The Ultimate Site Audit

Table of Contents

Introduction

You want to know your site's technical health, but performing an audit can seem like a never-ending chore. There's a doomy list of possible troubles a site audit might reveal.

Going through that list manually is tedious (and often groan-worthy), but it doesn't have to be.

With our stress-free Site Audit ↗ tool, you can quickly check for over 140 of the most common site health issues on any site. With fast crawling and customizable settings, the tool automatically detects almost every SEO obstacle between your site and the top of Google's results page.

You might also want some knowledge under your belt to understand exactly when, how, and why your site's health is squeaky clean. This PDF has info on all of the potential issues you'll face.

Find out why website problems crop up along with tips on how to fix them right here. All of the issues in this PDF are divided into three categories (just like our Site Audit tool).

ⓧ ERRORS: The most crucial issues that require immediate attention.

⚠ WARNINGS: These issues may have an impact on your website's performance and should never be neglected.

▤ NOTICES: Insignificant issues that might not pose a problem but still deserve attention.

In a short amount of time, you'll be able to conduct a complete audit of your site quickly and effectively.

Start your first audit ↗

Crawlability & Indexability

Crawlability & Indexability

There's no point in optimizing a virtually invisible website. For a site to appear in a search engine like Google, it needs to be crawled and indexed by it. Your website's crawlability and indexability are two of the most commonly unseen elements that can harm your SEO efforts if they're not in good shape.

To make sure both users and crawl bots can navigate your website, you need to build a well-organized site architecture. SEO-friendly here equals user-friendly, just as it should.

Streamline your website's structure, and make sure that valuable, converting content is available—and no more than three clicks away from your homepage.

Level up your website's crawlability.

Start your audit ↗

Robots.txt

Many factors can prevent search bots from crawling ↗ your site. A robots.txt file ↗ can help. This file communicates with search engines and tells crawlers which URLs to access. Although it's not crucial for a website's well-being to have a robots.txt, it can increase a site's crawling and indexing speed.

But watch out for mistakes. Google may ignore important pages of your site or crawl and index unnecessary ones if you don't. If formatted a certain way, it can block Google from crawling and indexing the whole site or parts of a site. Building a robots file can be straightforward. That said, format errors are quite common.

Watch out for:

- the wrong syntax
- an empty user-agent line
- mismatched directives

- listing multiple directories in a single line
- listing each file instead of shutting indexation for the whole directory

If you're not a web developer, these errors may seem foreign to you. Our Site Audit tool can demystify technical errors and help you fix them.

Simply put, consider a robots.txt as a guide to your website. By creating a simple file in txt format, you can lead bots to important pages by hiding those that aren't significant.

We recommend that you exclude from crawling:

- temporary pages

- pages without valuable content

- private pages that are only visible to certain users or administrators

Robots.txt are more of a suggestion, and sometimes bots can neglect it. Even if that's the case, it's always better to have one to help guide crawlers.

⊗ **Format errors in robots.txt**

⊗ **Invalid structured data items**

⚠ **Issues with blocked internal resources in robots.txt**

▤ **Issues with blocked external resources in robots.txt**

▤ **Robots.txt not found**

▤ **Pages blocked by X-Robots-Tag: noindex HTTP header**

▤ **Pages were blocked from crawling**

To learn more about robots.txt files, look into Google's manual on robots.txt ↗. If you want to validate an existing file, you can use this robots.txt tester ↗.

URL Structure

A [URL is more than just the address of a webpage ↗](). Crawlers and people alike will read URLs, so use relevant phrases in URLs to indicate what your page's content is about.

You can have the URL match the title, but know that search bots may consider all characters in a title. There are a few factors to think about before deciding on a URL.

Before You Pick a URL

Use hyphens or dashes instead of underscores to refrain from mix-ups. Google sees hyphens or dashes as the separation between words, not so with underscores.

