audit - Chrome Web Store

<https://chrome.google.com/webstore/.../oncrawl.../ngcfjikphejojdjnmiegbcjhemldbjhp...> ▼

★★★★★ Rating: 5 - 4 votes - Free - Chrome

Jan 22, 2018 - The **Oncrawl** extension provides valuable informations on the status of any web page by measuring its content (SEO tags, duplicate and near ...

A rich snippet including ratings, pricing (free), and browser compatibility information.

Pinpoint technical issues hindering UX

Talking about user experience (UX) in a technical SEO audit might seem strange, but there are actually a number of SEO factors that can be hiding in your website design.

Location of money pages

You want your best-performing pages (i.e. those that you most want visitors to see) to be strategically located in your website's structure. You don't want to feature the page to purchase your product or even a strong blog post five layers deep into your website's architecture. Rather, think hard on what your best content is and make sure that these "money pages" are easily accessible from both a user and crawler point of view.

Accessibility

Additionally, your UX could be damaging to other key ranking aspects, such as speed, size, mobile optimisation, local and international information. Pages that pose accessibility issues for users are generally also very difficult for crawlers to reach, read, or understand. Accessibility issues are elements of site design that make viewing or navigating difficult for a group of users: users of mobile devices, users on slow connections, users using visual aids, users from other countries...

Make sure that your design looks good, but that it also is fast to load and has a small overall file size. Creating a design that is mobile adaptive will also increase your search rankings. Further, having specific versions of your website designed for local versus international search, typically provided in the header or footer of your website, will help users and crawlers navigate to the right areas of your website and have a stronger understanding of your website structure.

Strategic business goals

Don't forget about your own strategic goals. Remember to prioritize technical improvements that offer the quickest wins for your business. You might want to drive more traffic to your website, generate more leads for your sales team, or complete more e-commerce sales. For example, if you're a local restaurant, you might want to skip investing in an international site, but improve mobile load times and structured data so that people on the road searching for somewhere to eat nearby will have better access to your site.

Whatever your business goals, match your SEO efforts to align with this.

Essential tools for technical auditing



Essential tools for technical auditing

Conducting a technical SEO audit would be impossible without the right tools. Here's a list of tools to use in conjunction with this ebook to conduct the perfect SEO audit.

1. [OnCrawl](#): In order to understand how Google crawlers will index your site, you have to use your own crawler first. Paired with an analysis log data from your server logs, this gives you a complete and accurate picture of what is happening on your website. OnCrawl is a data-driven SEO crawler and log analyser that helps you make smarter SEO decisions.
2. [Google Search Console](#): Built by Google, Search Console tools and reports help you measure your site's traffic and performance and fix issues. You can use this separately, or, allow OnCrawl to combine your GSC data with crawl data for additional insights.
3. [Google Keyword Planner](#): Identify potential keywords and their corresponding traffic with this free keyword planning tool from Google.
4. [HubSpot Website Grader](#): Use this website grading tool to determine how strong your website is and quickly find simple SEO changes to significantly improve your website.

When focusing on a specific issue, you'll find many additional tools that help you check details and validate implementations. Both HubSpot and OnCrawl recommend using other tools obtain the best understanding of your website.

In the list below, you'll find a few of the extra tools we find useful when running a technical audit. BONUS: unless noted, all of the tools in this list are free!

General tools:

1. [OnCrawl Free SEO Toolbox](#)
2. [QuickSprout Website Analyser](#)
3. [Moz Toolbar](#)
4. [AccuRanker guide](#)

Indexation tools:

1. Robots.txt tester: <https://technicalseo.com/seo-tools/robots-txt/>
2. XML sitemap generator: <https://technicalseo.com/seo-tools/xml-sitemap/>

Server performance tools:

1. HTTP status checker: <https://next.httpstatus.io/>
2. Site speed test: <https://tools.pingdom.com/>

Site structure tools:

1. Backlink analysis: <https://majestic.com/> (paid)

Tools for keyword performance:

1. Research and tracking: <https://www.semrush.com/> (paid)
2. Research: https://ads.google.com/intl/en_ie/home/tools/keyword-planner/
3. Research: <https://keywordtool.io/>
4. Research: <https://topics.seomonitor.com/query>

Tools for on-page optimization:

1. Page speed: <https://developers.google.com/speed/pagespeed/insights/>

Tools for duplicate content:

1. Internal duplicate content: <http://www.siteline.com/>
2. External duplicate content: <http://www.copyscape.com/>

Tools for mobile issues:

1. Mobile friendly test: <https://search.google.com/test/mobile-friendly>
2. AMP validator: <https://validator.ampproject.org/>

Hreflang testers and validators:

1. <https://technicalseo.com/seo-tools/hreflang/>
2. <https://www.aleydasolis.com/english/international-seo-tools/hreflang-tags-generator/>
3. <http://hreflang.ninja/>

Tools for structured data:

1. Validation tool (Google): <https://search.google.com/structured-data/testing-tool/u/0/>
2. Rich result test (Google): <https://search.google.com/test/rich-results>
3. Validation tool (Bing): <https://www.bing.com/toolbox/markup-validator>

Indexing

4

Indexing

Technical SEO can help you set limits on what can be seen by Google and indexed, and what can't.

One of the first goals of a technical audit is to verify that the appropriate pages of the website are **indexed**, or picked up by a search engine for use in search results, and that elements of your site that do not need to appear (or shouldn't appear) in search results are left alone.

Pages that are or that can be visited by the search engine's bots are **crawlable**. Being crawlable is only the first step. To be **indexable**, pages must also allow search engines to index them, must be available when requested from the server, and not redirect visitors to another page.

Indexed pages are available to be used in search results if the search engine's algorithms determine that the page is a good match for the search query. In other words: **in order to rank in SERPs, your pages must be indexed**. Unindexed and unindexable pages cannot appear in search results.

Which content is indexed?

A technical SEO audit is a process of analyzing and tracking key website performance metrics that impact all areas of your website. The goal of a technical SEO audit is to identify the shortcomings of your website's structure that negatively affect its performance and thereby hurt its SEO and rankings in search.

Who should audit?

The most reliable way to know which pages have been indexed by a search engine is to use information provided by the search engine itself. Google Search Console and Bing Webmaster Tools will both provide this information.

Another way to check the number of pages indexed by Google for your website by going to Google and entering "site:https://www.example.com". Replace https://www.example.com with the homepage of your site.

However, you can also gain further insight by tracking the pages that have been visited by a search engine's indexation bots using your log files. In the case of Google, track hits by googlebot-desktop and googlebot-mobile, and remember to verify the identity of the bots by looking up their IP address.

Compare the results with a breakdown of your website based on the type of URL.

Remember that pages are not the only thing on your website. Your server also provides--and search engines and users also view--other types of files. Together, these files are generally referred to as **resources**. They might include:

- PDFs
- Media files: images, videos...
- Resources that load with the page: Javascript files, css files, font files...

Sometimes you will absolutely want a crawler to index resources, such as images you want to appear in an image search, or Javascript that injects content into your page.

However, bots don't need to see or index files like fonts, default images provided by WordPress, or css libraries in your template that you don't use.

Do index:

- Landing pages
- Money pages
- Home pages
- JavaScript files that are necessary to see the page content
- Images you want to appear in image searches
- Translated pages
- Seasonal or temporary promotion pages
- News
- New pages
- Pages that old pages have been redirected to

Don't index:

- Pages with duplicate content that aren't the "main" page you want users to see
- Your content management system administration pages
- Previews of draft pages
- Files that should only be downloaded after filling out a form
- URLs for pages that have the same content, just sorted or filtered using queries
- Blank pagination or blank calendar views
- Confirmation pages for signups, downloads, purchases...
- Pages that are no longer available
- Pages that have been redirected

To inspect how your site presents pages for indexing and how it withholds pages from bots, you will also need to audit the elements below.

Sitemaps

Sitemaps in xml format are a way of providing a complete list of the pages you want a search engine to know about on your site. A search engine will attempt to visit all pages that appear on your sitemap. However, not putting a page on your sitemap does not prevent a search engine from learning it exists if there are links to it from other pages.

Your sitemap should be formatted according to the sitemap protocol, which you can find at <https://www.sitemaps.org/protocol.html>, and placed at the root (not in a folder) of the part of the website it covers.

As part of your audit, verify that your sitemap contains the URLs you want to index, and does not contain URLs you do not want to be indexed.

You should update your sitemaps regularly and submit them to search engines.

Robots.txt

Robots.txt files are text files that contain instructions to bots. They are the first thing a bot visits on a website. Rules in this file can be directed at all bots, or only at some.

A robots.txt file applies to the domain or the subdomain in which it is located, as well as to content in folders within that location. It does not apply to other subdomains, or to other domains on the same server.

An example of robots.txt rules might look like this:

```
user-agent: googlebot-news
disallow: /

user-agent: *
disallow: /wp-admin

user-agent: googlebot
allow: /blog
```

Each rule specifies the user-agent, or type of visitor, to which it applies. It then allows or disallows access by that type of visitor to a group of files, pages, or directories.

