THE ULTIMATE SITE AUDIT WITH SEMRUSH

semrush

TABLE OF CONTENTS

You have to regularly check your site's health and well-being, but performing a site audit can be very stressful, as the list of possible troubles your site may face is huge. Going through that list manually is a tedious chore, but luckily there is a tool that can sort out all of those issues for you.

The SEMrush Site Audit is a powerful instrument for checking your website's health. With fast crawling and customizable settings, it automatically detects up to 60 issues, covering almost every website disorder possible.

Along with this great tool you are going to need some knowledge under your belt for truly competent website analysis.

This e-book will guide you through everything from crawlability issues to on-page mistakes. Some of those may seem minor, but you have to make sure that they will not stack up and chain-react with devastating repercussions.

With the SEMrush Site Audit tool, our recent research and this PDF, you will be able to conduct a complete audit of your site quickly and effectively.

That is why we put together this PDF with the checklist of issues SEMrush Site Audit identifies. We also carried out a new study on the most common on-site SEO issues to read on our blog. We checked 100,000 websites and 450 million pages for 40 issues to find out which mistakes appear more often. In this research we present you with a the lineup of issues that might appear on your website as well as data on how often each mistake was detected.

This guide will provide you with explanations of why these problems crop up and tips on how to overhaul them. All of the issues in the PDF are divided into three categories by criticality, the same way as in the SEMrush Site Audit.

ERRORS
The most crucial issues that require immediate attention.

WARNINGS
These issues have a lesser impact on a website's performance but should never be neglected.

NOTICES
Insignificant issues that might not pose a problem but still need attending to.

CRAWLABILITY AND SITE ARCHITECTURE

First things first, there is no point in optimizing anything on your website if search engines can not see it. In order for a site to appear in a search engine like Google, it should be crawled and indexed by it. Consequently, the website's crawlability and indexability are two of the most commonly unseen elements that can harm your SEO effort if not addressed.

To foster better navigation and understanding for both users and crawl bots, you need to build a well-organized site architecture. SEO-friendly here equals user-friendly, just as it should. To achieve that you need to streamline your website's structure, and make sure that valuable, converting content is available and no more than four clicks away from your homepage.

## LEVEL UP CRAWLABILITY OF YOUR WEBSITE WITH THE SEMRUSH SITE AUDIT TOOL

Start your audit →

There are many reasons that can prevent search bots from crawling. Robots.txt can block Google from crawling and indexing the whole site or specific pages. Although it is 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, as they can cause Google to ignore important pages of your site or crawl and index unnecessary ones. Despite the fact that building a robots file is not that hard, format errors are quite common: an empty user-agent line, the wrong syntax, mismatched directives, listing each file instead of shutting indexation for the whole directory or listing multiple directories in a single line.

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 are of no significance to users and therefore crawlers. We recommend that you exclude from crawling temporary pages and private pages that are only visible to certain users or administrators, as well as pages without valuable content. Although, robots.txt is never a strict directive but more of a suggestion, and sometimes bots can neglect it.

FORMAT ERRORS IN ROBOTS.TXT

ROBOTS.TXT NOT FOUND
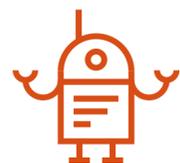
To learn more about robots.txt files, look into Google's manual on Robots.txt. If you want to validate an existing file, you can use Robots.txt Tester.
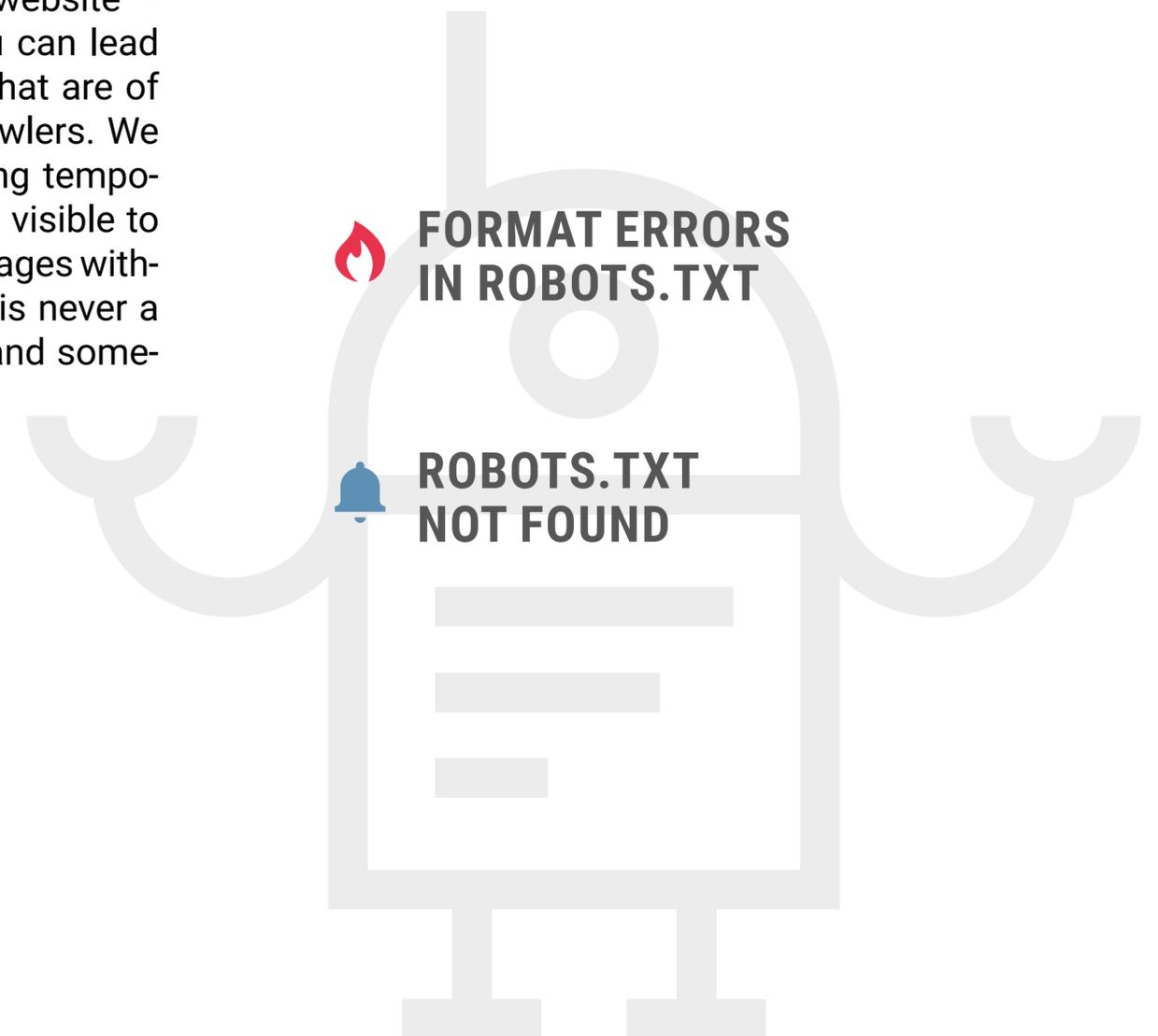
For an SEO specialist URL is more than just the address of a webpage. If left unattended, they can negatively affect indexing and ranking. Crawlers and people alike will read URLs, so use relevant phrases in URLs to indicate what the page's content is about. You can have the URL match the title, but know that search bots may consider underscores in URLs as part of a word, so it is better to use hyphens or dashes instead to refrain from mix-ups.

Do not use capital letters unless you have a very good reason. It just unnecessarily complicates readability for robots and humans. While the domain part of a URL is not case sensitive, the path part might be, depending on the OS your server is running on. This will not affect rankings, because a search engine will figure out the page no matter what, but if a user mistypes a case sensitive URL or your server migrates, you may run into problems in the form of a 404 error.

URL structure can signal the page's importance to search engines. Generally speaking, the higher the page is, the more important it seems. So keep the structure simple and put your prime content as close to the root folder as possible. Also keep in mind that having URLs that are too long or complex with many parameters is neither user- nor SEO-friendly. So, although it is officially acceptable to have up to 2,048 characters in a URL, try to keep its length under 100 characters and trim down dynamic parameters when possible.