Also, don't use capital letters unless you have a very good reason. Maybe you're working on a pre-existing site that uses capital letters consistently—this is a good reason. Otherwise, don't since the path part of the URL may be case-sensitive, leading to 404 errors on the user's end if they don't capitalize the URL correctly.

URL structure can signal the page's importance to search engines. Generally speaking, the higher the page is within this structure, the more important it seems. So keep the structure simple and put your prime content as close to the root folder as possible.

Keep in mind that having URLs that are too long or complex with many parameters is neither user nor SEO-friendly. So, although you can have up to 2,048 characters in a URL, **keep its length under 100 characters**. Try to trim down dynamic parameters (like with Google Ads) when possible.

⊗ Malformed links

⚠ Underscores in the URL

⚠ Too many parameters in URLs

⚠ URL is too long

▤ URL longer than 200 characters
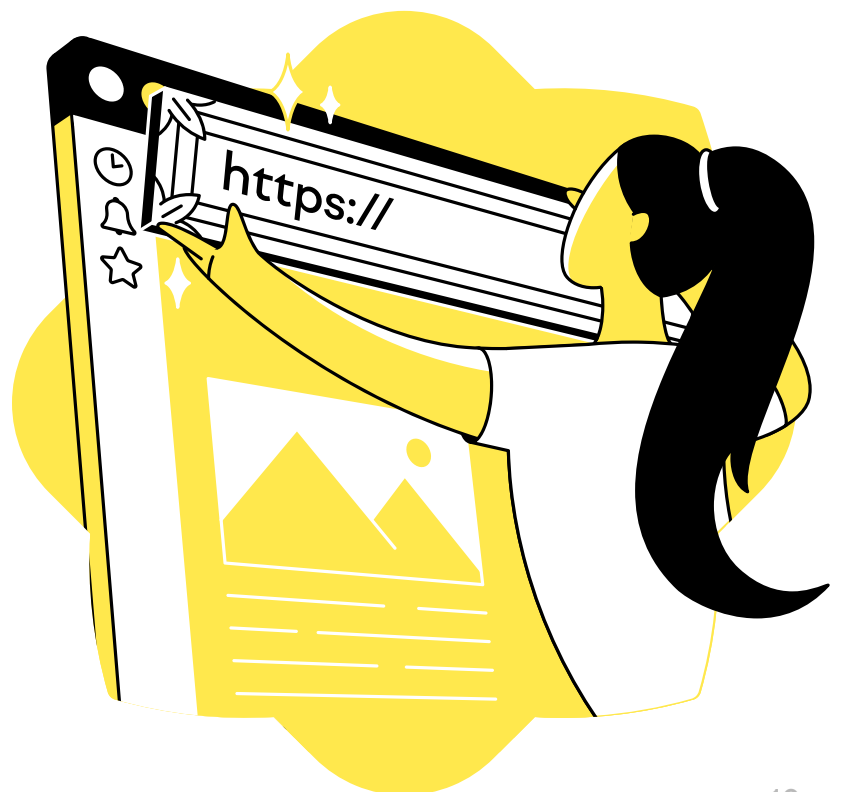
Links & Redirects

You need links on your website to steer users and to redistribute pages' authority from one page to the next. But you need to make sure your site's links work. Broken links and 4xx and 5xx status codes can ruin both the user experience and your SEO efforts.

Having too many links on a webpage is another way to dash your SEO efforts and UX. Your site's visitors won't go through all the links anyway and neither will bots. Also keep in mind that mistakenly using nofollow attributes ↗ can be harmful, especially when applied to internal links.

If you have broken external links, reach out to the website owners. Carefully review your own links, replace or remove broken ones, and in the case of server errors, contact web hosting support.

Internal Linking

In many SEO discussions, external links always seem to be the star of the show. Yes, backlinks are important to SEO, but internal linking ↗ can be just as important to both users and crawlers. Broken internal links can ruin UX and stop crawlers from navigating to important content on your website.

Pages that took you hours to craft may never come across a single pair of eyes if there are no links going to them. Typically, these are easy fixes that can help you rise through the ranks and provide valuable content.