Questions to ask when examining a robots.txt file:

- Do search engines' indexation bots see the same pages as human visitors?
- Can search engines' indexation bots access pages that should be indexed?
- Are search engines' indexation bots forbidden from accessing pages that should not be indexed?

Meta-robots instructions

In the <head> tag of each page, you can specify instructions to robots.

Like robots.txt files, these instructions can be general, in the case of

```
<meta name="robots" content="noindex" />
```

or specific to a given bot, such as

```
<meta name="googlebot" content="noindex" />
```

The following are valid content instructions:

all	No restrictions. This is the default value.
noindex	Do not index this page. It will not be part of the list of pages known to Google, and will not appear on SERPs.
nofollow	Do not follow any links on this page.
none	Same as noindex, nofollow.
noarchive	Do not save an archived version available to Google users.
nosnippet	Do not display a preview (an extract from the page or from the video) in the SERPs.
notranslate	Do not offer a translation of this page in SERPs.
noimageindex	Do not index the images on this page.

Questions you should ask when examining a meta-robots strategy:

- If some pages have meta-robots instructions, do they allow bots to index pages that should be indexed?
- If some pages have meta-robots instructions, are the pages that tell bots not to index really pages that should not be indexed?

Rel="alternate" / rel="canonical"

In the <head> tag of each page, you can use the <link rel="alternate"> and <link rel="canonical"> declarations to suggest to search engines that a page has multiple versions, and which of the versions should be indexed.

This can be used when you have a separate mobile version of a page. In this example, use a canonical link tag on the mobile version, and an alternate link tag on the desktop version. This prevents these pages to be treated as duplicate content and listed as competing against one another for rankings on the same keywords.

Redirections

Another way to send a message to the search engine to not index pages is to instruct your server to return an HTTP code to the bot indicating that the pages have been replaced, either permanently (301) or temporarily (302), or that it is no longer available (410).

HTTP codes, particularly redirections, will be discussed in the next chapter on server performance.

Requesting indexing

Most search engines allow you to submit URLs that you would like them to index. Until recently, this function was open to anyone. Increasingly, though, it is becoming reserved to webmasters, who can submit URLs for their own sites only. Between July and September 2018, both Bing and Google retired their public submission forms.

Since then, Google requests that webmasters submit fetch request for an individual URL or submit a sitemap in xml format containing the URLs they are requesting indexing for.

Fixing problems

If you have discovered indexing errors, list them for correction.

Remember the following techniques:

Getting pages indexed:

- Sitemaps
- Allow lines in robots.txt
- Requesting indexing by search engines

Refusing indexation:

- Disallow lines in robots.txt
- Noindex meta-robots on page
- Rel-alternate on page
- Page redirections (301, 302, 410)

Sources:

Verifying Google's crawlers: <https://support.google.com/webmasters/answer/80553>

Robots.txt: https://developers.google.com/search/reference/robots_txt

Meta tags for robots: https://developers.google.com/search/reference/robots_meta_tag

Sitemaps: <https://www.sitemaps.org/>

Getting your site indexed: https://support.google.com/webmasters/answer/34397?hl=en&ref_topic=3309469

Submitting URLs: <https://blogs.bing.com/webmaster/september-2018/Anonymous-URL-Submission-Tool-Being-Retired>

Submitting URLs: <https://developers.google.com/search/docs/guides/submit-URLs>

Server performance



Server performance

Server performance is one of the factors that has the most impact on your rankings, for two extremely important reasons. First, Google has recognized that server performance influences crawl frequency and site authority. And second, sites that are perceived by users as slow or buggy will be clicked on less, or will promote “pogo-sticking”, in which a user returns to the search results using the back button in order to choose a different result.

Fewer clicks and pogo-sticking don’t only mean fewer visits. They also mean losing your spot in the SERPs: these types of user behavior are also ranking factors.

Auditing server performance aims to verify that:

- Your server is up and available (no 5xx errors).
- Your server uses appropriate security protocols (HTTPS).
- Your server responds to requests in a timely manner (page speed).
- Your server provides page content correctly (page speed and page status).
- Your server’s errors are user-friendly (custom 404 pages).
- Your server provides the same page content to visitors and to bots, as required by Google.

A good web-based crawler will provide most of the information you need to audit your server. To compare results provided to bots and results provide to humans, though, you are best served by the information in your server’s log files. Web crawlers intended for use in SEO audits, like OnCrawl, will provide this functionality in addition to crawls.

HTTPS

HTTPS is a standard for sending information over the internet--such as requests, URLs, and the contents of URLs--that protects the data sent during communications between a visitor and a server.

HTTPS provides protection through encryption of transferred data, checks on the integrity of transferred data, and authentication that the server responding is who it claims to be.

This communication protocol has become the expected standard, and less secure protocols, such as HTTP, can penalize your website.

Your audit should determine whether or not your site uses appropriate standards and protocols.

If your site has not migrated to HTTPS yet, Google provides a [migration guide](#).

Speed

The speed of your website is a ranking factor, particularly for searches run from a mobile device. In short, if two sites are identical except for site speed, the faster site will rank better in SERPs.

The goals of evaluating page speed during a technical SEO audit are to determine the threshold page speed at which bot and organic visits drop off, to produce an exhaustive list of the pages below this threshold, and to identify the factors causing these pages to load slowly.

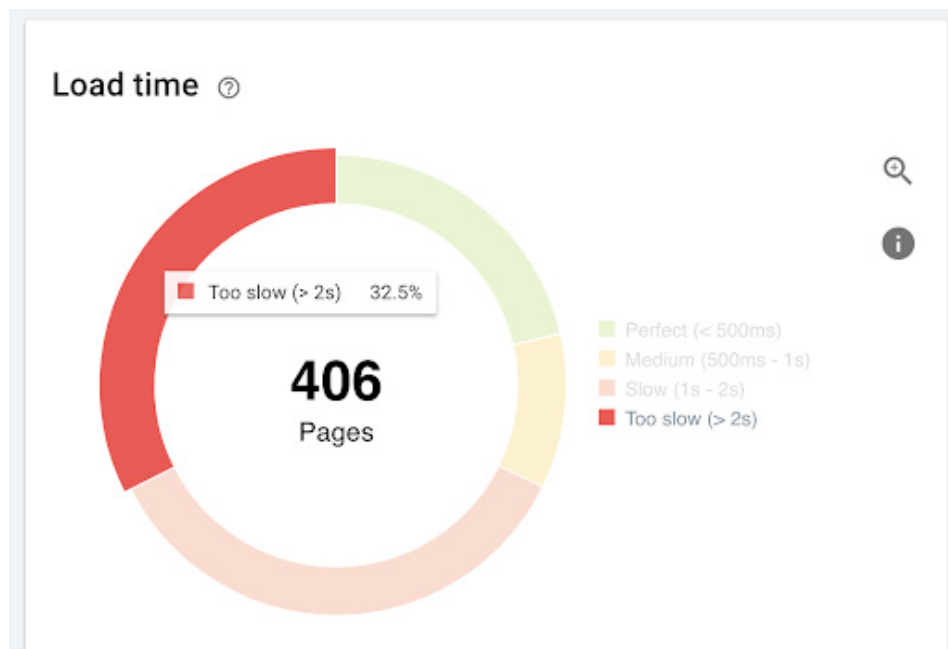
By cross-analyzing your log files and crawl results, you can obtain direct answers to the first two questions, as well as collect the necessary information to answer the third.



The slower a page is (more red), the more likely it is to have fewer organic visits (bars further left). Source: OnCrawl and log data.



Generally speaking, the faster a page loads, the more frequently it is crawled. Improving pages slower than 1s is more worthwhile for this side than increasing the speed of pages that already load in less than 1s. Source: OnCrawl and log data.



406 pages ADD COLUMNS

FULL URL	DEPTH	LOAD TIME (IN MS)	LOAD TIME EVALUATION
https://...	2	2.520	Too slow (> 2s)
https://...	2	2.567	Too slow (> 2s)
https://...	2	2.609	Too slow (> 2s)
https://...	2	4.565	Too slow (> 2s)
https://...	2	3.789	Too slow (> 2s)
https://...	2	3.690	Too slow (> 2s)

Accessing a list of pages determined to be too slow by a crawl. Source: OnCrawl.

Assuming your server is responding correctly, the factor with the greatest impact on load time is page size. The length of the page and all of the resources required to produce the content contribute to the size of the page. These resources include images, videos, embedded PDFs, css, fonts, and JavaScript that must be loaded in order to display the final page.

Once you've identified pages that are too large, you can reduce them using the following strategies:

Content issues:

- Compress HTML
- Minify JavaScript and CSS
- Reduce image sizes
- Move scripts that prevent page rendering to a later point in the loading process, if possible

Caching:

- Use a browser caching
- Use a CDN (content discovery network)

Server response:

- Improve server reactivity
- Reduce the number of redirections

Page status

Every request for a URL is answered by the server with a page status that indicates the availability of the content for that URL. Missing pages (404) and server errors (503, for example) are indications to search engines that your site's information is not always available as a response to a user's search. Your rankings for a given page can be negatively impacted by persistent status codes in the 4xx and 5xx series.