⚠ **UNDERSCORES IN THE URL**

⚠ **TOO MANY PARAMETERS IN URLS**

🔔 **URL IS TOO LONG**
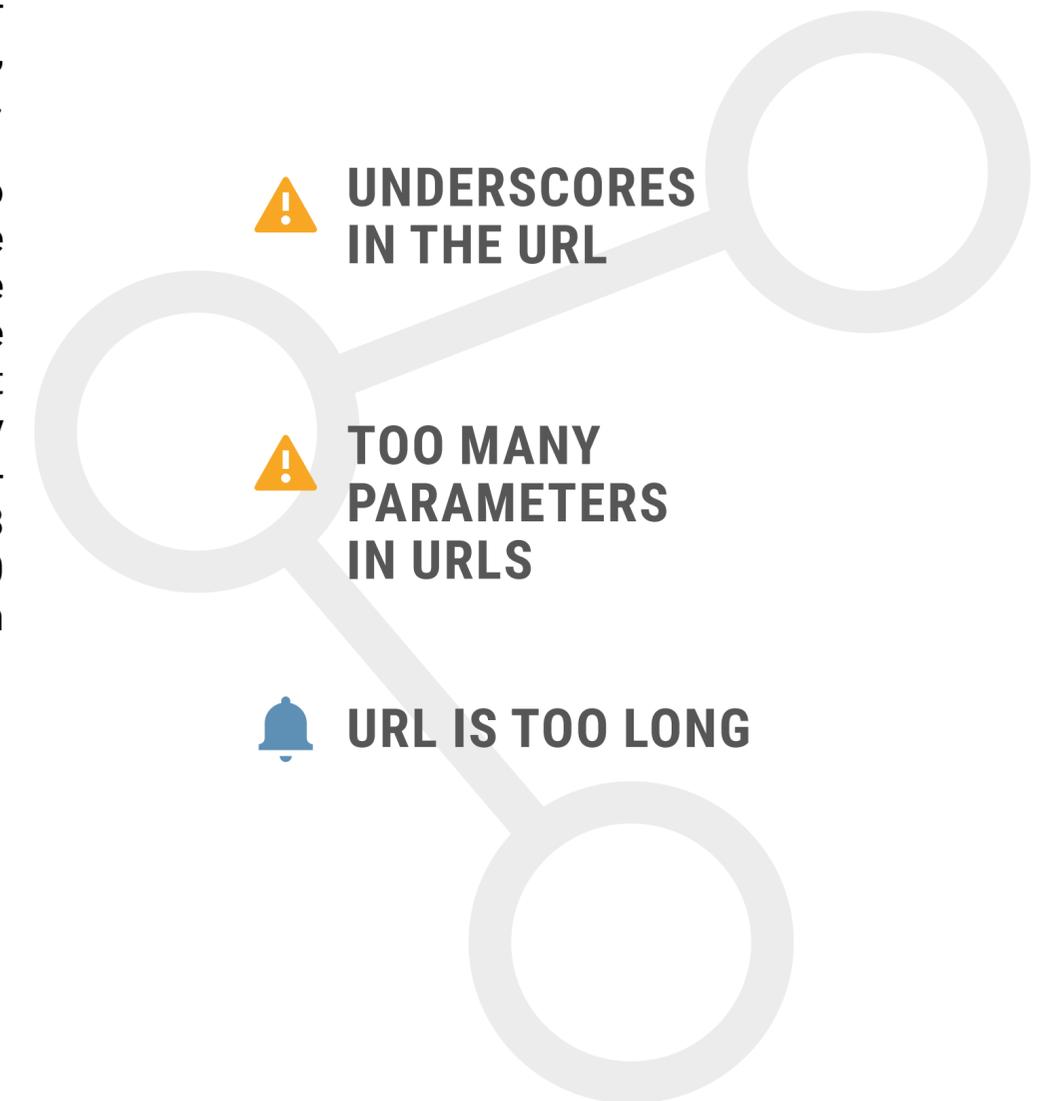
Having links on your website is necessary for steering users and redistributing pages' link juice. But broken links and 4xx and 5xx status codes can notably deteriorate user experience and your SEO efforts. Having too many links on a page as well makes it look spammy and unworthy to both users and crawlers, which will not go through all the links anyway. Also keep in mind that mistakenly used 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 inoperative ones, and in the case of server errors, contact webhosting support.

Another concern here is dealing with temporary redirects. They seem to work in the same manner as permanent ones on the surface, but when you use 302/307 redirects instead of a 301 redirect, search engine keeps the old page indexed and the pagerank does not transfer to the new one. Take into account that 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.

🔥 **4XX ERRORS**

🔥 **5XX ERRORS**

🔥 **BROKEN INTERNAL LINKS**

🔥 **WWW DOMAIN CONFIGURED INCORRECTLY**

🔥 **REDIRECT CHAINS AND LOOPS**

🔥 **BROKEN EXTERNAL LINKS**

🔥 **BROKEN CANONICAL LINK**

🔥 **MULTIPLE CANONICAL URLS**

⚠️ **TEMPORARY REDIRECTS**

⚠️ **INTERNAL LINKS WITH NOFOLLOW ATTRIBUTES**

⚠️ **TOO MANY ON-PAGE LINKS**

🔔 **EXTERNAL LINKS WITH NOFOLLOW ATTRIBUTES**

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. But you have to be careful when using canonical tags. Make sure that the rel="canonical" element does not 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.

Redirect chains and loops will confuse crawlers and frustrate users with increased load speed. You also lose a bit of the original pagerank 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 linking on your website periodically.

🔥 **4XX ERRORS**

🔥 **5XX ERRORS**

🔥 **BROKEN INTERNAL LINKS**

🔥 **WWW DOMAIN CONFIGURED INCORRECTLY**

🔥 **REDIRECT CHAINS AND LOOPS**

🔥 **BROKEN EXTERNAL LINKS**

🔥 **BROKEN CANONICAL LINK**

🔥 **MULTIPLE CANONICAL URLS**

⚠️ **TEMPORARY REDIRECTS**

⚠️ **INTERNAL LINKS WITH NOFOLLOW ATTRIBUTES**

⚠️ **TOO MANY ON-PAGE LINKS**

🔔 **EXTERNAL LINKS WITH NOFOLLOW ATTRIBUTES**

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. Almost every site contains some utilitarian pages that have no place in search index and the sitemap is a way of highlighting the landing pages you want to end up on the SERPs. Sitemap does not guarantee that the listed pages will be indexed, and those that are not mentioned will be ignored by search engines, but it does make the indexing process easier.
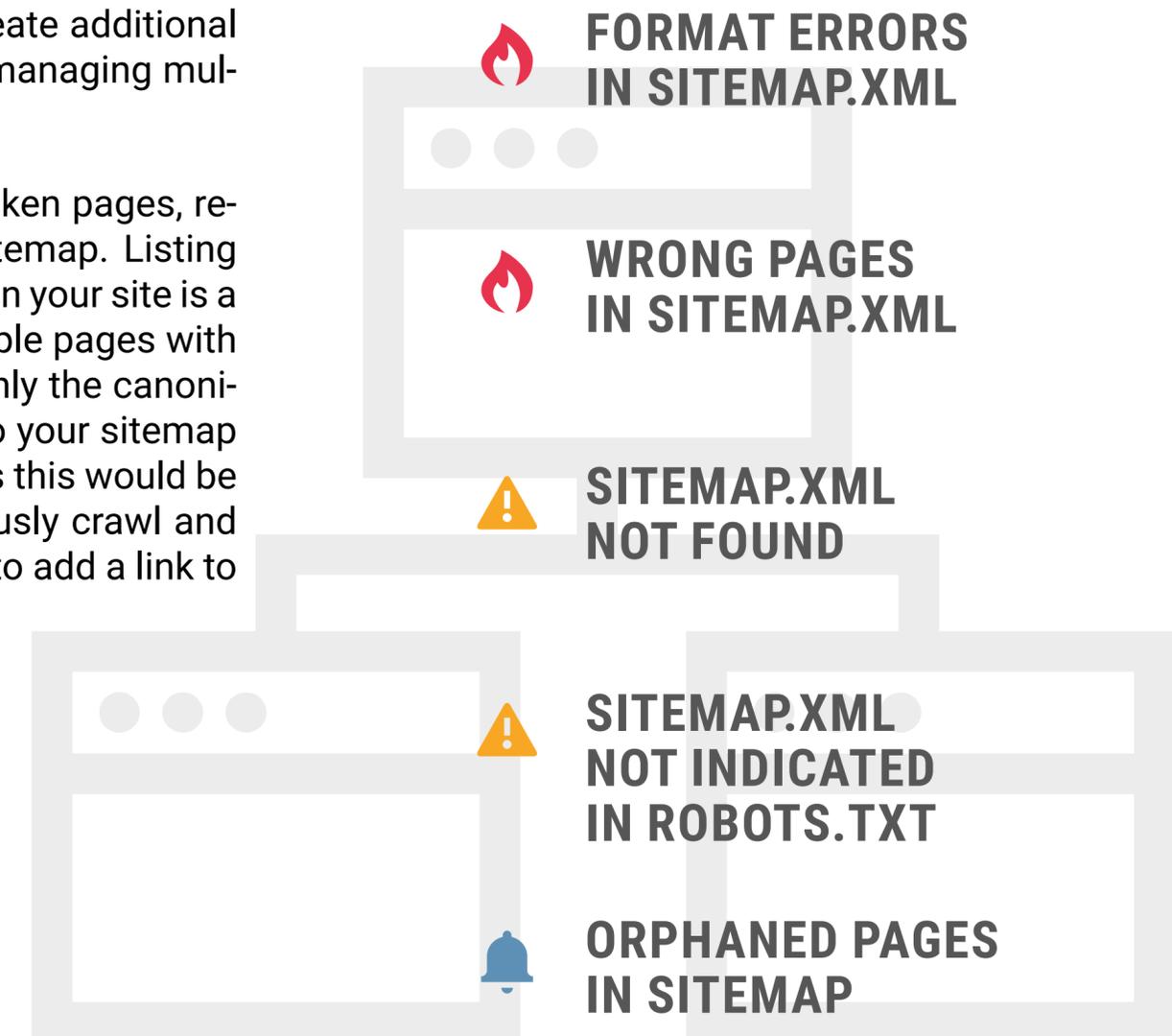
You can create an XML sitemap manually, or generate one using a CMS or a third-party tool. Search engines only accept sitemaps that are less than 50 MB and contain less than 50,000 links, so if you have