ⓧ Broken internal links

▤ Pages with only one incoming internal link source

▤ Pages that need more than 3 clicks to be reached

Read Before You Redirect

Another concern here is dealing with temporary redirects. When you use 302/307 redirects instead of a 301 redirect, the search engine keeps the old page indexed and the PageRank ↗ does not transfer to the new one (bye bye PageRank).

Search bots may consider your website with WWW and without WWW as two separate domains. So, you need to set up 301 redirects to the preferred version **and** indicate it in Google Search Console.

Canonicals: Use Them Right

If you have multiple versions of a page, you need to use the rel="canonical" tag to inform crawlers of which version you want to show up in search results. Otherwise, the search engine may not know which page to index, and your rankings could suffer. But, you have to be careful when using canonical tags.

Make sure that the rel="canonical" element doesn't lead to a broken or non-existent page; this can severely decrease crawling efficiency. And if you set multiple canonical tags on one page, crawlers will most likely ignore all of them or pick the wrong one.

Redirects: Use Them Sparingly

[Redirect chains and loops ↗](#) will confuse crawlers and frustrate users with increased load speed. You also lose a bit of the original page rank with each redirect. That is a big no-no for any website owner.

However, redirection mistakes tend to slip through the cracks and pile up, so you have to check these periodically.

⊗ **Pages returning 4XX status code**

⊗ **Pages returning 5XX status code**

⊗ **WWW domain configured incorrectly**

⊗ **Redirect chains and loops**

⊗ **Pages with a meta refresh redirect tag**

⊗ **Broken canonical link**

⊗ **Multiple canonical URLs**

⚠ **Pages with temporary redirects**

⚠ **Internal links with nofollow attributes**

⚠ **Too many on-page links**

⚠ **Broken external links**

▤ **External links with nofollow attributes**

▤ **URLs with a permanent redirect**

Sitemap

Submitting a sitemap ↗ to Google Search Console is a great way to help bots navigate your website faster and get updates on new or edited content.

A sitemap is a way to help highlight the landing pages you want on the SERPs versus what you don't. That said, a sitemap doesn't guarantee that the pages will or won't be indexed.

You can create an XML sitemap manually, or generate one using a CMS or a third-party tool ↗. Search engines only accept sitemaps that are less than 50 MB and contain less than 50,000 links.

If you have a large website, you might need to create additional sitemaps. You can learn more about managing multiple sitemaps from this guideline ↗.

Listing pages that aren't linked internally on your site is a bad practice as well. If there are multiple pages with the same content, you should only leave the canonical one in your sitemap.

Don't add links to your sitemap that are blocked with the robots file. This would be like telling a search bot to simultaneously crawl and not crawl the page. But do remember to add a link to your sitemap to robots.txt.

To learn more about the correct implementation of a sitemap, look into [the official guide ↗](#).