Sometimes status codes other than 200 ("ok") are intentional, for instance when you've modified the URL of a page and used a 301 status code to redirect to the new URL. 3xx codes provide critical information to search engines when you have made changes to your website. Google has clearly stated that 3xx codes pass along "link juice" and page rank to the URL they redirect traffic to, which should ultimately replace the old URL in the SERPs.

The most common HTTP status codes you will see include:

2xx status codes: okay

200 OK

3xx status codes: wrong location

301 Moved permanently

302 Moved temporarily

304 Not modified

307 Temporary redirect for resources

4xx status codes: client (browser) error

400 Bad request

401 Unauthorized

404 Not found

410 Gone or removed

5xx status codes: server (website) error

500 Internal server error

503 Service unavailable

550 Permission denied

Your audit should discover:

- All pages that cause server errors (5xx). These errors should be corrected.
- All pages that return 4xx errors. While these pages do not necessarily negatively affect your SEO, they frustrate users. They should be corrected or redirected with a 3xx status code.
- All pages redirected using a 3xx code, and the page to which they are redirected.

Redirect chains and loops

Series of site modifications can sometimes produce chains of 3xx codes where page A is redirected to page B, which is redirected to page C...

After 5 links, Google will no longer crawl the redirections, and after approximately 20 links, a browser will display an error to human users, though this number varies from browser to browser.

Certain chains loop back on themselves: page C is redirected--not to page D, but back to page B. This creates an never-ending loop.

A crawl of your website including all pages with a 3xx status will allow you to spot redirect chains and loops.

Your audit should:

- Identify all pages that are part of redirect chains and loops
- Identify all loops, and all chains longer than 5 links
- Identify links pointing to pages with a 3xx status in a redirect chain. These should be modified to point to the final page in the chain.

Custom 404 page and soft 404

404 errors are impossible to eradicate completely: even if all of your links and pages work perfectly, a user might still type a URL incorrectly, or when you modify a URL, you might be unable to get an external link updated. Custom 404 pages help keep users and bots on your site, even when an error occurs.

However, avoid redirecting to or dynamically replacing content with custom “empty” pages, “error” pages or unrelated pages. This is called a “soft 404.” When this happens, the user sees what looks like a 404 or error page, but the browser (and any bots) are told that the URL functions correctly. These waste a search engine’s crawl energy and frustrate users.

You audit should check that a custom 404 page containing at least one link to keep users on your site is present and functional on your website.

Sources:

HTTPS: <https://support.google.com/webmasters/answer/6073543?hl=en>

Page speed (2018): <https://webmasters.googleblog.com/2018/01/using-page-speed-in-mobile-search.html>

Page speed (general): <https://www.oncrawl.com/oncrawl-seo-thoughts/page-speed-how-it-impacts-your-seo-and-how-to-improve-it/>

Speed benchmarks (2017): <https://think.storage.googleapis.com/docs/mobile-page-speed-new-industry-benchmarks.pdf>

Site structure



Site structure

Site structure is an important element in a technical SEO audit, particularly given the importance of site structure on how Google sees your website.

Site structure includes how your URLs are composed, how pages are linked within your site, and how a user (or a search engine) understands the purpose of your site, its subjects and its categories.

<https://www.nytimes.com/2018/09/26/technology/personaltech/the-sun-sets-the-wind-dies-but-energy-data-is-relentless.html>

We know from many sources, including the Google Webmaster Guidelines for how to build a Google-friendly site, that Google analyzes your site's architecture. For example, Google advises you to "Design your site to have a clear conceptual page hierarchy."

Google uses what it understands about the organization of your site to decide what your site is about. Google may use your site structure to offer thematic category pages to users in search results, such as in this SERP listing for the European Union's official site.

EUROPA - European Union website, the official EU website

europa.eu/ ▼

European Union website - EUROPA is the official EU website that provides access to information published by all EU institutions, agencies and bodies.

Results from europa.eu



European Union

About the EU - EU law - Countries - EU
Newsroom - Brexit - Contact

European Commission

European Commission. български
čeština dansk Deutsch eesti ...

Countries

28 EU member countries, candidate
countries and other ...

About the EU

What the EU does - its aims and
challenges.

European Environment Agency

About us - Air pollution - Data and
maps - Countries - ...

Living in the EU

Key facts and figures on living in the
EU: size and population ...

Your Europe

Information, help and advice on EU
rights for EU nationals and ...

VAT Number Validation ...

VIES VAT number validation. You can
verify the validity of a VAT ...

EURES - The European job ...

EURES - The European job mobility
portal - European ...

EU Newsroom

Press releases | Newsroom -
Highlights - Audiovisual - ...

Additionally, pages within your site that link to one another transfer popularity and authority. A strong internal linking structure will indicate to Google which pages you consider most important. Strategies based on internal site structure have been proven to be successful in getting certain pages on your site ranked higher; consequently, they are often used to promote a given set of pages during seasonal or other temporary periods.

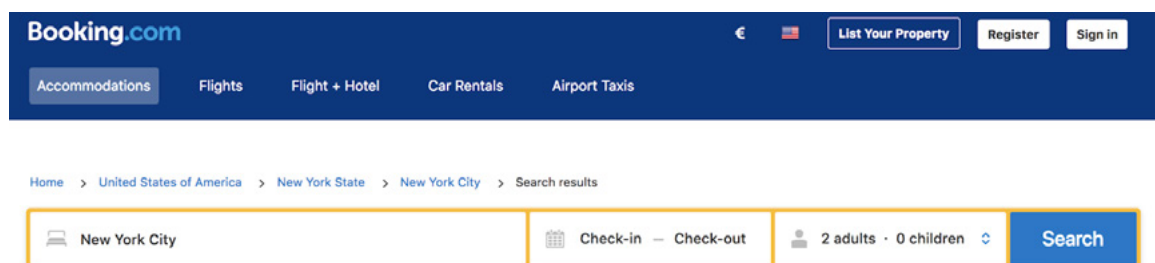
Your audit should look at different elements of site organization in order to determine the current structure and to spot opportunities to improve it.

Breadcrumbs

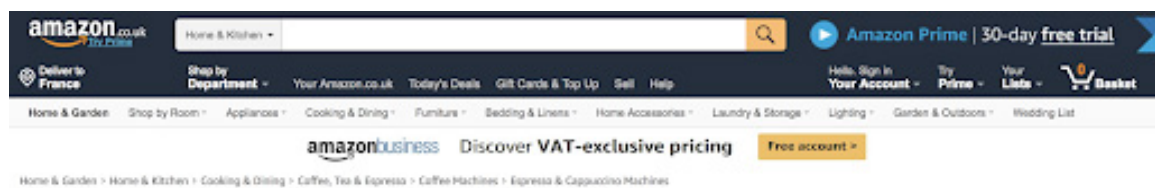
Breadcrumbs are a text-based list of increasingly narrow categories leading to the current page, usually found below the page header, at the top of the page content.

In many cases, they serve as a type of navigational structure, though they do not have to reproduce your menu navigation.

They can also be used to show a hierarchy of attributes leading to the current page, such as on Booking.com, where accommodation pages are classified by location, despite it being very difficult to follow a navigational path from the homepage, to US accommodations, to those in New York state, then New York City.



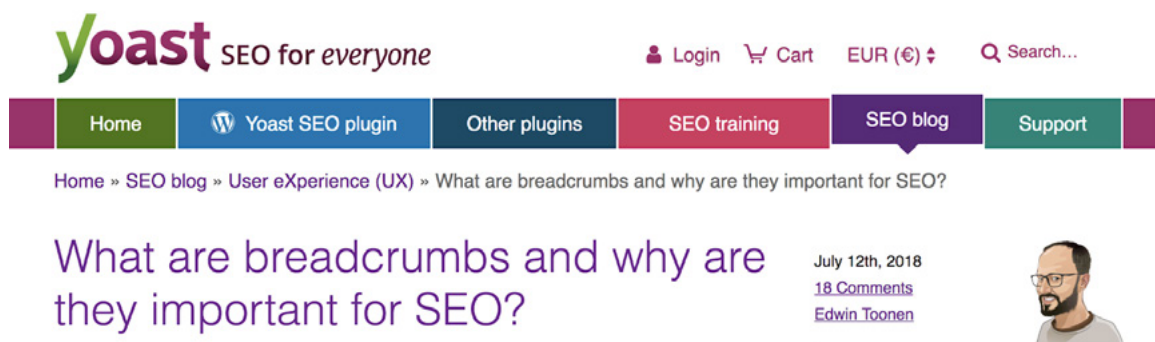
This type of breadcrumbs is often found on e-commerce sites, where the attributes are often product categories, such as in this case on amazon.co.uk.



Breadcrumbs are useful for SEO because they present important structural categories or attribute categories for your site, which helps search engines understand your content. They are also presented in a structured fashion, making them easy for search engines to find and understand.

In addition, they provide a better user experience, keeping users on your site and increasing a search engine's confidence that your site answers user questions. This is even more true for mobile users, where breadcrumbs are often easier to see and follow than large main navigation menus.