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

Obviously there should not be any broken pages, redirects or misspelled links in your sitemap. Listing pages that are not linked to internally on your site is a bad practice as well. If there are multiple pages with the same content, you should leave only the canonical one in sitemap. Do not add links to your sitemap that are blocked with the robots file, as this would be like telling a searchbot to simultaneously crawl and not crawl the page. But do remember to add a link to your sitemap to robots.txt.

To learn more on the correct implementation of the sitemap, look into the official guide.

FORMAT ERRORS IN SITEMAP.XML

WRONG PAGES IN SITEMAP.XML

SITEMAP.XML NOT FOUND

SITEMAP.XML NOT INDICATED IN ROBOTS.TXT

ORPHANED PAGES IN SITEMAP

ON-PAGE SEO

On-page SEO is about improving the rankings of specific pages by optimizing their content and HTML behind them. You need to fastidiously craft all the ingredients of a page in order 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 is 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 content displayed on the page and meta tags concealed in the code.

FIND ALL ON-PAGE SEO MISTAKES WITH THE SEMRUSH SITE AUDIT TOOL

Start your audit →
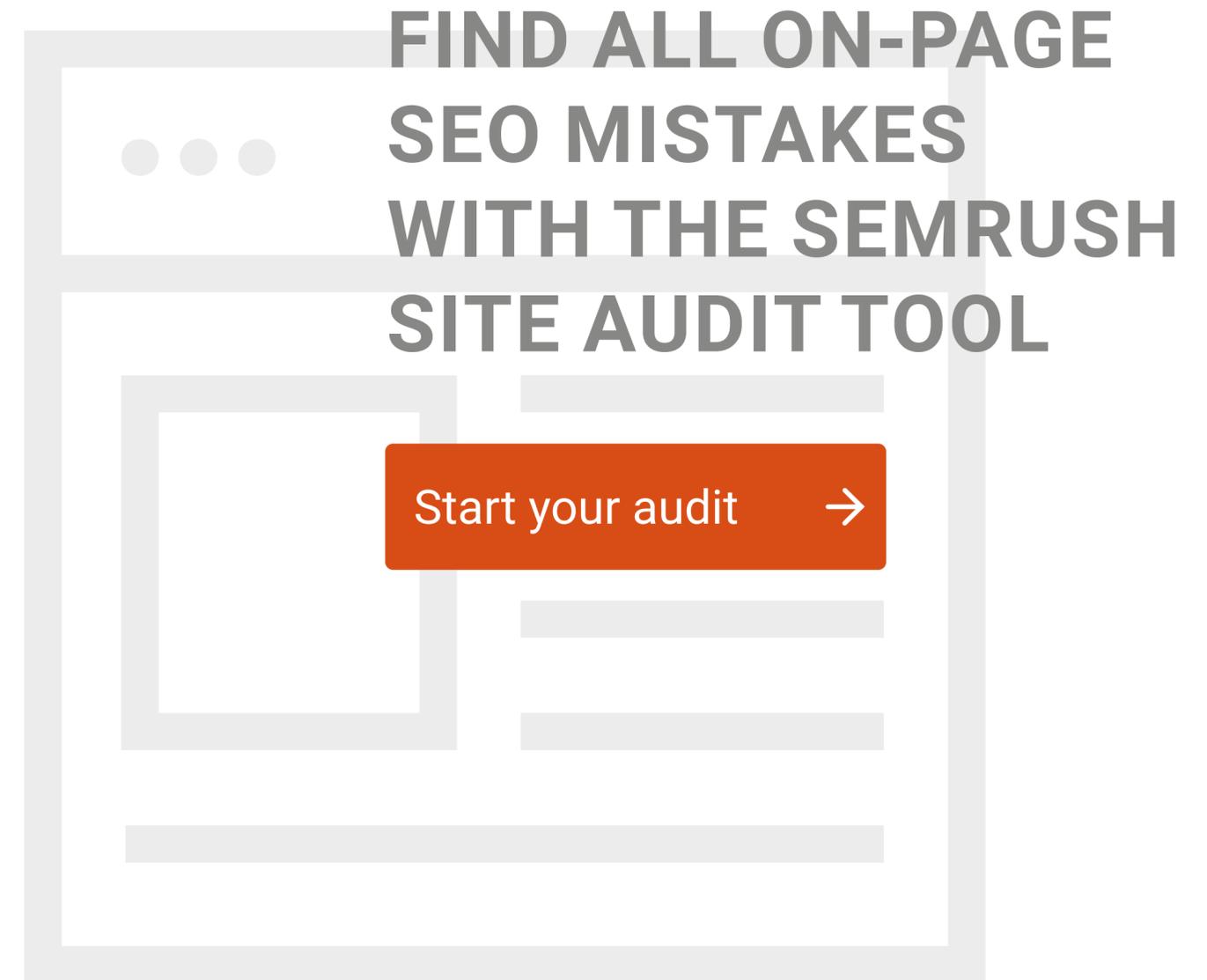
It is well known that good SEO means good content. Rehashed or even copied content is rarely valuable to users and can significantly affect rankings. So you have to inspect your website for identical or nearly identical pages and remove or replace them with unique ones. We advocate that pages have at least 85% unique content.

If under certain circumstances duplicate content is appropriate, in order to avoid cannibalization (multiple pages targeting the same keywords) you have to indicate secondary pages with a rel="canonical" tag that links to the main one. It is a common distress of the e-commerce portals where product pages and variously sorted lists of products appear as duplicates. And sometimes when a URL has parameters, it might get indexed as a separate page, thus creating a duplicate. To prevent that from happening, you need to 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 do not create content just for the sake of content.

A low text-to-HTML ratio can also indicate a poor-quality page. Our advice here is that a page should contain more than 10% of the actually displayed text in relation to the code. But again, if you think that a low word count is acceptable for a specific page, then no worries. But be cautious if this issue is triggered on a page with a lot of content. While having excessive HTML code is not critical, you should try streamlining it to contribute to faster loading and crawling speed.