⊗ Format errors in sitemap.xml

⊗ Wrong pages found in sitemap.xml

⊗ Sitemap.xml files are too large

⚠ Sitemap.xml not found

⚠ Sitemap.xml not indicated in robots.txt

▤ Orphaned pages in sitemap

▤ Orphaned pages from Google Analytics
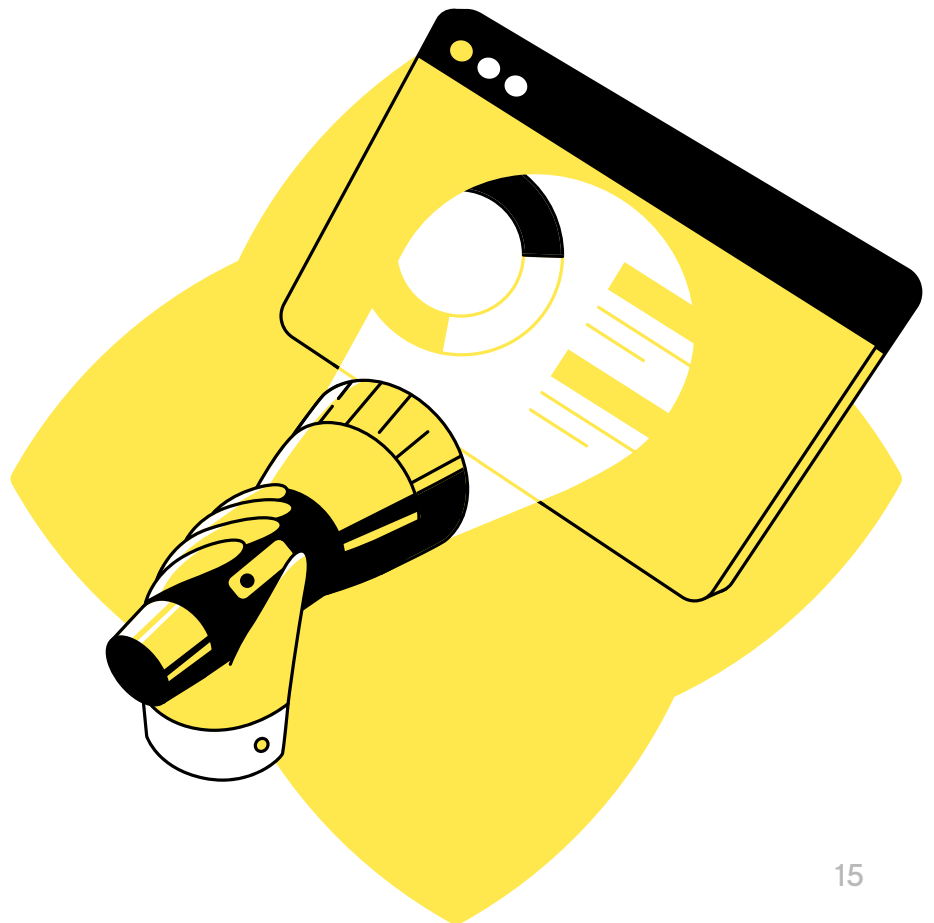
On-Page
SEO

On-Page SEO

On-page SEO is about improving the rankings of specific pages by optimizing their content and the HTML behind them. You need to carefully craft all the ingredients of a page to earn more relevant traffic. [Great written and visual content ↗](#) combined with the perfect backstage work leads to user satisfaction and search engine recognition.

It's also fair to say that well-executed on-page optimization is a legitimate path to the off-page success of your website. Using strong content as a basis for link building will take less effort to reach excellent results. And the best part is that all those elements are in the palm of your hand—you can always adjust the content displayed on the page and meta tags concealed in the code.

Find all on-page SEO mistakes with this site audit.

[Start your audit ↗](#)

Content

[Good SEO means good content ↗](). First of all, that means avoiding the urge to fill your site with rehashed or even copied content that can hurt your rankings. Remove or replace duplicate content pages with unique ones to avoid this. We advocate that pages have at least 85% unique content.

Sometimes you can't avoid duplicate content. This is the case with ecommerce, where you can sort items by price or popularity. You can, however, avoid [cannibalization ↗]() (multiple pages targeting the same keywords) by indicating secondary pages with a rel="canonical" tag that links to the main one.

Other times, when a URL has parameters (think different sizes and colors of clothes on an online boutique), it might get indexed as a separate page, thus creating a duplicate. To prevent that from happening, add a self-referential canonical tag that directs to the clean version of the URL.

Another important issue is your pages' word count. Long-form content tends to have more value, and generally, we recommend putting at least 200 words on a page. But obviously, not every page needs a lot of text, so use common sense and don't create content just for the sake of content.

Note that a low text-to-HTML ratio can indicate poor quality. A page shouldn't

contain more than 10% of the displayed text in relation to the code. This is especially important for a page with lots of content since it likely means there's unnecessary code bogging download times and crawling speed.