You can also use structured data to display them, which allows Google to display them in the SERP snippet for your page, as has occurred here for the navigational breadcrumbs in the Yoast post on breadcrumbs:



Structured data will be discussed further in the chapter on Schema.org.

Breadcrumbs are particularly useful in these types of sites:

- e-commerce websites
- large blogs
- knowledge bases
- large or complex sites
- web applications

Your audit should identify where breadcrumbs are used on your site in order to answer the following questions:

- Are breadcrumbs used on your site?
- Are breadcrumbs used consistently?
- Are breadcrumb category names short and easy to understand?
- Are breadcrumbs used in all appropriate places?

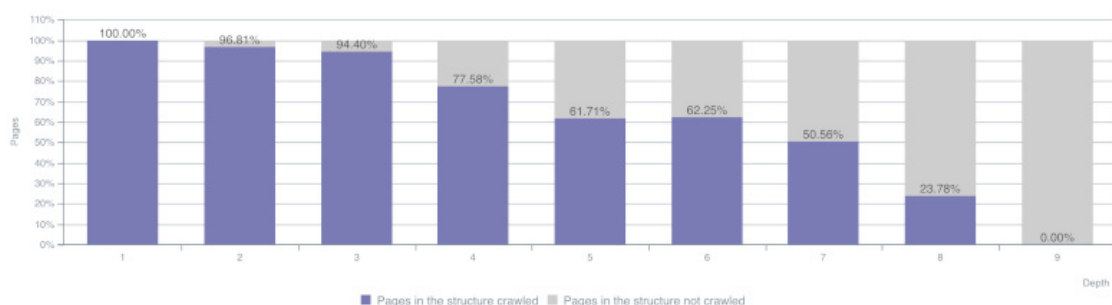
Page depth

Page depth refers to the number of links a user has to follow to reach a given page. This is another concept that impacts SEO both from a UX perspective and from a search engine's perspective.

You can get an overview of your users' behavior in Google Analytics, by looking at Behavior > Behavior Flow. You'll see 100% of sessions on the starting page, but perhaps only 8% on the second step, 3% on the third step, less than 1% on the fourth step, and so on. The more a user has to click, the less likely they are to arrive at the final page.

This is also true for search engine's web crawlers. In an analysis of hundreds of websites, OnCrawl has found that the further the page from the homepage, the less Google crawls it. This has since been confirmed by Google: distance from the homepage affects page importance, which is used as a ranking factor.

Pages crawled and not crawled by depth ⓘ



Your audit should determine:

- How many and what type of pages are found at which depth
- Whether page depth negatively impacts your pages' indexation and ranking by comparing the number of pages crawled to their depth, and the number of ranked pages to their depth.

Internal linking strategy

If your audit finds opportunities for improvement, your linking strategy will come into play.

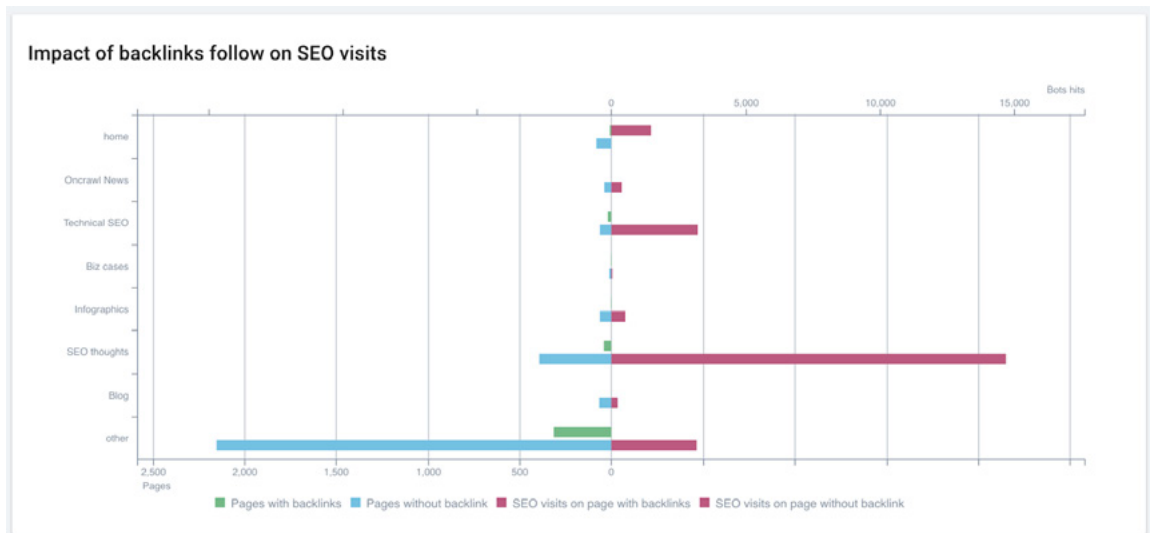
Backlinks

Backlinks are links to URLs on your site from external sources, that is, other sites.

Backlinks are used by search engines to establish the authority of a website on a given subject. Because authority is increasingly important in order to rank for competitive keywords, collecting legitimate backlinks from reliable, on-topic sources is an essential part of a comprehensive SEO strategy.

These links bring a boost of confidence and traffic to your site. They should land on pages that are part of your site structure (linked to by another page in your site), and they should distribute their "link juice" to other related pages on your site (by linking to them). Pages that receive backlinks should give a good impression of your site: they should load correctly and at an acceptable speed.

Cross-analysis between crawl data and data from a backlink tool will provide insights for your audit. This can include the impact that backlinks have on organic traffic, on bot behavior, and their distribution across your site according to factors such as types of pages or page depth.



Using OnCrawl and Majestic to view the relationship between backlinks and organic traffic.

You should also note that spammy backlinks (backlinks from bad, dangerous, off-topic sites) can hurt your SEO.

Your audit should aim to discover:

- What pages receive backlinks?
- Are the pages that receive backlinks functional:
 - Is their HTTP status 200?
 - Do they receive links from other pages in your site?
- What sites link to yours?
- Is there a correlation between pages with backlinks and performance?
- Do internal links transfer the boost received by your page from a backlink to additional pages on your site?

Internal links and popularity

As mentioned in previous sections, internal links, or links within your site, are the tools to create site architecture and to transfer popularity and confidence from one page to another.

The best practices for internal linking are as follows:

- “Limit the number of links on a page to a reasonable number (a few thousand at most).” (Google Webmaster Guidelines) Don’t forget that this includes all of the links in your header, menu, sidebar, and footer.
- Link from main pages (such as your homepage) to category pages, and from category pages to all of the pages in that category, creating “hubs” of content by type, theme, or category.
- Create links between categories.
- Link from popular pages that receive a lot of visits to distribute popularity.
- Link to pages whose popularity to want to promote.

Your audit should tell you:

- What groups of pages link to what other groups of pages?
- Which groups of pages receive the most links?
- What is the relationship on your site between links to the page (internal links) and bot hits? Use the optimal profile to improve pages that underperform.
- What is the relationship between internal links and organic visits? Use the optimal profile to improve pages that underperform.
- Are there pages on your site that receive too many links?
- Are there pages on your site that receive no links?

Broken links

Broken links are links that do not allow a user to arrive at their intended destination. This is usually due to a page status of 4xx (such as in incorrect URL or a missing page), or 5xx (server errors).

Above and beyond the UX issue that this represents, search engines will demote and eventually deindex pages that can't be found: these pages don't provide answers to users' search queries.

Some SEOs will also consider a 3xx status as an indicator of a broken link: although the redirection helps get the person where they intended to go, the link isn't doing the job by itself. Google doesn't index a page that it knows only as the target of a 3xx redirection, either, so it's also good practice to have internal links point to the updated page with a 200 status. This ensures that the updated page that you're sending users and bots to gets indexed.

Your audit should create a list of links to fix by listing:

- Internal links leading to a page with a 4xx or 5xx status
- Internal links leading to a page with a 3xx status

Subdomains

For the purposes of crawls, Google officially treats subdomains (<https://subdomain.example.com/page/>) and subfolders (<https://www.example.com/subfolder/page/>) the same.

Many SEOs prefer using subfolders because, outside of crawls, Google does consider separate subdomains to be their own "entity". SEOs perceive it to be more effort to maintain and promote rankings for subdomains, which may be treated as separate sites. This means they'll be ranked separately, and influence won by one subdomain won't necessarily transfer over to another.

Others prefer to use subdomains, notably when dealing with vastly different types of content.

Your audit should treat subdomains as part of the site you are auditing unless you have a good reason not to. During your audit you should explore the same topics for subdomains as for the main site.

Orphan pages

Orphan pages are pages that are not included in the structure of your site. They have no “parents”, or pages that link to them.

This means that they don’t benefit from any of the signals that a website can give to a page to show it’s important, such as placing links to it from important pages, increasing the number of links to it from pages with similar themes, or allowing users to reach it from the homepage in only a few clicks. Consequently, orphan pages can never rank as well as they should.

Orphan pages are quick wins that allow you to improve a page’s ranking and performance by integrating it into the site’s structure.

Orphan pages can be created, for example, when websites are restructured, when pages expire, or when pages or links are deleted outside of an audit.