DUPLICATE CONTENT

LOW TEXT-TO-HTML RATIO

LOW WORD COUNT ON PAGE
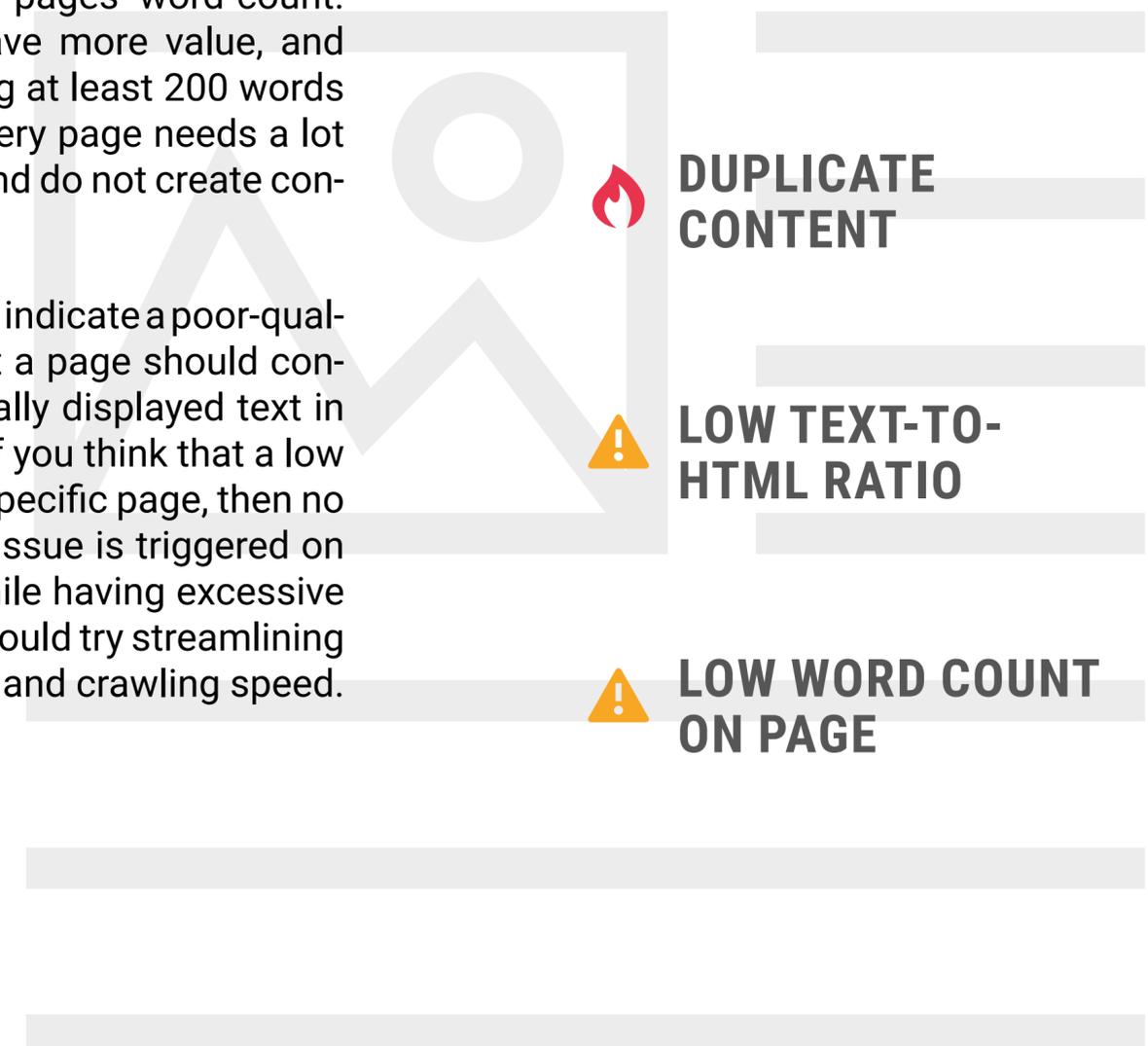
TITLE TAG

The importance of your title tag is pretty obvious – generally it is the first impression you will make on a person browsing the search engine results page. So you need to create captivating, but more importantly, individual meta titles for every page to orient searchers and crawlers. Duplicated titles can confuse users as to which webpage they should follow.

Make your title tags concise. Brevity is the soul of wit and all, but in particular, you need to do this because titles that are too long might get automatically cropped. However, you should recognize that short titles are usually uninformative and rob you off the opportunity to inject more delicious wordage to lure customers. To keep your title tags balanced, you should typically strive for about 50-60 characters. However, the space allotted for titles on results page but it is actually about pixels these days, so keep an eye out for wide letters like "W."

After you have written a perfect, simple and descriptive title, you still need to watch out for Google's reaction, since it might not find your masterpiece relevant to the page or the query, and completely rewrite it. There is also a chance that Google will add your brand name to the title, casually cutting off its ending.

MISSING OR EMPTY TITLE TAGS

DUPLICATE TITLE TAGS

TOO LONG TITLE TAG

TOO SHORT TITLE TAG

H1 TAG
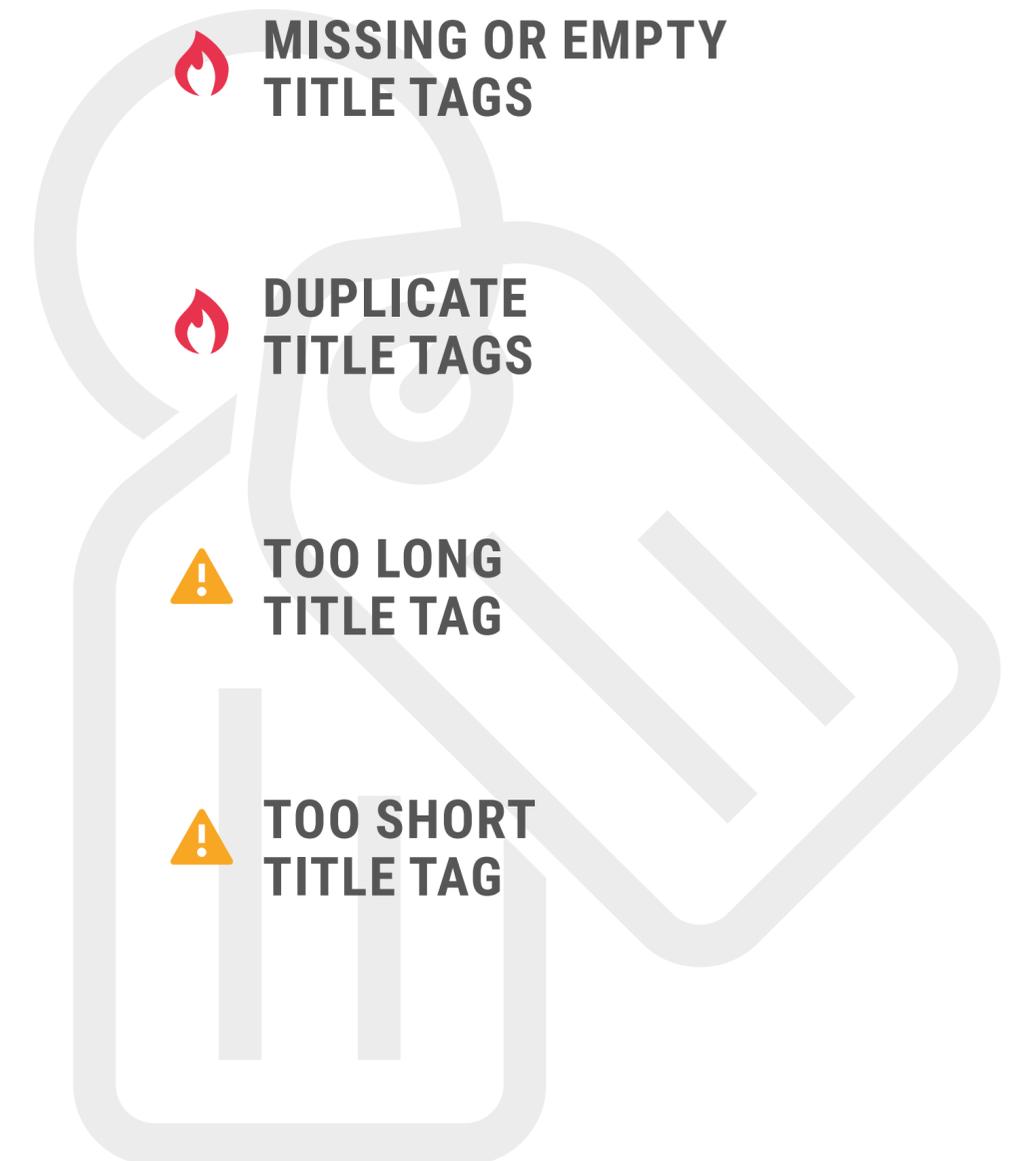
A website's H1 heading is less important than its title tag, but it still helps crawlers and users, and stands out visually on a page. The H1 and the title tag can be identical, which is acceptable from a technical standpoint but not a good SEO practice. When your H1 and title are the same you are missing the chance to diversify semantics with varied phrases and it makes your page look overly optimised. Give some thought to your H1s – make them catchy, yet simple and relevant.

