

# THE ULTIMATE SITE AUDIT WITH SEMRUSH

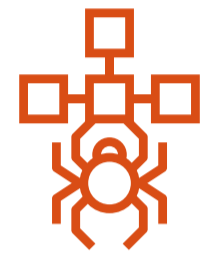


***SEMRUSH***

# TABLE OF CONTENTS



## Introduction



## Crawlablity and Site Architecture

Robots.txt

URL structure

Links & Redirects

Sitemap



## On page SEO

Content

Title tag

H1

Meta descriptions

Images



## Technical SEO

Page speed

Old technology

Mobile



## HTTPS Implementation



## International SEO

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**

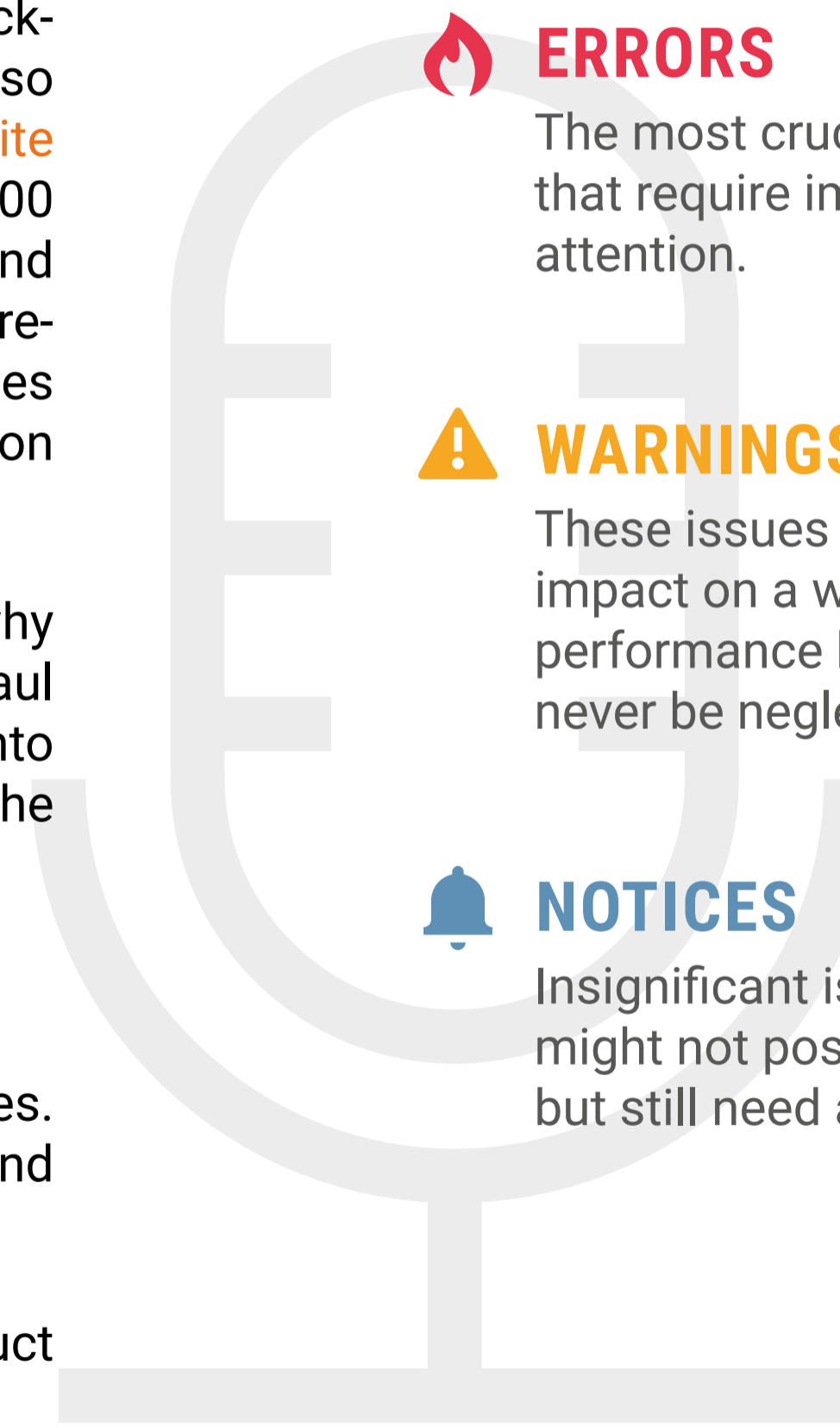
The most crucial issues that require immediate attention.