Your audit should:

- Identify all orphan pages by comparing crawled pages
 - to indexed pages
 - to pages that receive visits, and
 - to pages that are listed in the website’s sitemaps.
- List **active orphan pages**: orphan pages that are ranked and receive organic visits. These pages already perform relatively well and can benefit most from improvements.

Sources:

Webmaster Guidelines: <https://support.google.com/webmasters/answer/35769>

Breadcrumbs: <https://developers.google.com/search/docs/data-types/breadcrumb>

Page depth: <https://www.oncrawl.com/technical-seo/perfect-page-depth-level/>

Internal linking: <https://research.hubspot.com/topic-clusters-seo>

Subdomains: <https://youtu.be/uJGDyAN9g-g>

Orphan pages: <https://www.oncrawl.com/deal-orphan-pages/>

Keyword Performance



Keyword performance

Understanding your keyword performance is a crucial part of conducting a technical SEO audit. The biggest success factor in keyword performance is the amount of organic traffic that is reaching your web pages. To understand how your keywords are performing in terms of organic traffic, you'll want to dig into your keyword research and your ranking URLs to identify the technical characteristics of your most successful pages and those of the areas of your website with room for improvement. When a technical issue, such as page speed beyond a certain threshold, affects nearly all of your pages that need improvement, correcting this will produce impressive results.

But to spot the technical issues related to keyword performance, you need to target appropriate keywords and follow your rankings for selected keywords.

Let's get started by looking at keyword research.

Keyword research

Whether you're at the early stages of creating a business or are already an established brand, conducting keyword research is fundamental to successful technical SEO. Your keywords will evolve as your business does, so continual evaluation over time is necessary and should be included in the regular technical audit of your website SEO.

Here is a [6-step process](#) to conduct keyword research:

1. Make a list of important and relevant topics based on what you know about your business.
Think broadly about your business. What topics do you find yourself consistently blogging about? What topics come up in conversation when discussing your business? Use these questions to help guide you to create umbrella terms that are related to your company.
2. Fill in those topic buckets with specific, related keywords.
These are keyword phrases that you think are important to rank for in search. What search queries do you want your business to come up for in the SERPs (search engine results pages)?
3. Research related search terms.
Fill out your list with additional, related keywords by using Google search or other keyword explorer tools.
4. Check for a mix of head terms and long-tail keywords in each bucket.
Head terms are broad (and likely popular) keywords, while long-tail keywords are elongated or related versions of the head term. Long-tail keywords typically include more words and are easier to rank for due to less competition. [Here is an example.](#)
5. See how your competitors are ranking for these keywords.
There are two key things to look out for here. Keywords that your competitors are ranking for are important for you to also pursue. However, keywords on your list that your competitors aren't ranking for can provide a great opportunity for you. Use tools such as [SEMrush](#) to run competitive reports.
6. Narrow down your list with [Google Ads Keyword Planner](#).
Take a look at the estimated search volume for the potential keywords on your list to identify ones that have way too little (or way too many) searches.

Now that you have a comprehensive list of keywords, you'll want to see how your web pages are currently ranking for your chosen keywords to find your website's strengths and opportunities.

Ranking URLs

URLs that rank, or appear in position in the SERPs, are a confirmation that your SEO strategy works for the keywords you've chose. The ranking for each URL is its SERP position (first, tenth, or twenty-third result, for example) on a given day.

You should keep track of how your website pages rank for each of your keywords. This will help you to identify which URLs are performing well and show you where you can improve. Shifting your technical SEO efforts to keywords that are underperforming will help you increase your ranking in organic search.

You can also use basic keyword ranking data from Google Search Console to understand how different technical factors influence your keyword performance.

Your audit should examine potential relationships between keyword performance and factors such as:

- Type of page on your site, whether a product page, a blog page, a dedicated landing page...
- Keyword length
- Branded vs. unbranded keywords
- Page depth, or distance from the homepage in number of clicks
- Number of followable internal links pointing to the page
- Page load time
- Number of words
- Title tag optimization
- Measures of internal popularity, such as OnCrawl's Inrank
- Length of meta description
- Use of structured data
- Crawl frequency

Organic traffic

You should also keep track of how effective your ranking URLs are at drawing organic traffic to your site. Ordering your URLs by their organic traffic is the best way to determine the success of your keyword strategy. You can do this directly in OnCrawl using data on organic visits taken from your log files.

You can also do this by hand if you have Google Analytics on all of your ranking URLs. Here's a step-by-step list of how to order your URLs using Google Analytics.

1. Open your Google Analytics account.
2. To display your organic traffic, select Acquisition > Campaigns > Organic Keywords
3. Adjust the number of keywords and the period of time of which you'd like to analyze. (Example: 30 days, 500 rows)
4. To display the keywords that are being displayed for these landing pages, select Secondary Dimension > Keyword
5. Click Export to export this data into an Excel or Google spreadsheet.
6. Delete the unnecessary columns. You'll want to keep Keyword, Landing Page, Sessions, Pages/Visit, and Bounce Rate.
7. Add columns for Potential Sessions and Performance.
8. Add filters to your header column and filter your keywords by sessions.
9. Open [Google Ads Keyword Planner](#)
10. Copy your keywords in column one of your spreadsheet and paste them into the keyword tool in the same order.
11. Change the matching type to [exact match] in the keyword tool.
12. Export the results by clicking Download > All Search Results > CSV
13. Import this dataset into the Potential Sessions column of your spreadsheet.
14. Set the Performance column to equal Visits/Potential Sessions and convert to a percentage.
15. Analyse the Performance column to see how much of the potential traffic you're driving to your website for each keyword.

Your audit should:

- Indicate the organic visits for each ranking URL.
- Be able to order pages that draw the most visits to pages that draw the fewest.

Sources:

Keyword Research: <https://blog.hubspot.com/marketing/how-to-do-keyword-research-ht>

Head vs Longtail Keywords: <https://research.hubspot.com/charts/example-of-head-term-vs-long-tail-keywords>

On-page optimization



On-page optimization

On-page optimization covers all of the actions that can be carried out on a given page to improve SEO performance. These are issues that affect each page individually, rather than the site as a whole.

On-page elements include the page content itself as well as meta elements in the page's source code that are used by search engines to obtain information about the page.

HTML tags used in SEO

HTML tags are used to structure your content and help crawlers and humans understand it.

The most important HTML tags for SEO currently include the meta tags used in SEO (including Title Tag, Meta Description Attribute, Meta Robots Attribute) and headings (H1, H2...).

Title tag

Title Tags are the most important tags to consider when creating and auditing your on-page content. Titles were originally developed to label browser windows and tabs. Today, title tags still appear in the label of the browser tab and are therefore one of the few SEO meta tags that are seen by your audience. Title tags also have a large impact on your search engine results and page ranking.

Here are the best practices:

- Include the main keyword phrases you think your audience will type. Even keywords that appear after the cutoff point in your title tag count towards rankings.
- Titles are cut off differently depending on the type of device used by the person searching. In most cases, respecting the old 60-character title limit will allow your title to display correctly almost all of the time.
- Prefer long-tail phrases as you'll have more success optimizing performance.

Your audit should:

- Make sure each page has a unique title
- Make sure the first 60 characters of the title give an indication of what the page is about

Meta description

The description or preview of your page that shows in the SERPs is taken, about a third of the time, from the Meta Description Attribute that you can use to provide a short summary of your content.

Because Google attempts to tailor page descriptions to search queries, the description that is displayed in the SERPs may vary for each search term. When not using the meta description, Google most often displays a relevant sentence from the page content.

Here are the best practices:

- Insert a customized and unique description tag. Otherwise, search engines will most likely display the first sentences of your content.
- Aim for a meta description length of around 70-155 characters, but this may vary as Google's practices evolve. For example, during several months in early 2018, Google accepted descriptions of up to 320 characters.
- Don't use quotes as Google will cut them off. If you must include a quote, use a single quotation mark (') rather than a double quotation mark (").
- Don't keyword stuff.
- Don't use identical or boilerplate text. This makes Google less likely to display your description.
- Include facts about the page content and separate them clearly from one another, even if that means deviating from sentence form.

Your audit should:

- Verify that all of your pages have a meta description
- Identify boilerplate and duplicate meta descriptions
- Find meta descriptions that are too short or that may be too long

Meta robots

Meta Robots Attribute tells search engines what to do with your web pages. There are two types of this tag.

- **"Index/noindex"** tells search engines to show your page in the search results or not.
- **"Follow/nofollow"** tells search engines if they should consider and follow your links to the next page or not.

Here are some of the ways to combine them:

- If you want bots to index the current page and to visit pages that it links to, use "Index, Follow".
- If you want bots to index the current page, but not visit any pages that you link to on it, use "Index, NoFollow".
- If you want bots to not index the current page but still follow links that appear on the page, use "NoIndex, Follow".
- If you want bots to not index the current page and not to follow links to the pages it links to, use "NoIndex, NoFollow".

Your audit has already covered meta robots tags when looking at the indexability of pages, but should now make sure that page tags are used correctly.

