

Duplicate content

Table of contents

[Learning objectives](#)

[Introduction](#)

[1. What is duplicate content?](#)

[2. How to find duplicate content?](#)

[3. How to solve duplicate content issues](#)

[Key takeaways](#)



Learning objectives

After this lesson, you'll be able to:

- define duplicate content;
 - find duplicate content issues for your shop;
 - solve duplicate content issues for your shop.
-

Introduction

If you want your products to appear in the search engines, you need to write content for them and give them a unique place on your website. Otherwise, you'll run into the SEO problem called duplicate content. What is that, and is there a way to solve it? That's what we'll explore in this lesson!

1. What is duplicate content?

Duplicate content is content that lives in several locations, i.e. URLs. It can harm your rankings, and many people say that copious amounts of it can even lead to a penalty by Google. That's not true, though. There is no penalty, but having loads of duplicate or copied content can get Google to influence your rankings negatively.

Duplicate content is a bit like being at a crossroads where road signs point in two different directions for the same(ish) destination. So, you wonder: which road should you take? To make matters worse, the final destination is different too, but only ever so slightly. As a reader, you don't mind because you get the content you came for, but a search engine has to pick which page to show in the search results because, of course, it doesn't want to show the same content twice.

If this sounds too abstract, here's an example. Let's say your page about 'keyword X appears at

<http://www.example.com/keyword-x/>

and the same content also appears at

<http://www.example.com/article-category/keyword-x/>.

Then let's say several other sites have picked up your content, and some of them link to the first URL, while others link to the second. This is a problem because those links both promote different URLs. If they were all **linking to the same URL**, your chances of ranking for 'keyword X would be higher.

2. How to find duplicate content?

Most of the time, if you have duplicate content issues, they will be on your site. Sometimes, however, other websites use your content, with or without your consent. They don't always link to your original article, and therefore the search engine doesn't 'get' it and has to deal with two versions of the same content.

In addition, very often, ecommerce sites have duplicate content issues with their product descriptions. In that case, we are not talking about copied or scraped content. Rather, **resellers often use the manufacturer's description on their product pages**. The problem is, if there are many of those resellers, there will be a lot of the same content all over the web (Image 1). So, let's explore how you can find duplicate content.

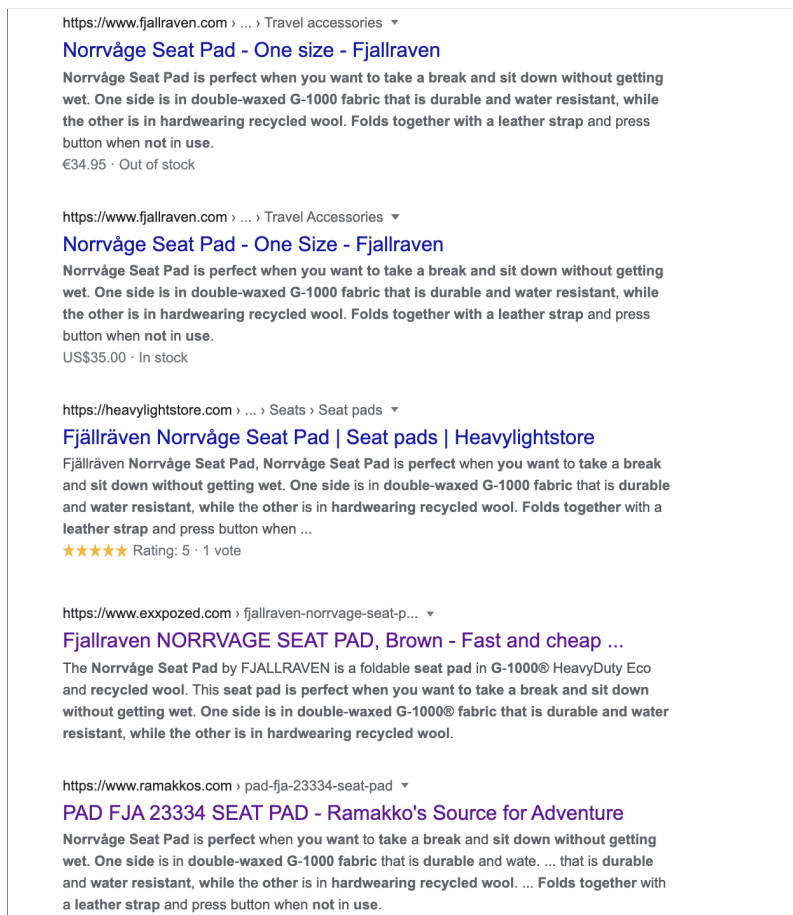


Image 1: An example of shops using the same manufacturer's description for their product

Finding duplicate content on your site

You might not know whether you have duplicate content issues on your site.