## **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.

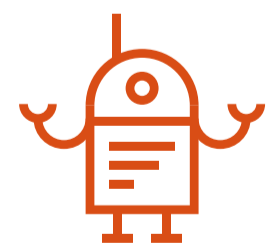


**LEVEL UP  
CRAWLABILITY  
OF YOUR WEBSITE  
WITH THE SEMRUSH  
SITE AUDIT TOOL**

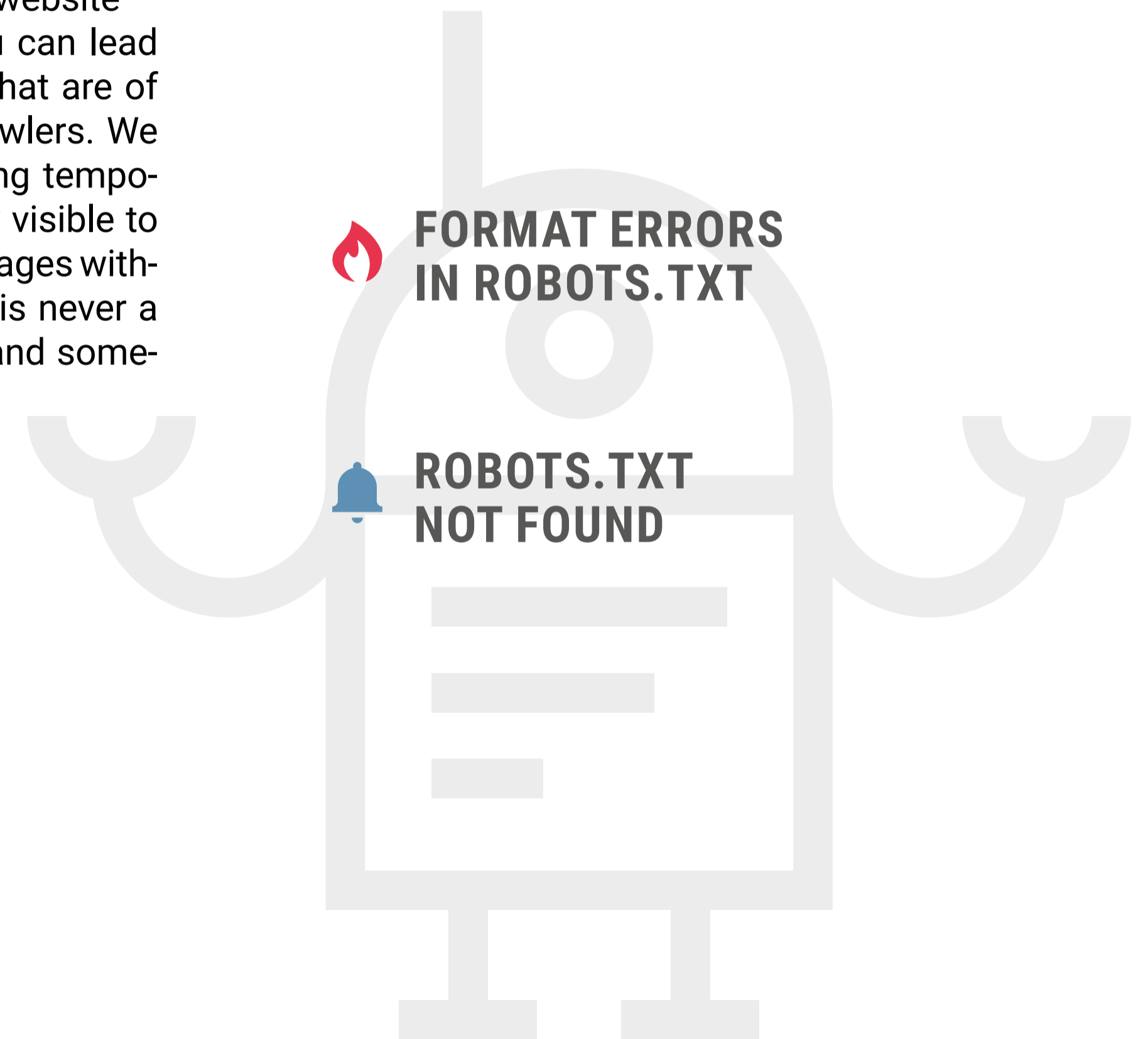
[Start your audit →](#)

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.