Headings

Headings are titles and subtitles of content sections visible on your page. They help guide readers, including bots, through the structure of the content on your page.

- Use only one <h1> per page. This is often used for the title within the content of your page, but it doesn't have to be the same as the <title>.
- Without resorting to keyword stuffing, use keywords and related words in your headers to help Google establish the topic of your content.

URL issues

Most advice on creating URLs is aimed at making URLs easier to read by human beings. This includes using human-readable language in the URL, using dashes as punctuation between words in your URL, adding keywords in the URL, and keeping the URL relatively short.

Although this advice falls under the category of user experience (UX) improvements, it's still advice you should follow for SEO. While they don't make it more likely that your pages will appear as search results, these optimizations increase the probability that a user will click on your URL when it does appear.

In terms of pure technical SEO, though, Google's advice is to:

- Use a simple URL structure.
- Avoid URLs containing multiple parameters, which can unintentionally create large numbers of possible URLs and therefore cause problems for Google's bots. This problem can also occur with [faceted navigation](#), or navigation that allows users to essentially personalize a page through sorting and filtering.
- [Avoid using # in your page's URL](#). The hash character was historically reserved for anchors to locations within the page, and Google still considers pages with identical content before the # to be the same page.
- If you're using a technology such as ajax to generate content, you can [use a hasbang \(!\) instead](#).

Finally, unless you have major issues to correct, it's best to avoid modifying existing URLs unless you absolutely have to.

Your audit should help you answer questions such as:

- Do my URLs use too many parameters?
- Does Google spend too much time crawling my faceted navigation pages?
- Do the URLs I want to be indexed contain a hash (#)?
- Is my URL structure simple and easy to read?

Media size

The media you include on your page will influence the page's performance, particularly in terms of page speed. Media may include images, videos, css, JavaScript, fonts, pdfs, and any other files that are downloaded in order to display the page. A page with many, large additions to its text content will take much longer to load than a page with optimized images and fewer resources.

Typical strategies to reduce page size include:

- Removing resources that are downloaded, but not used
- Compressing data
- Replacing embedded fonts with web fonts
- Various techniques for optimizing images by reducing their size, such as: replacing images with CSS effects, using appropriate image resolutions, compressing images, using the appropriate image format, lazy-loading images, etc.
- Replacing animated GIFs with videos

Your audit should:

- Use data on page size, page speed and page performance to identify at what size the page gets to be too big to perform well. For reference, OnCrawl uses the standard measures for an ideal page size of under 100KB and an ideal load time (TTLB, or the time it takes to load all of the page) of under 500ms.
- Discover particularly large images, videos, and other media files.

Content length

Deciding on the optimal content length for SEO is a long-debated topic. Some have argued that shorter content with only a few hundred words perform best, while others have said that long-form content of 2,000 words or more will perform better. More recently, people have claimed that the number of words in your content doesn't matter.

Your audit should establish whether or not this is true for your website by determining, for each type of content on your website, whether a relationship exists between the number of words on a page and:

- Page depth
- Instances of ranking pages in the search results pages.

In reality, every website and each type of content is different. The optimal number of words that should be included in your on-page content for each type of page can be determined by looking at the pages that perform best for each type of page on your site.

Regardless, you should always aim to create content that is well written and serves the purpose for which you are writing for.

Here are three key takeaways to consider when writing on-page content:

- Don't be afraid to write short- and long-form content. It is most likely that both have a place within your blog or website.
- Follow principles of good writing:
 - If you can write something more concisely, do! Don't over-complicate your writing just to add to your word count.
 - Do some research on blogs or websites that you admire and assess how their content compares to yours. Try experimenting with new ideas and track how they perform.
- Use your audit to measure the impact of word count and adjust your content based on the result.

Finally, keep in mind that the quality of the words you write and how well you use HTML tags to optimize this content for web crawlers play a significant role in how your content performs.

Sources:

Meta descriptions: https://support.google.com/webmasters/answer/35624?visit_id=636758086798903067-3237243957&rd=1

Meta tags: <https://www.oncrawl.com/oncrawl-seo-thoughts/meta-tags-impact-on-your-seo-and-good-practice/>

Info specific to title tags: <https://www.oncrawl.com/oncrawl-seo-thoughts/12-things-might-not-known-title-tags/>

Content length: <https://blog.hubspot.com/marketing/how-long-should-your-blog-posts-be-faq>

Duplicate content



Duplicate content

Duplicate content is page text that appears identically on different URLs.

This can occur on two different pages when:

- your content is too thin compared to your boilerplate (header, footer, sidebar, and other page structure information), although many search engines reduce the weight carried by these sections and concentrate on page-specific content.
- you reuse standard text as headers, page titles, or meta descriptions, making the search results snippets for different pages look identical, even though the pages are not.
- you have two identical versions of a page for historical or testing reasons.
- two similar pages, such as the product page for a 15mm nail the product page for a 17mm nail, contain minimal differences. In this hardware example, the only difference might be the millimetres indicated in the product title.

Finally, even if you haven't reused content elsewhere, duplicate content can also be produced by a CMS (content management system, such as WordPress), which may have multiple ways of addressing the same page.

It's important to remember that, for a search engine's purposes, a single page that can be reached at the four following URLs doesn't look like a single page. Instead, it appears to be four different pages with the same content:

- <https://www.example.com/about/>
- <http://www.example.com/about/>
- https://www.example.com/about?utm=my_parameter
- <https://www.example.com/category/article/798/>

The problem with duplicate content

The problem presented by duplicate content is fairly simple: **How should a search engine know which address is the one you want to be indexed?**

Keep in mind that indexing all of the URLs will make them compete against one another for search result rankings and user clicks. This lowers the position of all of the pages in the search results compared to the position you could earn for a single URL.

Finding duplicate content

During your audit, you should use a crawler or other tool that can identify duplicate content on your website.

Content types that should be checked for duplication and near-duplication include:

- Page titles
- H1 tags
- H2 tags
- Meta descriptions
- Full page text

Fixing duplicate content: different pages with duplicate content

In the case of duplicate elements on URLs that lead to pages that are different, the best solution is to modify the duplicated elements to make them unique.

- Page titles
- H1 tags
- H2 tags
- Meta descriptions
- Full page text of URLs that lead to different pages, like the similar product pages discussed above

Your audit should produce a list of pages with duplication issues, the type of issue, and the pages on which the same content can also be found.

Remember to re-check after modification to make sure the pages you have modified no longer have issues with duplicate content.

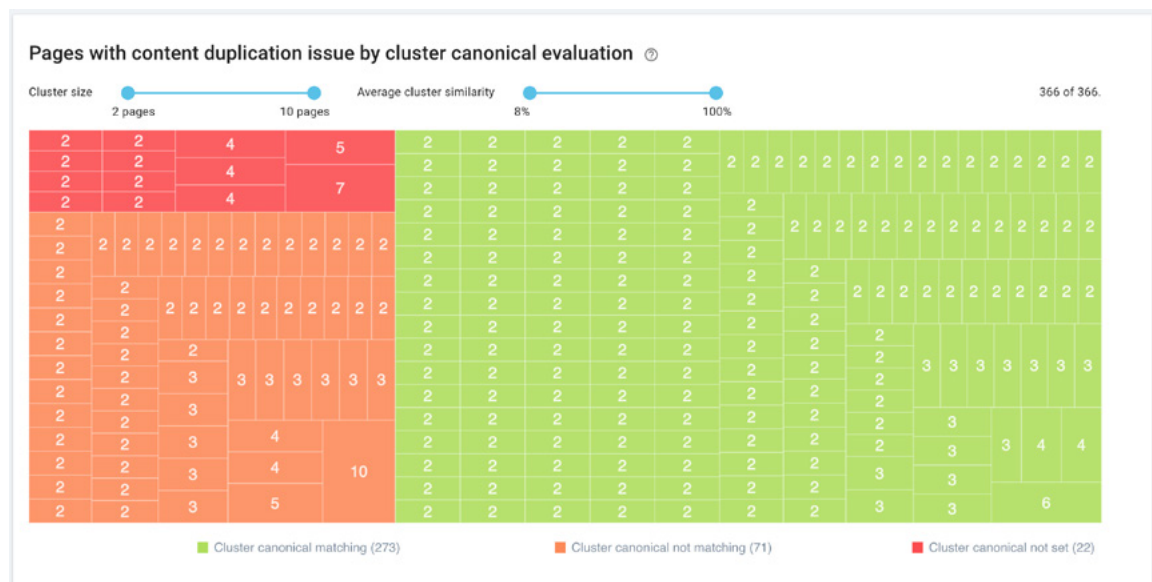
Fixing duplicate content: the same page with multiple URLs

When duplicate content comes from the fact that a single page has multiple URLs, the best strategy is to inform search engines which URL is the right one, and which URLs are extra pointers to the same page.

This can be done using the `<link rel="canonical">` tag in the header. This tag should indicate, on all URLs that point to a single page, the URL that you want indexed for that page.

Your audit should include a list of canonicalized URLs, and the URL they point to. Make sure that:

- All pages with multiple URLs have a canonical URL.
- All URLs pointing to the same page list the same canonical URL.
- Indicate whether the canonical URL has a mobile version by adding a `<link rel="alternate">` to the canonical URL's header.