It's usually not a huge deal if you can't help having more code than text, but it could separate you and a competitor's domain from that first SERP position.

⊗ Missing or empty title tags

⊗ Issues with duplicate title tags

⚠ Duplicate H1 and title tags

⚠ Title tag is too long

⚠ Title tag is too short

H1 Tag

A page's H1 heading helps crawlers and users and stands out visually on a page. The H1 and the title tag can be identical, which is acceptable from a technical standpoint but not a good SEO practice.

When your H1 and title are the same you miss the chance to diversify semantics with varied phrases, making your page look suspiciously optimized. Give some thought to your H1s—make them catchy, yet simple and relevant.

Search bots use the H1 to get a hint as to what is your page about. Don't distract them by putting multiple H1s on a single page. Instead, use an H2–H6 hierarchy for descending subsections.

Structured text is better at holding readers' attention and better for UX overall. So create scannable content, and make sure that your headings and subheadings correlate with the topic of a page and its sections.

⚠️ Missing H1 heading

⚠️ Duplicate H1 and title tags

🗏 Multiple H1 headings

Meta Description

If your page's title tag is the proverbial book cover, then your meta description ↗ is the back cover that sells it for a click. Of course, a missing meta description won't affect your rankings—Google will make one up for you. But the results likely won't help your potential CTR.

In some cases, it might be inconvenient and unnecessary to come up with a unique description for each page. Here, you should concentrate on the most important landing pages and leave all the rest auto-generated.

Creating a loud-and-clear summary of a page is an art, but keep in mind that copy-pasted meta descriptions are worse than not having any at all. Duplicates might stop a crawler's ability to distinguish the relevance and priority of a page.

You can use SEOmofo ↗ to preview the appearance of your title, description, and URL in the snippet on Google's SERP.

⊗ Duplicate meta descriptions

⚠ Missing meta descriptions

Images

Top ranking content in an image SERP can bring a chunk of a target audience to your website. Despite this, image SEO ↗ is still neglected by some website owners.

Let's look solely at the SEO aspects of an image—its alt attribute and its availability. Seeing appealing and informative images on a website is awesome, but broken links and no longer existent sources can spoil all the fun.

Plus, Google may decide that your page is poorly coded and maintained if it contains broken images. You need to regularly inspect your site for such occurrences and reinstate or erase faulty elements, especially if your imagery is doing the selling. With missing pictures, it's hard to reach an audience for clothing shops, food delivery, hotels, etc.

Google Cares About Your File Name

An alt attribute should give a clear depiction of a picture, and while it is an opportunity to add more keywords to a page, beware of keyword stuffing. Keep the alt attribute simple and accurate to what is seen in the image. Another tip many website owners don't know is that the file name of an image also matters since search engines will read it when crawling a page.

Try to give your files relevant names and create descriptive alt attributes, because besides helping you rank in image searches, it also helps people who are visually impaired.

⊗ **Broken internal images**

⊗ **Broken external images**

⚠ **Images missing alt attribute**

▤ **Resources formatted as page link**

Technical
SEO

Technical SEO

Technical SEO deals with code-related factors that affect user experience, and, as a result, rankings. Common issues include slow page loading speed, use of outdated technologies, and lack of optimization for mobile devices. These are aspects of a website audit that you need to pay extra attention to because poor page performance can reverse all the good SEO work that you've done.

On the other hand, fixing technical issues can be highly rewarding. Many technical issues have a site-wide nature, so fixing them usually benefits your whole website. Oftentimes just a little tweaking can drastically increase your traffic and save you a lot of money.

Improve your website performance with the Site Audit tool.

[Start your audit ↗]

Page Speed

Page speed is a big ranking factor affected both by the server-side and page performance. And, for that reason, it's a big bounce rate cultivator. To increase page speed and avoid long loading times, you need to keep your code to a minimum.

One way to achieve this is by using compression schemes like gzip or deflate. Condensing HTML, CSS, and Javascript can greatly benefit load speed, but there are drawbacks of complicated setup and issues with older browsers.