Search bots use H1 to get a hint as to what is your page about, so do not distract them by putting multiple H1s on a single page, instead use an H2-H6 hierarchy for descending subsections. These subheadings are far less important than the H1 and are placed mostly for users rather than crawlers. Structured text is better at holding readers' attention, and a clear layout ensures easier information consumption and, creates an overall better user experience. 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 HEADING**
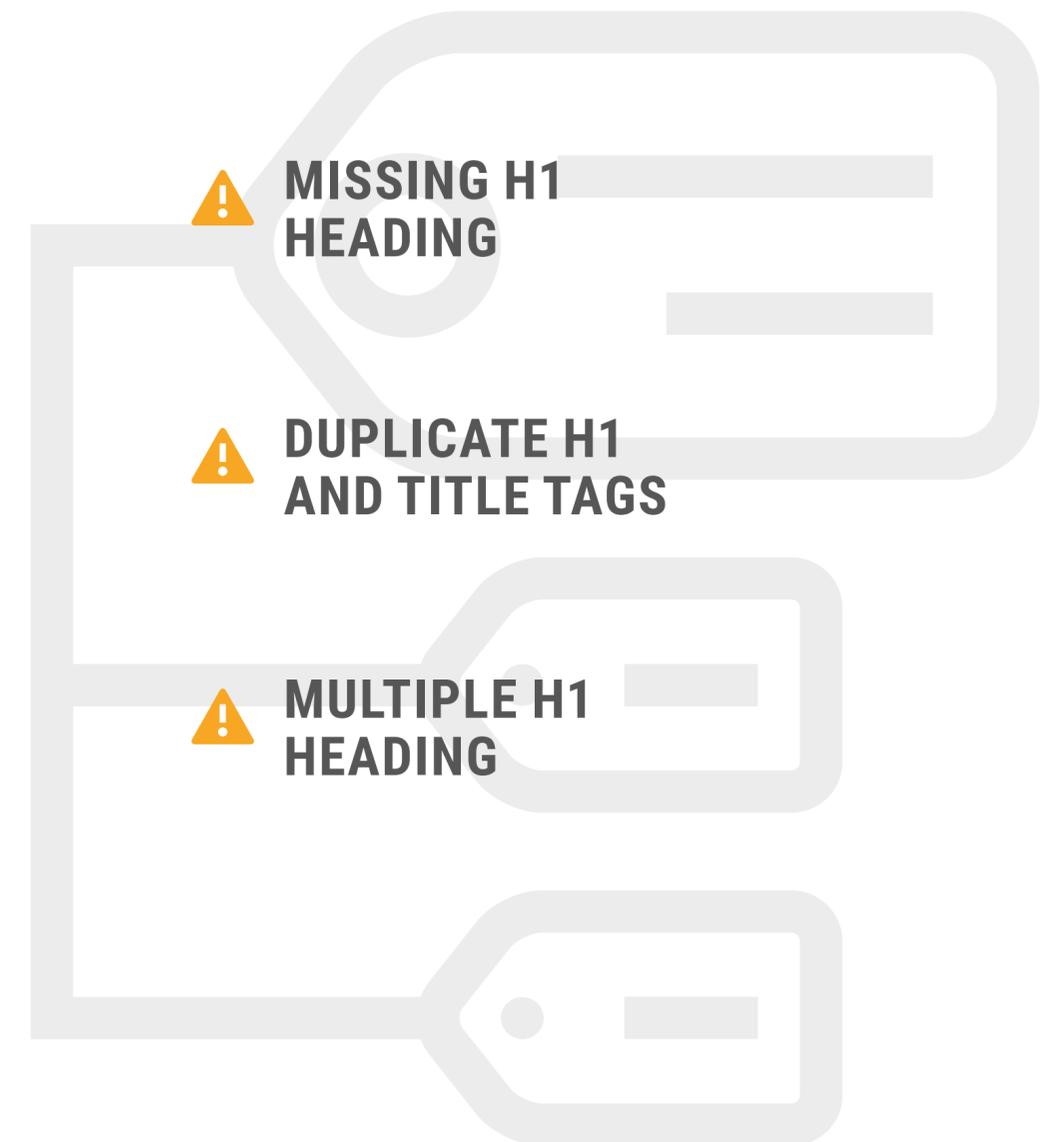
META DESCRIPTION

If your page's title tag is the proverbial book cover that it is judged upon in the search results, then your meta description is the back cover that sells it for a click. Of course, a missing meta description will not affect your rankings – Google will make one up for you. But the result will probably not be the most relevant or flashy, which may, in turn lower your potential CTR.

Although, on many occasions, it might be inconvenient and unnecessary to come up with a unique description for each page. In that case you should concentrate on the most important landings and leave all the rest with auto-generated.

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

DUPLICATE META DESCRIPTIONS

MISSING META DESCRIPTION

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

# IMAGES

Image searches are nothing new, and while top ranks in an image SERP can bring a chunk of a target audience to your website, image SEO is still neglected by some website owners. We will talk more about image optimization in the following section on page speed. For now let's look solely at the SEO aspects of an image; which are 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

is hard to reach an audience for clothing shops, food delivery, hotels, etc.

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 do not 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 will also greatly aid visually impaired people.

BROKEN INTERNAL IMAGES

**BROKEN EXTERNAL IMAGES**

MISSING ALT ATTRIBUTE
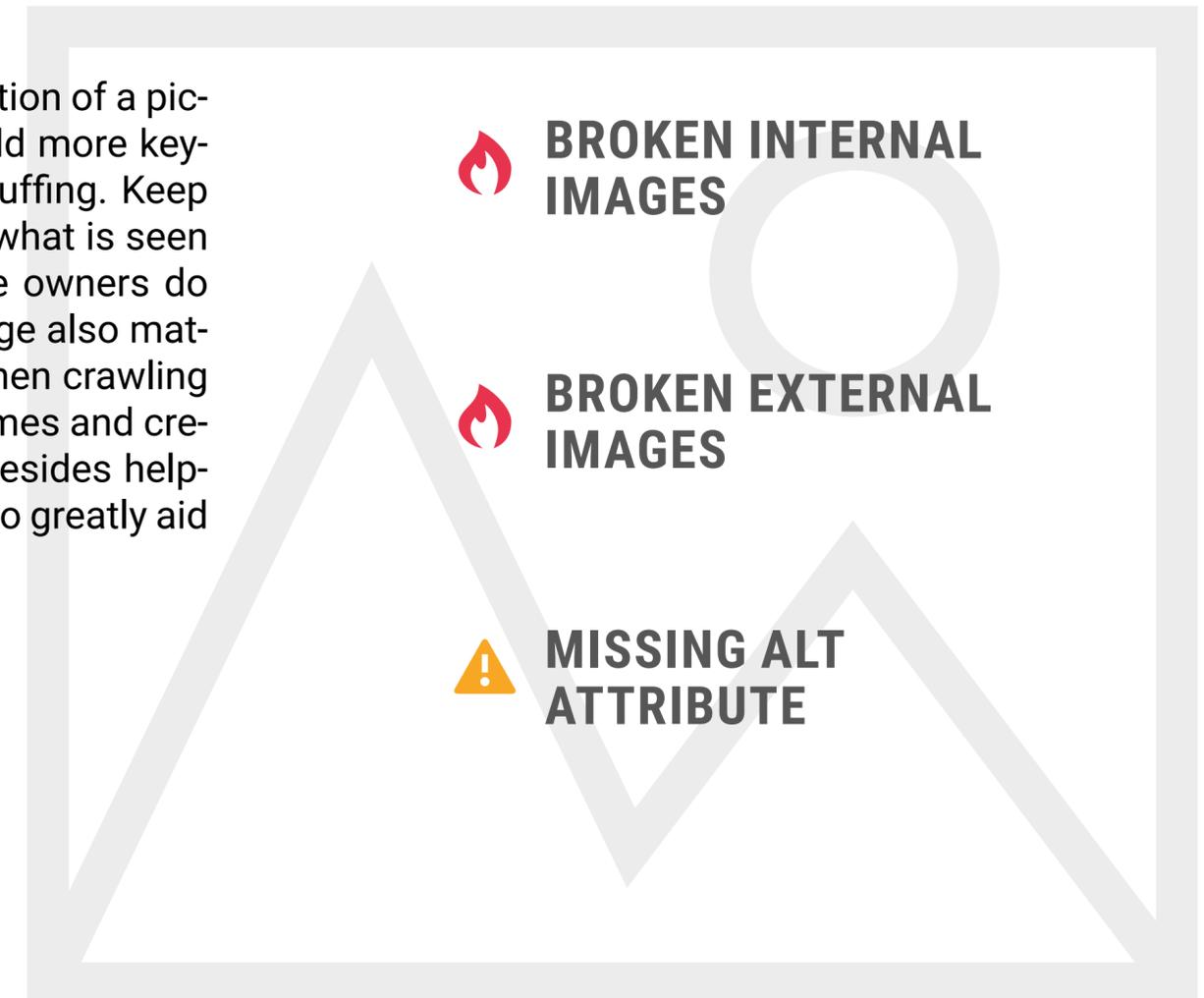
TECHNICAL SEO

Technical SEO deals with things apart from content that affect user experience and rankings. This include a slow page loading speed, utilization of outdated technologies and inadequate optimization for mobile devices. These are aspects of a website audit that you need to pay extra attention to, because poor page performance can bring to naught all the good SEO work that you have done.