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](#).

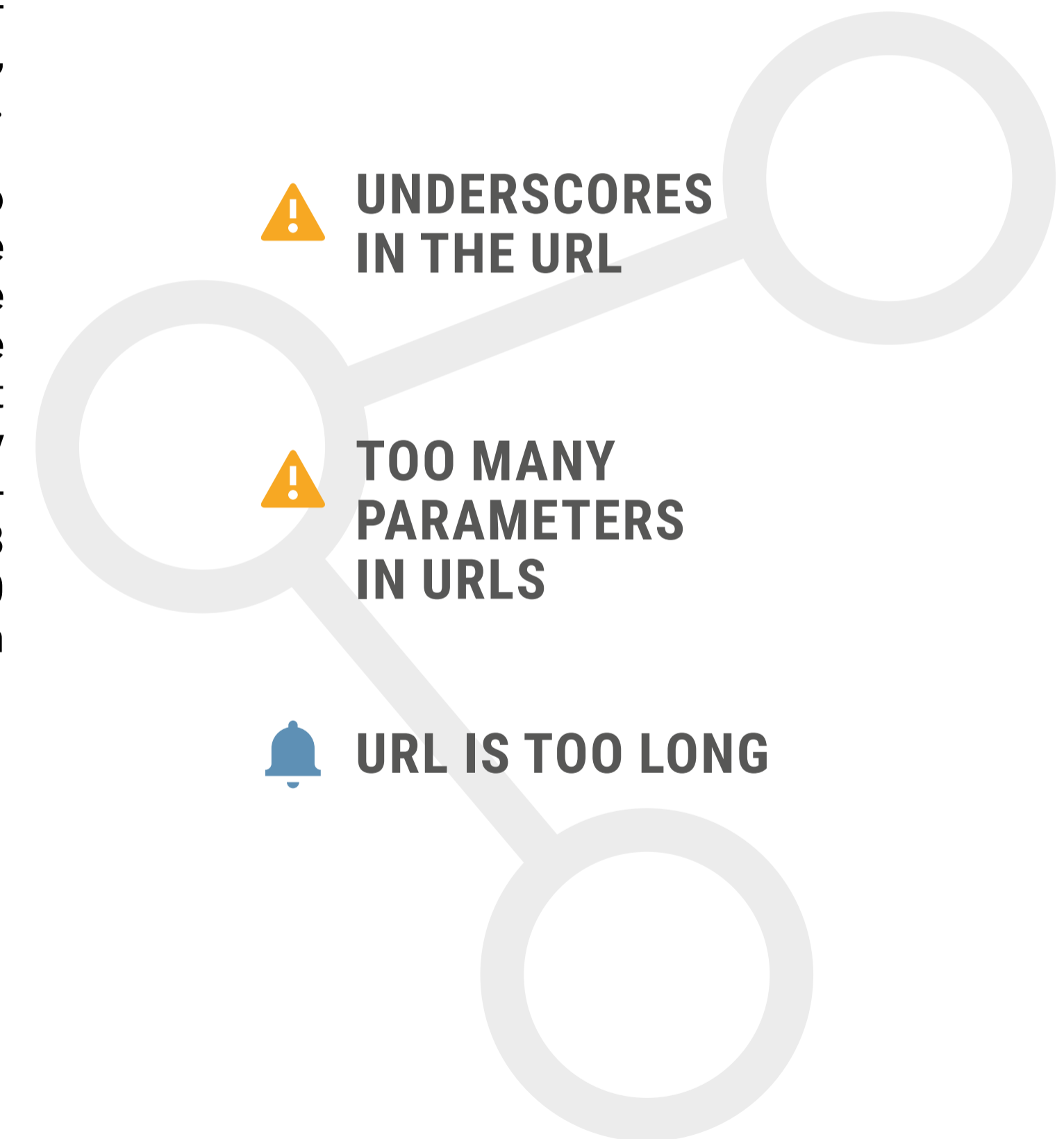


For an SEO specialist **URL is more than just the address of a webpage**. If left unattended, they can negatively affect indexing and ranking. Crawlers and people alike will read URLs, so use relevant phrases in URLs to indicate what the page's content is about. You can have the URL match the title, 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.



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**

**This is just an introductory excerpt**

You can download the full version here:

<https://www.semrush.com/ebooks/ultimate-site-audit-with-semrush/>