Images usually take up the most weight on a page, so optimizing them is essential for increasing a page's speed.

Don't Let Images Slow You Down

Fine-tune your images using a graphic design tool of your choice. Try to achieve the smallest file size you can while maintaining image quality. Examine the possibility of using vector graphics. It's a great way to slim down simple geometrical images. Also, don't use large images if you can.

Lastly, since page speed is even more important for mobile devices, you have to configure the viewport and rescale images for different screens.

⊗ Large HTML size

⊗ Slow page load speed

⚠ Uncompressed pages

⚠ Issues with uncompressed JavaScript and CSS files

⚠ Pages have a JavaScript and CSS total size that is too large

⚠ Pages use too many JavaScript and CSS files

⚠ Issues with unminified JavaScript and CSS files

⚠ Issues with uncached JavaScript and CSS files

▤ Issues with broken external JavaScript and CSS files

For more extensive information consult
[Google's recommendations for page speed optimization ↗](#).

Core Web Vitals

Yes, there are plenty of search engines to choose from. But, Google is probably the one that you think of when you have questions that need answers! It's important to make your website appealing to Google because it impacts your site's rankings, but also because it helps UX.

Core Web Vitals ↗ are a series of ranking factors out of the 200 known ones that Google looks at when it decides where to place your site on the SERP. The metrics Google's algorithm bases it on are:

- **LCP (Largest Contentful Paint)**— The amount of time it takes to load the largest element on your page from the part of a page a user can see (typically is an image or video)

- **FID (First Input Delay)**—How long it takes for your website to react to a user's interaction (such as when they click on a button)

- **CLS (Cumulative Layout Shift)**— How many times elements of a webpage shift (like if a page shows a certain font at first and then switches to another once the code fully loads)

Google provides a great breakdown of the importance of these metrics ↗. Our Site Audit tool can help you fix any problems that your site may have in this area by looking at the same metrics. The only difference is that we use Total Blocking Time (TBT) instead of FID, but it basically measures the same factors

▤   Pages that take more than 1 second to become interactive

Old Technology

The death of Flash was a long time coming, and for good reason. From an SEO perspective, Flash software slowed down a page's performance and handicapped crawling. Adobe stopped supporting its technology at the end of 2020.

As for installing widgets and plugins from external domains with iframes—they can come in handy and won't affect your rankings if implemented properly, but can also hurt your website's user experience and complicate its indexing.

For a browser to understand how to properly render the content, you should always specify which version of HTML or XHTML a page is written in with the tag. Give it special attention if you're using an older version of code.

⚠ Frames used

⚠ Doctype not declared

⚠ Incompatible plugin content

Mobile

We're all optimizing for mobile devices ↗, right? With mobile usage skyrocketing, it's crucial to check that all your pages have viewport tags and can scale for various screen sizes. Opting out will scare away all your visitors and will worsen your rankings, especially considering Google's concept of mobile-first indexing.

Accelerated Mobile Pages ↗ (AMP) are a great method to align your website with mobile-friendliness. AMP started as a way for publishers to serve fast-loading content from a search engine results page, but now it's also a platform for ecommerce and advertising. The project constantly evolves, and now AMP pages can show up in featured snippets.

The process of implementing AMP is entangled in a lot of intricate details. There are many potential AMP mistakes ↗, and so that you don't feel overwhelmed, these are sorted into groups:

- HTML issues

- Style and layout issues

- Templating issues

The Site Audit tool covers all three of these groups and detects over 40 AMP errors. You can crawl a whole site on demand whenever you like and see all of the broken pages in one report. You can also choose to crawl AMP versions first.

Since the AMP format involves code restrictions and usage of custom AMP tags, errors in the HTML issues group are very common. The styles and layouts also require specific AMP standardization.

This group of errors should also be prevented so that a page can be properly indexed and served. And if your page includes template syntax, it will not work correctly unless that syntax is used in the AMP tags specifically designed for templates.