On the other hand, the outcome of fixing technical issues can be highly rewarding. Most technical mistakes have a site-wide nature, so fixing them usually benefits not only a single page but the whole website as well. Oftentimes just a little tweaking can drastically increase your traffic and save you a lot of money.
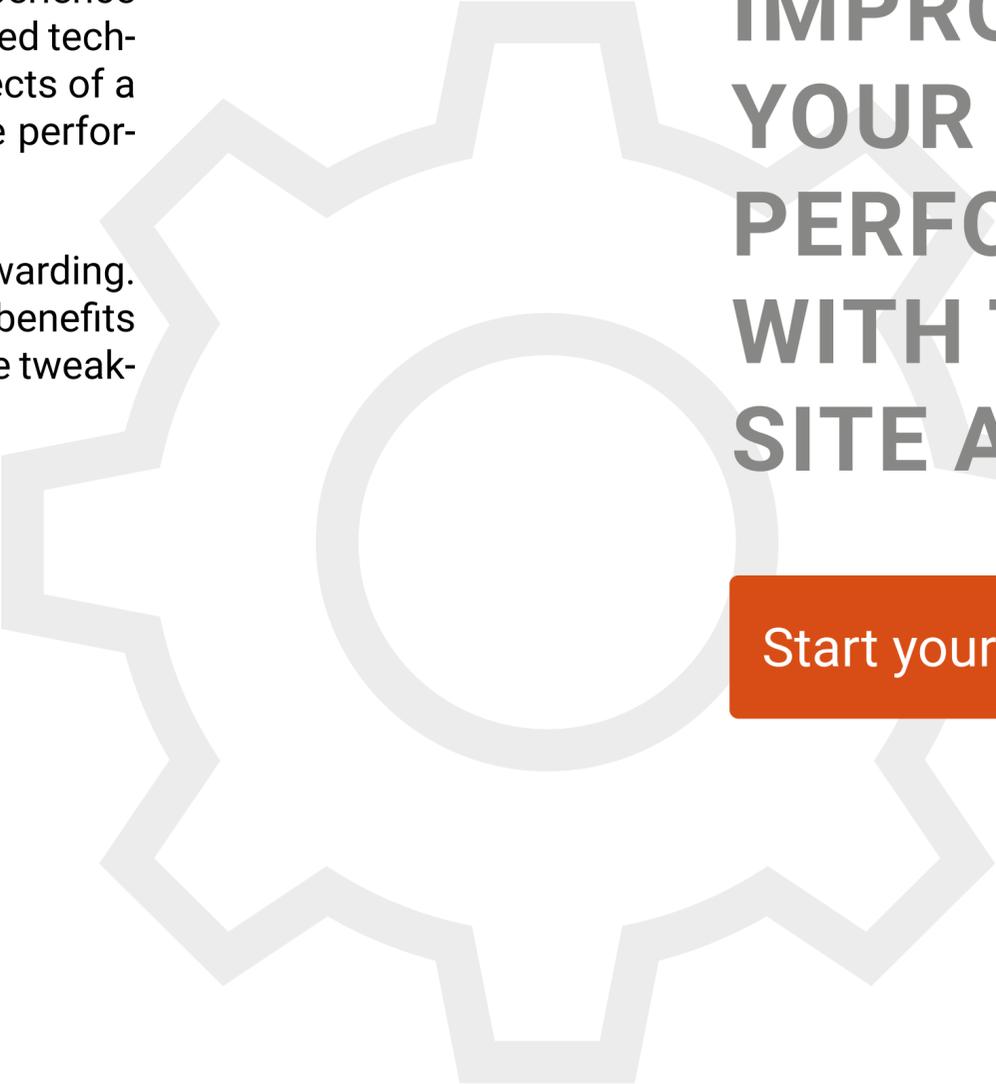
IMPROVE YOUR WEBSITE PERFORMANCE WITH THE SEMRUSH SITE AUDIT TOOL

Start your audit →

PAGE SPEED

Page speed is a big ranking factor affected both by the server side and page performance. And it is a big bounce rate cultivator for obvious reasons. So you need to optimize HTML, reduce scripts and styles, and try to keep page size to a minimum. One way to achieve this is using compression schemes like gzip or deflate. Condensing HTML, CSS and Javascript can greatly benefit load speed, but there are drawbacks of complicated set up 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. There is a lot to contemplate – image quality and resolution, its format and more, but before looking at all that, you have to consider if visual content is actually necessary for your page. If the answer is yes, then fine-tune your images using a graphic design tool of your choice. Try to achieve the smallest filesize you can while maintaining acceptable image quality. Examine the possibility of using vector graphics. It is a great way to slim down simple geometrical images. If a large image file is not absolutely necessary to the message of the page, then you can consider removing it to improve page speed. Lastly, since mobile page speed is even more important, you have to configure viewport and rescale images for different screens.

For more extensive information consult Google's recommendations for page speed optimization.

LARGE HTML SIZE

**SLOW PAGE LOAD SPEED**

UNCOMPRESSED PAGE
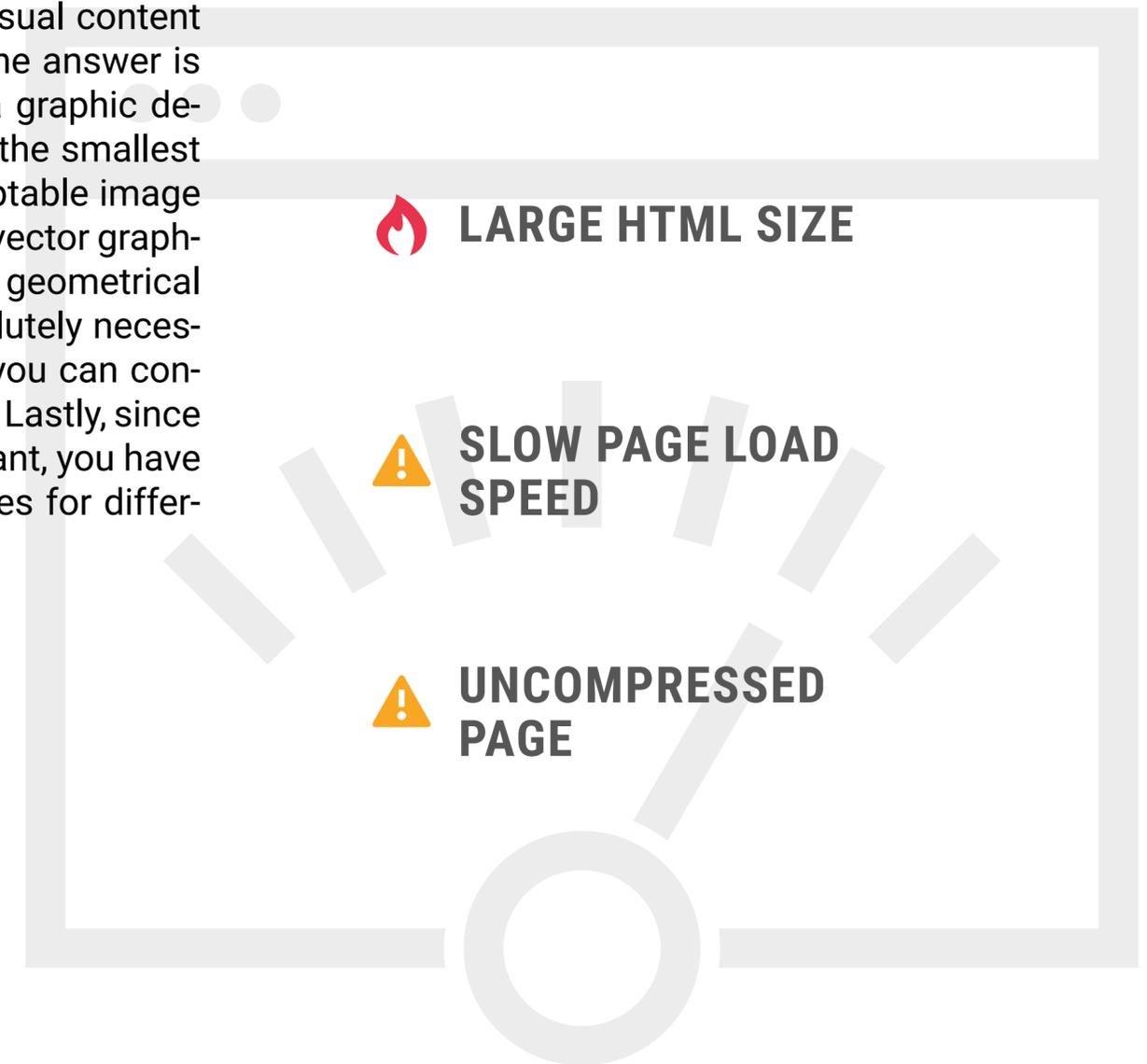
Evolution of the Internet never stops. And just as some species become extinct for others to thrive, some technologies have to go for the sake of progress. The death of Flash was a long time coming, and for good reason. From an SEO perspective (although it might give a more vibrant look to your website) Flash impoverishes a page's performance, and handicaps crawling. Adobe announced that it will stop supporting its technology by the end of 2020.