Using OnCrawl to manage groups of duplicate content.

Sources:

<https://webmasters.googleblog.com/2013/04/5-common-mistakes-with-relcanonical.html>

<https://support.google.com/webmasters/answer/139066?hl=en>

<https://moz.com/blog/rel-canonical>

Mobile issues

10

Mobile Issues

Mobile search is increasingly important. As early as 2015, Google confirmed that more searches were performed on mobile devices compared to desktop devices. By mid-2018, that number has risen to nearly 60% for searches on Google, contributing to Google's move to the indexation of mobile pages for use in search results in September 2018.

Not only are an increasing number of users visiting sites and searching using mobile devices, but Google itself is also crawling as a mobile bot. This results in an advantage for sites with pages that display quickly and appropriately to mobile users.

Mobile search today involves:

- Increased importance of ranking high on a results page, thanks to smaller screen real estate and to a growing use of voice search
- Indexation of mobile pages, for sites with separate mobile and desktop experiences
- Increased importance of page speed
- Pre-cached pages offered by search engines, such as Google's AMP
- [Up to 80% difference in SERP positions](#) for desktop versus mobile searches

In general, your audit's first task is to make sure your site is mobile-friendly. Google offers a [single-page testing tool](#).

Mobile first indexing

In 2018, Google began indexing mobile pages instead of desktop pages. This means that the search results snippet (including the title, description, and cached version) are taken from the mobile page, if you have one, instead of the desktop page. This will also become the principal content used when ranking the URL.

If you use AMP, mobile first indexing will index the mobile version of your standard site, not the AMP page.

If you have separate pages or dynamic pages for mobile users, keep reading!

For sites with separate or dynamic mobile pages, Google suggests:

- Using identical content on your mobile and desktop sites
- Making sure any structured data is present on both mobile and desktop sites
- Metadata is present on both mobile and desktop sites

Mobile first indexing is preceded by a spike in crawls by the googlebot mobile user-agent, as Google crawls all of the known URLs in your site over several days or weeks. The desktop crawls decline. This is usually visible as a spike in the old Search Console crawl stats, or by monitoring googlebot hits in your server log files.

Your audit should:

- Check the crawlability and indexability of mobile pages.
- Use data in your server log files to check whether your website has moved to mobile first indexing yet or not.
- Make sure that the link `rel=canonical` and `rel=alternate` are correctly implemented to avoid issues with duplicate content on mobile and standard versions of a page.
- Make sure that any link `rel=hreflang` point to the correct type of page: that is, that hreflang tags on mobile pages only point to mobile pages, and hreflang tags on desktop pages only point to desktop pages.
- Check your robots.txt instructions for both mobile and desktop sites. Unless you've intentionally created an exception, bots should ideally receive identical instructions for each version of the site.

Responsive mobile pages

If your pages shift content around as you resize their width in order to view them on smaller devices, you have responsive pages. Responsive pages respond to the window size in order to provide the best layout for the user's screen size. This means that you only have one page, and this page is provided to all users, no matter what type of device they use.

This strategy is often praised as a good SEO strategy, as it requires less additional work to make sure that mobile versions of pages are properly set up and ready to be indexed by a search engine.

Google AMP

AMP or Accelerated Mobile Pages is an open source project supported by Google and Twitter. It can be used by any app, web viewer or browser. Google, Twitter, LinkedIn, Reddit and other platforms are currently using it. The main purpose is to create a fast mobile user experience.

AMP consists of HTML, JS, and cache libraries that, thanks to specific extensions and AMP-focused properties, speed up load time for mobile pages, even if they have "rich" content like videos, PDFs, audio or infographics.

AMP is based on light-weight coding and content hosting by Google in its AMP Cache. This cached version actually lets AMP pages to be preloaded and can be then instantly delivered to users for search results. It also ensures that all AMP pages are actually valid AMP. Users can identify AMP search results by the blue lightning bolt symbol next to the page link.

AMP pages have the advantage of:

- Ensuring proper display across devices
- Benefiting from reliable hosting by Google
- Fast loading appropriate for mobile data connections and for the processors in mobile devices
- Possible preferential treatment by Google in the SERPs

Using data scraping during a crawl, you can identify pages with an AMP version.

Your audit should:

- Check whether or not AMP versions exist for your web pages.
- Verify that AMP pages display correctly.

Sources:

<https://adwords.googleblog.com/2015/05/building-for-next-moment.html>

Mobile first: <https://webmasters.googleblog.com/2018/03/rolling-out-mobile-first-indexing.html>

Mobile first best practices: <https://developers.google.com/search/mobile-sites/mobile-first-indexing>

AMP: <https://www.oncrawl.com/oncrawl-seo-thoughts/google-amp-project-what-to-know/>

AMP: <https://www.oncrawl.com/oncrawl-seo-thoughts/amp-best-practices/>

AMP: <https://developers.google.com/search/docs/guides/about-amp>

International issues



International issues

International sites provide content adapted to different regions or languages.

For ease of reference here, we'll talk about different languages, but this content can just as easily be content in the same language with differences due to national customs, laws and regulations, currencies, etc.

Site structure for international sites

International content can be organized in different ways on a website.

Google strongly recommends not using analysis of IP addresses to redirect visitors to the "right" version of your site.

No matter what structure you use, your audit should:

- Make sure that your site allows Google to explore and index all of your pages. If you automatically redirect visitors from your homepage or dynamically change content, keep in mind that most googlebots crawl with a US-localized user agent.
- Analyze the internal link structure to ensure links from one page to another don't force users to change languages
- Verify that the language each page is written in is obvious. Google expects to find a single language on each page: navigation in one language and page content, or side-by-side translations can confuse Google.

Pell-mell organization

Nothing separates pages in one language from pages in another.

These sites are not yet optimized, and Google (and users) may have difficulty finding the pages that apply to them.

Your audit should find and identify pages for each language.

URL parameters

Each language is identified with a URL parameter, such as **<https://www.example.com?loc=en>**.

This practice is not recommended because it poses difficulties for Google's crawlers and for many site monitoring tools. It can also create difficulties for users.

Directory organization

Each language is located in its own directory, such as **<https://www.example.com/en>** and **<https://www.example.com/de>**. This is one of the best ways to organize a single website that is translated into multiple languages, particularly when there is little localization, or modifications for local customs, regulations, or cultures.

Your audit should make sure all language directories can be crawled by Google's bots.

Subdomain organization

Each language is located in its own subdomain, such as **<https://en.example.com>** and **<https://de.example.com>**.

Your audit should:

- Include pages on all language subdomains.
- Make sure that each subdomain's rules for bots allow pages to be discovered by googlebot.

Domain organization

Each language has its own website domain, such as **<https://www.example.co.uk>** and **<https://www.example.de>**. This type of organization is recommended when targeting different countries.

Your audit should include all of the sites for different languages.

Handling content in multiple languages

Google will attempt to identify the language of your pages based on many indicators, including:

- Information specified in Google Search Console's International Targeting Report, if it has been used. However, this function is still unavailable in the new Search Console as of late 2018.
- Country-level domain extensions (ccTLDs) such as <https://www.example.de>, if you use them.
- Hreflang declarations.
- Website's server location.
- On-page information, such as addresses, phone numbers, and currencies.

Consequently, if you are auditing an international site, you will need to audit hreflang declaration implementation.

Hreflang declaration is important:

- to ensure that Google can correctly identify the language of your content.
- to avoid creating duplicate content signals when you offer content for two countries in the same language.
- to help transfer the popularity of existing content to new translations.

Hreflang declarations are link declarations that identify alternate versions of the current page (its translations) using a two-letter language code (in ISO 639-1 format) and an optional country code (in ISO 3166-1 Alpha 2 format). For each page that is translated, hreflang declarations list links to the page itself and to all of the translated versions, no matter where on the internet they are located.

To prevent people on other websites from randomly listing their pages as translations of yours, hreflang declarations must be reciprocal: if page A lists page B as a translation, page B must also list page A as a translation.

Use hreflang="default-x" for generic, non translated pages or landing pages that exist to allow you to choose the language or region.

Hreflang declarations can be implemented:

- Either in the **<head>** of the page. These statements must occur before any elements that might prematurely close the <head>. For an English and a German translation set, both pages should indicate the following hreflang declarations:

```
<link rel="alternate" hreflang="en" href="https://www.example.com/en/">
<link rel="alternate" hreflang="de" href="https://www.example.com/de/">
```

- Or in your **HTTP headers** for non-HTML content. Hreflang in HTTP headers uses the following format:

```
Link: <https://www.example.com/en/>; rel="alternate"; hreflang="en",
<https://www.example.com/de/>; rel="alternate"; hreflang="de"
```

- Or in an **xml sitemap**. Every translated page should list a hreflang declaration for all URLs that are translations of the same content, including itself. All of these URLs must also appear in the same sitemap. This should look like:

```
<url>
  <loc>http://www.example.com/en/</loc>
  <xhtml:link rel="alternate" hreflang="de" href="http://www.example.com/
de/" />
  <xhtml:link rel="alternate" hreflang="en" href="http://www.example.com/
en/" />
</url>
```

You audit should:

- Verify the presence of hreflang declarations.
- Make sure all hreflang declarations are full URLs (<https://www.example.com/de/>) and not short or relative links.
- Make sure that all pages that are translations of the same content all list one another and themselves in their hreflang declarations.
- Check that all pages listed in hreflang declarations exist.