These three groups of checks are available for Business plan users. You will get a detailed list of every AMP error for each page, with descriptions and fixing tips with the Site Audit tool.

You will also see the exact line of code containing a mistake; this feature is not present in any of the other tools, including Google's AMP test tool ↗.

Another important thing to consider if you have an AMP version of a page is ensuring it has a canonical tag, and is referenced on the non-AMP version. That way you will avoid duplicate content issues. If you only have an AMP page, add a self-referential canonical tag.

⊗ **Missing viewport tag**

⊗ **Missing the viewport width value**

⊗ **AMP pages with HTML issues**

⊗ **AMP pages with style and layout issues**

⊗ **AMP pages with templating issues**

⊗ **AMP pages have no canonical tag**

⊗ **Issues with broken internal JavaScript and CSS files**

HTTP Implementation

Security is a necessity for every website. You have to protect yourself and your users from any malicious intent on the internet if you want to have their trust. HTTPS protocol ensures that all the data transferred through your website is authentic, encrypted, and intact. And, of course, there's a perk of Google's favoritism toward secured pages.

HTTPS is a ranking factor ↗ that will become more and more considerable in the future, because safety issues have no expiration date. But behind all those security benefits there are also quite a lot of risks associated with moving your site to HTTPS and maintaining a secured protocol.

Secure your HTTPS implementation.

Start your audit ↗

When shifting your website to the secured protocol, you can come up against multiple mistakes ↗. Beware of missing redirects and canonicals to HTTPS URLs, as these can lead to lower rankings and cannibalization. Use a 301 redirect or rel="canonical" on the HTTP version to indicate that your primary version is on HTTPS now. Mind all the elements of a page, and only add HTTPS content to HTTPS pages to ward off security and UX issues. And remember to update your website's internal linking and your sitemap with HTTPS URLs.

Keep an eye on your SSL certificate— it should be up to date, valid, and registered to the correct domain or your users will get upsetting notifications, which will certainly increase bounce rate. It's recommended that you implement HTTP Strict Transport Security (HSTS) to force your user's browsers to only use secure connections. Also, it is good to have a server supporting SNI (Server Name Indication) so that there would be a possibility to use multiple certificates at the same IP address.