As for installing widgets and plugins from external domains with IFrames – they can come in handy and will not affect your rankings if implemented properly, but can also hurt your website's usability 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 <!DOCTYPE> tag. Give it special attention if you are using an older version of a code.

⚠ **FLASH CONTENT USED**

⚠ **FRAMES USED**

⚠ **DOCTYPE NOT DECLARED**
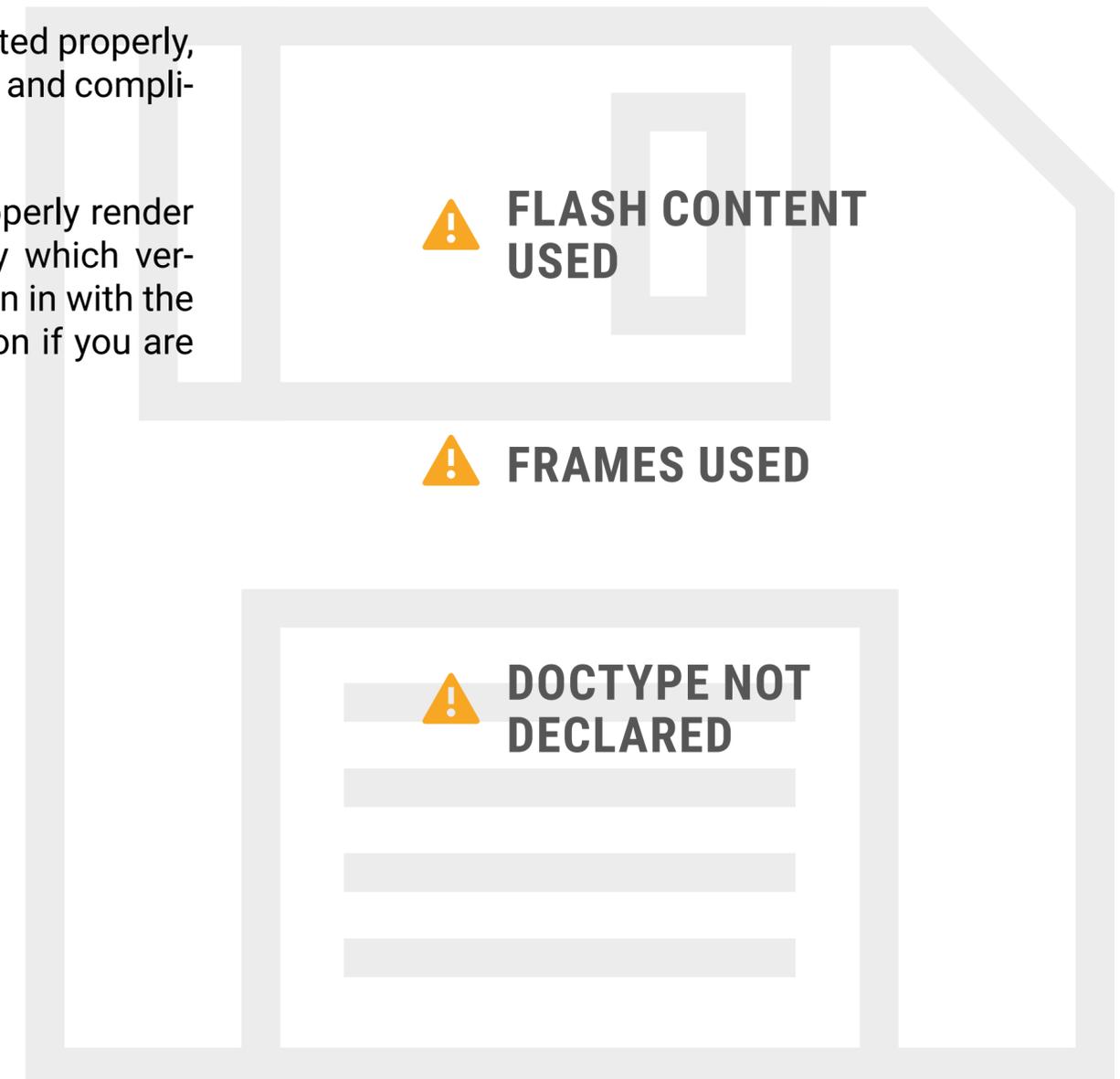
We are all optimizing for mobile devices, right? So checking that all your pages have viewport tags and can scale for various screen sizes is imperative. If a page does not have a viewport meta tag, mobile browsers will not be able to find the optimized version of the page and will show the desktop version with the font too small or too big for the screen and all the images jumbled. There are no two ways about it – this 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 is also a platform for e-commerce 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, and templating issues. SEMrush covers all three of these groups, and detects over 40 AMP errors. You can crawl the whole site on demand whenever you like and get all of the broken pages in one report. You can also choose to crawl AMP versions first.

MISSING VIEWPORT TAG

AMP PAGES WITH HTML ISSUES

AMP PAGES WITH STYLE AND LAYOUT ISSUES
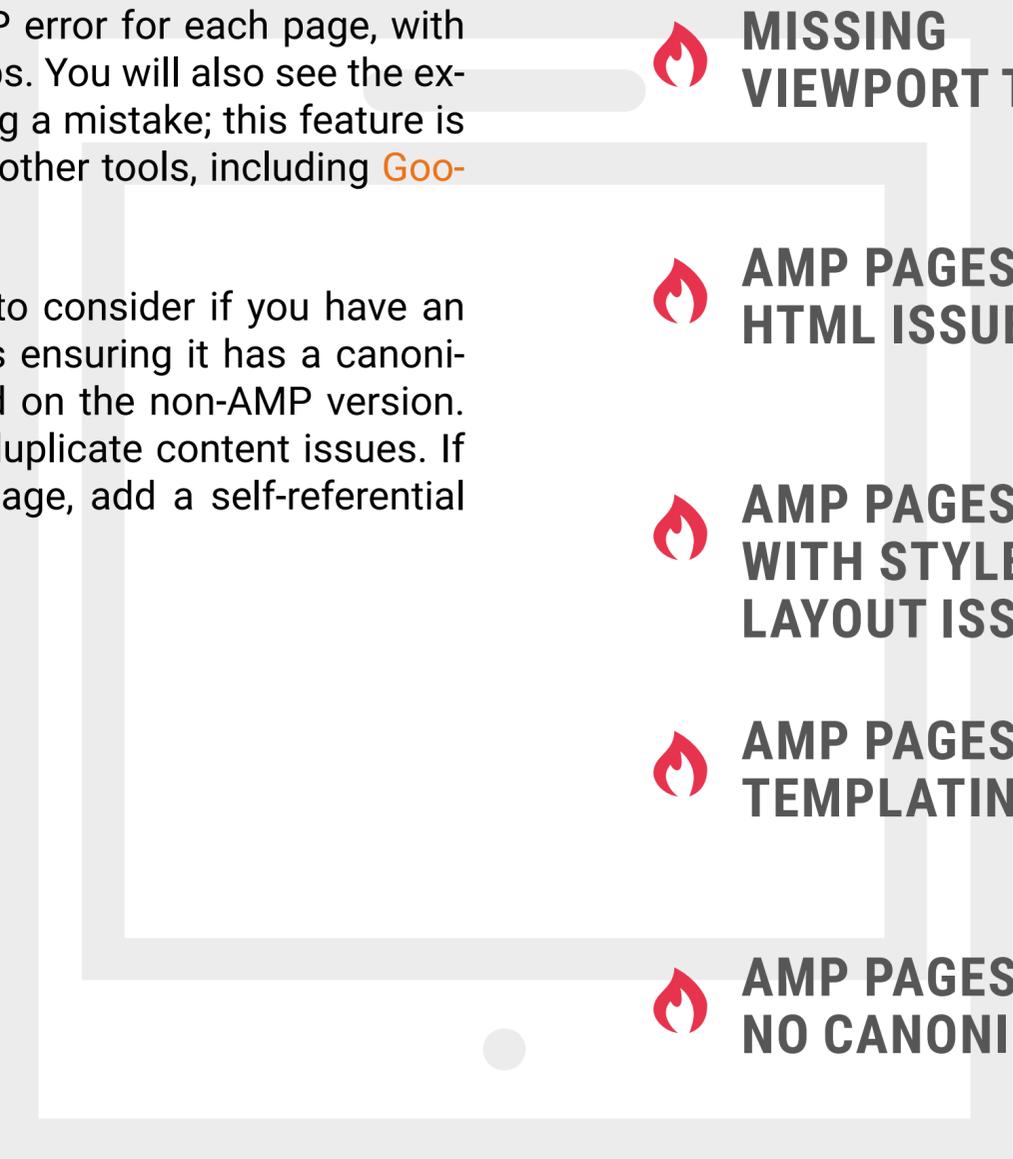
AMP PAGES WITH TEMPLATING ISSUES

AMP PAGES HAVE NO CANONICAL TAG

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 standardisation. 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. In SEMrush Site Audit you will get a detailed list of every AMP error for each page, with descriptions and fixing tips. 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**

🔥 **AMP PAGES WITH HTML ISSUES**

🔥 **AMP PAGES WITH STYLE AND LAYOUT ISSUES**

🔥 **AMP PAGES WITH TEMPLATING ISSUES**

🔥 **AMP PAGES HAVE NO CANONICAL TAG**

HTTPS IMPLEMENTATION

HTTPS is a necessity for every website. You have to protect yourself and your users from those pesky, malicious people on the Internet by ensuring that all the data transferred through your website is authentic, encrypted and intact. And of course there is a perk of Google's favouritism toward secured pages.