Sources:

<https://support.google.com/webmasters/answer/182192?hl=en>

<https://support.google.com/webmasters/answer/189077?hl=en>

<https://www.oncrawl.com/oncrawl-seo-thoughts/hreflang-and-seo-5-mistakes-to-avoid/>

<https://www.oncrawl.com/product-updates/hreflang-metrics-language-clusters/>

Schema.org

12

Schema.org

Structured data is markup that allows content to be processed as entities, which allows machines to “understand” the relationships between pieces of content. It provides a way of allowing bots and other algorithms with ways of recognizing familiar types of content and providing better results in response to search queries.

[Schema.org](https://schema.org) provides the standards for online structured data.

Google uses structured data to create:

- Knowledge Panel results
- Carousels of videos, images, ingredients, top stories
- News results
- Map results
- Rich results in SERPs that include additional information such as: Star ratings, Show times, Prices, Calories in recipes, Search boxes, Corporate information

These enrichments make a URL stand out in the search results, place it in the right-hand bar, or award it a spot in the coveted “position zero”, before the traditional search results begin.

The advantage conferred by a rich result varies by the type of enrichment and how it occurs on the SERPs. In some cases, rich snippets can increase click-through rates by up to four times; others report lower click-through rates for snippets in position zero--although this position is the result used by voice searches and therefore reach a larger audience even if they receive fewer direct page visits.

Here is a list of structured data entities:

Supported by Google:

- Breadcrumb
- Corporate contact
- Carousel
- Logo
- Sitelinks searchbox
- Social profile
- Article
- Book
- Course
- Dataset
- Event
- Fact Check
- Job Posting
- Local Business
- Media (limited access)
- Occupation
- Podcast
- Product
- Recipe
- Review
- Software App (in beta)
- Speakable (in beta)
- Top Places List (in beta)
- Video

Supported by Bing:

- Breadcrumbs
- Businesses and organizations
- Events
- People
- Products and offers
- Recipes
- Reviews
- Reviews (aggregated)
- ClaimReview / Fact Check

You audit should:


- Identify pages with structured data.
- Identify the types of structured data used on the website.
- Check that the structured data you use is valid.
- Compare the performance (SERP rankings, bot visits, organic visits) of pages with structured data to those without, in order to determine its value on your site

Here are a few examples:

Google restaurant

All Maps Images News Videos More Settings Tools

About 4,510,000,000 results (0.54 seconds)



Map data ©2018 Google

Rating ▾ Cuisine ▾ Price ▾ Hours ▾

Brothers Railroad Inn
4.5 ★★★★★ (101) · \$\$ · Italian
113 S Penn Ave
Closed · Opens 5PM

El Pueblito Mexican Restaurant
4.5 ★★★★★ (317) · Mexican
1721 N Pennsylvania Ave
Closed · Opens 11AM

Down Home Family Restaurant
4.4 ★★★★★ (191) · \$ · Family-friendly
1515 N Pennsylvania Ave
All you can eat · Comfort food · Small plates

☰ More places

Perfect Apple Pie Recipe - Pillsbury.com

<https://www.pillsbury.com/recipes/...pie/1fc2b60f-0a4f-441e-ad93-8bbd00fe5334> ▾



★★★★★ Rating: 4.5 - 777 votes - 3 hr - 230 cal

First, refrigerated **pie** crusts should be **baked** in either glass or dull-metal **pie** pans. A dark pan will cause too much browning while a shiny pan will prevent your ...

pie ingredients

 Meringue	 Condensed milk	 Rhubarb	 Pumpkin	 Graham cracker crust	 Graham cracker	 Evaporated milk	 Bisquick	 Pumpkin pie spice	 Key lime
---	---	--	--	---	---	---	---	--	---



Cat

Animal



The cat, often referred to as the domestic cat to distinguish from other felids and felines, is a small, typically furry, carnivorous mammal. It is often called house cat when kept as indoor pet or feral/feral domestic cat when wild. It is often valued by humans for companionship and for its ability to hunt vermin. [Wikipedia](#)

Lifespan: 2 – 16 years (In the wild)

Gestation period: 58 – 67 days

Family: [Felidae](#)

Daily sleep: 12 – 16 hours

Mass: 7.9 – 9.9 lbs (Adult)

Did you know: Outdoor cats tend to be farsighted and indoor cats tend to be nearsighted. [fullyfeline.com](#)

Breeds

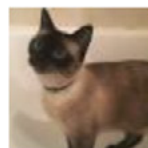
[View 15+ more](#)



Russian Blue



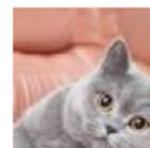
Persian cat



Siamese cat



Scottish Fold

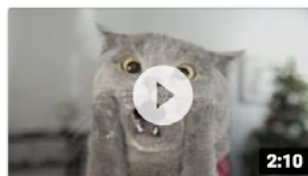


British Shorthair



You should have a CAT -
Funniest cat videos
ever!

Tiger Funnies
[YouTube](#) - Nov 27, 2016



Cat Left Home Alone

AaronsAnimals
[YouTube](#) - Dec 21, 2017



CATS will make you
LAUGH YOUR HEAD
OFF - Funny CAT
compilation

Tiger FunnyWorks
[YouTube](#) - May 31, 2017



Top stories



Do dogs understand the words we say to them?

Futurity.org

18 hours ago



New York City Halloween Dog Parade Is Back on Track

Wall Street Journal

12 hours ago



Dog named Magic found alive two days after brick wall collapses on him

NBC News

22 mins ago



Dog - Wikipedia

<https://en.wikipedia.org/wiki/Dog> ▼

The domestic **dog** is a member of the genus *Canis* (canines), which forms part of the wolf-like canids, and is the most widely abundant terrestrial carnivore.

Species: *C. lupus* **Kingdom:** Animalia

Genus: *Canis* **Phylum:** Chordata

Biology · Intelligence, behavior and ... · Ecology · Roles with humans

Sources:

Documentation (Google): <https://developers.google.com/search/docs/guides/search-features>

Documentation (Bing): <https://www.bing.com/webmaster/help/marking-up-your-site-with-structured-data-3a93e731>

Drawing conclusions from your audit

13

Drawing conclusions from your audit

Once you've completed your audit, what you do next is incredibly important to gain value from all your hard work.

Your audit alone will not do you much good without taking the proper steps afterwards to resolve any issues and make the necessary important. Follow these steps when drawing conclusions from your technical audit:

Actionable items

First, you'll want to identify the errors or improvements revealed in your audit and determine the actions required to resolve them. For example, one error that you may identify is that your website contains many 404 errors. You may find that many of these are caused by broken internal links. List the actionable next steps: in this case, list the action to redirect these 404 errors to a working webpage. This, in turn, will improve your Page Rank and your webpage ranking across your website. Follow this process for each error in your audit.

Prioritization

Now that you've identified the next steps for your actionable items, it is important to prioritize your work and resources before taking action. For example, fixing a broken link on the main page of your website will have a huge impact with relatively few resources. On the other hand, restructuring your website navigation may also have an impact on your rankings, however, it will require more resources to complete.

Keep this cost-benefit analysis in mind when choosing where to start improvements after completing your technical SEO audit.

Recall strategic business goals

Lastly, remember that all the work you put into resolving technical SEO issues should be connected to achieving your overarching business goals. Keep in mind what it is that you're trying to achieve, and what the end result should be.

For example, by reorganising your blog in using a topic cluster strategy you may want to increase traffic to your blog. Reorganising your blog is only one way to do this, however. Removing duplicate content or improving your keyword research strategy are also other ways to drive more visitors to your blog.

By keeping your business goals in mind, you'll reveal multiple efforts that will help you to reach your goals. Prioritise these efforts because they'll have the greatest impact on your business's bottom line.

Conclusion

14

Conclusion

Now that you know how to conduct a technical SEO audit, what do you do next?

Use this guide while working through all of the steps listed throughout this ebook to navigate through performing your audit. Once you've conducted your first audit, be sure to use your insights to make the necessary improvements to your website to drive more traffic to your content.

Don't forget, conducting a technical SEO audit is not only a one-time action. Be sure to continually track your webpage performance, experiment with new ideas, and conduct additional audits in the future. Best practice is to audit your website at least once a year, or whenever you've made significant structural or design updates to your website, whichever comes first.



Start Growing with HubSpot Today

With tools to make every part of your process more human and a support team excited to help you, getting started with inbound has never been easier.

[Get HubSpot Free](#)



Unleash your SEO Potential

The technical SEO platform that helps you make smarter SEO decisions, for enterprise SEO audits and daily monitoring.

[Start Your Free Trial](#)



HubSpot



OnCrawl