⊗ **Mixed content**

⊗ **Non-secure pages**

⊗ **No redirects or canonicals to HTTPS URLs from HTTP version**

⊗ **SSL certificate registered to an incorrect domain name**

⊗ **Old security protocol version**

⊗ **Expiring or expired SSL certificate**

⊗ **Subdomains don't support secure encryption algorithms**

⚠ **HTTPS pages lead to HTTP page**

⚠ **Homepage doesn't use HTTPS encryption**

⚠ **HTTP URLs in sitemap.xml for HTTPS site**

⚠ **Subdomains don't support SNI**

▤ **Subdomains don't support HSTS**

International
SEO
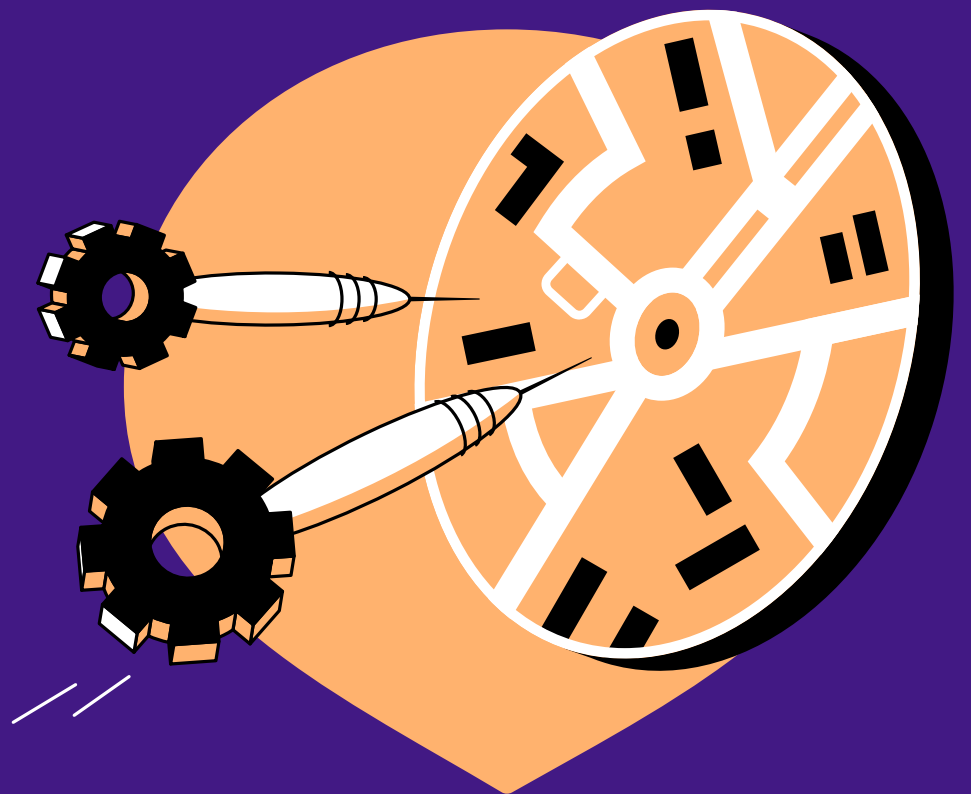
International SEO

The internet makes the world small, globalization never stops, and international SEO is becoming more relevant than ever. Both big and small companies can get traffic from all over the world by creating sites in more than one language.

Though, maintaining a multilingual website creates a specific set of potential problems. It's hard enough to [get the hreflang attribute right ↗](#) so that your audiences in different locations will get the version of your page with the correct language. Besides that, you also need to signal to the search engine which results should be provided for which users and explain to it that you're not just scattering duplicates around.

Fix all hreflang implementation issues.

[Start your audit ↗](#)

The main SEO problems international websites face are duplicates and redirects. Adding rel="alternate" hreflang="x" [tags will help Google ↗](#) figure out which version of a page to show based on a user's location.

Watch out for broken or conflicting URLs, and make sure that all alternative versions are referenced on each page, including self-reference, otherwise, a search bot might not understand or may ignore those annotations.

Also keep in mind that you can only have a self-referential canonical tag on a page with a hreflang, otherwise you will be giving conflicting instructions to the crawler.

When configuring an international multilingual website ↗, you first need to specify the correct language and country codes for matching pages. Language code should precede and be separated with a hyphen from a country code. Remember that you can designate a language without a country, but not the other way around. It's also important to declare encoding ↗ so that browsers will know which set of characters to use.

Even if you feel that all of your page's redirects and hreflang tags are perfectly implemented, it's still a good idea to add an option for users to select their language. And, **please**, don't use automatic translators for everyone's sake.

⊗ Hreflang implementation issue

⊗ Hreflang conflicts within page source code

⊗ Incorrect hreflang links

⚠ Pages with no hreflang and lang attributes

⚠ Pages without character encoding declared

▤ Hreflang language mismatch issues

About Semrush

Since 2008, Semrush has served more than 10 million marketing professionals around the world. Our goal is to help our users understand their competitors' marketing strategies, uncover growth opportunities, build brand reputation, and produce valuable content that delivers on brand goals.

By the Numbers (as of May 2022)

21B	keywords
142	geographic databases
808M	domain profiles
10M	people have already tried Semrush

Trusted By

TESLA   DECATHLON   P&G   SAMSUNG   Forbes   IBM   Walmart

We love your feedback!

We hope this ebook can help you polish your website and achieve high SEO rankings! Our team is always working to perfect our product's existing features and build new ones based on your feedback. Please share your thoughts and suggestions with us anytime at site-audit-feedback@semrush.com.