HTTPS is a ranking factor which 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.
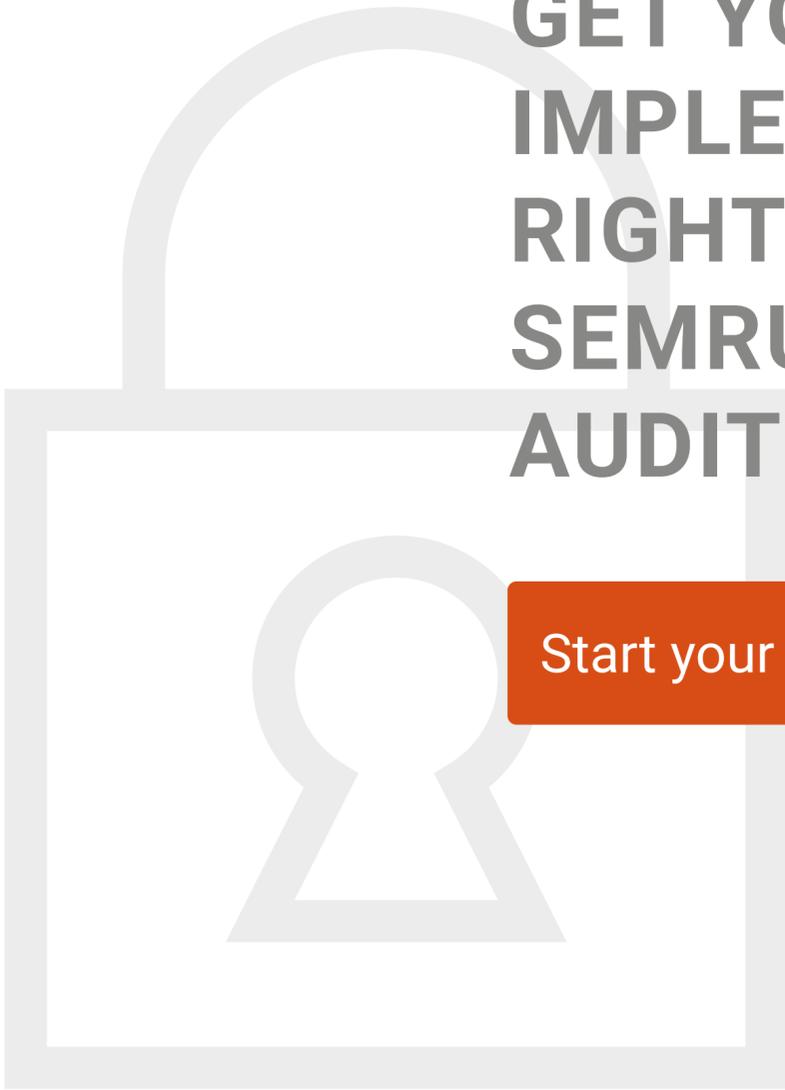
GET YOUR HTTPS IMPLEMENTATION RIGHT WITH THE SEMRUSH SITE AUDIT TOOL

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 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 is 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 PAGE

NO REDIRECTS OR CANONICALS TO HTTPS URLS

SSL CERTIFICATE REGISTERED TO AN INCORRECT DOMAIN NAME

OLD SECURITY PROTOCOL VERSION
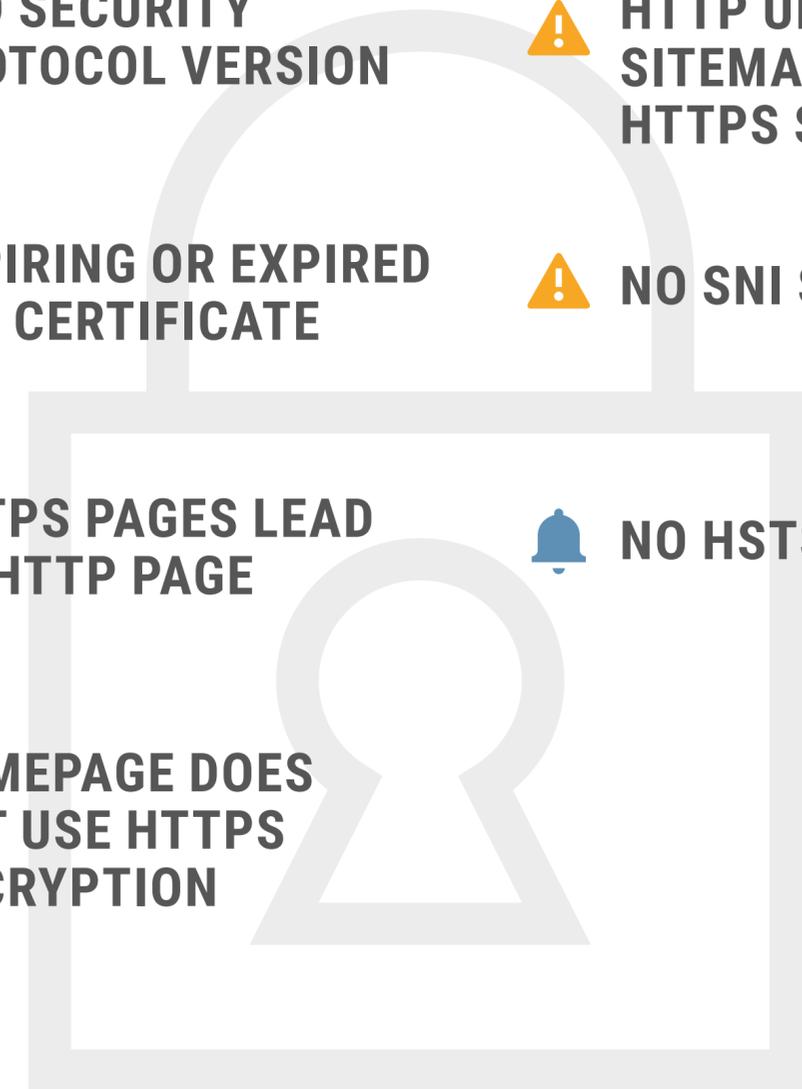
EXPIRING OR EXPIRED SSL CERTIFICATE

HTTPS PAGES LEAD TO HTTP PAGE

HOMEPAGE DOES NOT USE HTTPS ENCRYPTION

HTTP URLS IN SITEMAP.XML FOR HTTPS SITE

NO SNI SUPPORT

NO HSTS SUPPORT

INTERNATIONAL SEO

The Internet makes the world small, globalization never stops, and international SEO is becoming more relevant than ever. Creating sites in more than one language is not a prerogative of big corporations, and smaller web portals can also gain a lot by geographic expansion.

Maintaining a multilingual website creates a specific set of potential problems. It is 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 it that you are not just scattering duplicates around.

FIX ALL HREFLANG IMPLEMENTATION ISSUES WITH THE SEMRUSH SITE AUDIT TOOL

Start your audit →

When configuring a 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 is also important to declare encoding so that browsers will know which set of characters must be used.

The main SEO problems of an international website 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 searchbot 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.

Even if you feel that all of your page's redirects and hreflang tags are perfectly implemented, it is still a good idea to add the option to select a language. Be careful if you are using automatic translators to create content. The result might be unreadable and even nonsensical, which will be noted by crawlers and, obviously, by readers.

🔥 **HREFLANG IMPLEMENTATION ISSUE**

🔥 **HREFLANG CONFLICTS WITHIN PAGE SOURCE CODE**

🔥 **INCORRECT HREFLANG LINKS**

⚠ **LANGUAGE IS NOT SPECIFIED**

⚠ **NOT DECLARED ENCODING**

🔔 **HREFLANG LANGUAGE MISMATCH ISSUES**

We hope this PDF will help you polish your website! Our team is continuously working to perfect our products' existing features and to develop new ones. Please share your thoughts and suggestions by reaching out at:

site-audit-feedback@semrush.com