Using Google is one of the easiest ways to check that. If you'd want to find all the URLs on your site that contain keyword X, type the following search phrase into Google:

```
site:example.com intitle: "Keyword X"
```

Google will then show you all pages on your site that contain that keyword. The more specific you make that `intitle` part of the query, the easier it is to weed out duplicate content.

Alternatively, you can **use tools**. For example, [Siteliner](#) searches for duplicate content on your own site. This tool comes in a free and a premium version. The free version of the Siteliner duplicate content check is limited to 250 pages every 30 days.

Finding duplicate content across the web

You can use the same method to identify **duplicate content across the web**. Let's say the full title of your page was 'Keyword X – why it is awesome', you'd search for:

```
intitle: "Keyword X – why it is awesome"
```

And Google would give you all sites that match that title. So, you'll be able to see if anyone has copied your content without your consent. Sometimes it's even worth **searching for one or two complete sentences from your page**. If you have a certain page that you'd like to check, simply go to that page. Copy a text snippet, preferably from a section that you think might be attractive for others to copy, or that might appear on some other sites. Insert the exact snippet in Google between double quotation marks. The results you get will tell you if the content you search for appears on multiple URLs.

There are also tools that can help you discover if someone is copying your content. One of the best-known duplicate content checkers is probably [CopyScape.com](#). This tool works pretty easily: insert a link in the box on the homepage, and CopyScape will return a number of results, presented a bit like Google's search result pages.

3. How to solve duplicate content issues

There are different ways to solve duplicate content issues on your site. For example, you can merge duplicate pages together and redirect them to the main URL. Or, you can use the `rel="canonical"`. Let's explore these options in more detail.

Redirecting duplicate content

Did you find any pages with the same, or similar content on your site? One course of action could be to **merge those pages**. Make one really awesome page out of the two (or three) you have written. Then, pick one page that you want to keep, and delete the others. Ideally, pick the URL that got the most traffic. However, this is very important, **don't delete the pages without redirecting them!** But wait, what is redirecting?

The name 'redirect' pretty much says it all. When you redirect a page, **you send visitors traveling to a specific page to an alternative one**. There are many types of redirects. But, when you merge several pages together, it's best to **use the so-called 301 redirect**. A 301 redirect will permanently send visitors and search engines to the new destination.

So, how do you create such a redirect? Well, the easiest and the quickest way is to use the [Yoast SEO Premium plugin](#). The built-in Redirect manager assists you whenever you change the URL of a post, page, or any of the taxonomies that may result in a possible 404 if you don't properly redirect visitors. The plugin also tells you when you're about to create a redirect that will result in a [redirect loop](#). This looping is something you want to avoid at all costs.

Using `rel="canonical"`

Let's assume you have two versions of the same page, each with exactly – 100% – the same content. The only difference is that they're in separate sections of your site. And because of that, the background color and the active menu item are different – but that's it. Both versions have been linked from other sites, so the content itself is clearly valuable. So which version should search engines show in results?

For example, these could be their URLs:

```
https://example.com/wordpress/seo-plugin/
```

```
https://example.com/wordpress/plugins/seo/
```

The situation described above occurs pretty often, especially in a lot of e-commerce systems. A product can have several different URLs depending on how you got there.

However, sometimes you may have the same product in different colors on different URLs. In that case, it's best to **create a category for those products** and make sure to add the category in the URL. That would then be the "main" URL for all those products. That **chosen "main" URL is called a canonical URL**.

This is precisely what rel=canonical was invented for.

How to apply rel=canonical

Applying a rel=canonical may sound very technical, but it is not very complicated. Here is what you need to do:

1. **Pick one page as the canonical version.** This should be the version you think is the most important. If you don't care, pick the one with the most links or visitors.
2. **Add a rel=canonical link from the non-canonical page to the canonical one.** So if we picked the shortest URL as our canonical URL, the other URL would link to the shortest URL in the <head> section of the page – like this:

```
<link rel="canonical"
href="https://example.com/wordpress/seo-plugin/" />
```

What this does is "merge" the two pages into one from a search engine's perspective. It's a "soft redirect" without actually redirecting the user. Links to both URLs now count as the single, canonical version of the URL.

Setting the canonical URL

With the [canonical URL tag feature](#) of the [Yoast SEO for WordPress plugin](#), you can easily add a canonical URL to a post or page. You only need to do this if you want to change the canonical to something different from the current page's URL. Yoast SEO **already renders the correct canonical URL for almost any page type in a WordPress install**. For posts, pages, and custom post types, you can edit the canonical URL in the advanced tab of the Yoast SEO metabox (1 in Image 2) — or in the sidebar of the WordPress block editor (2 in Image 2).

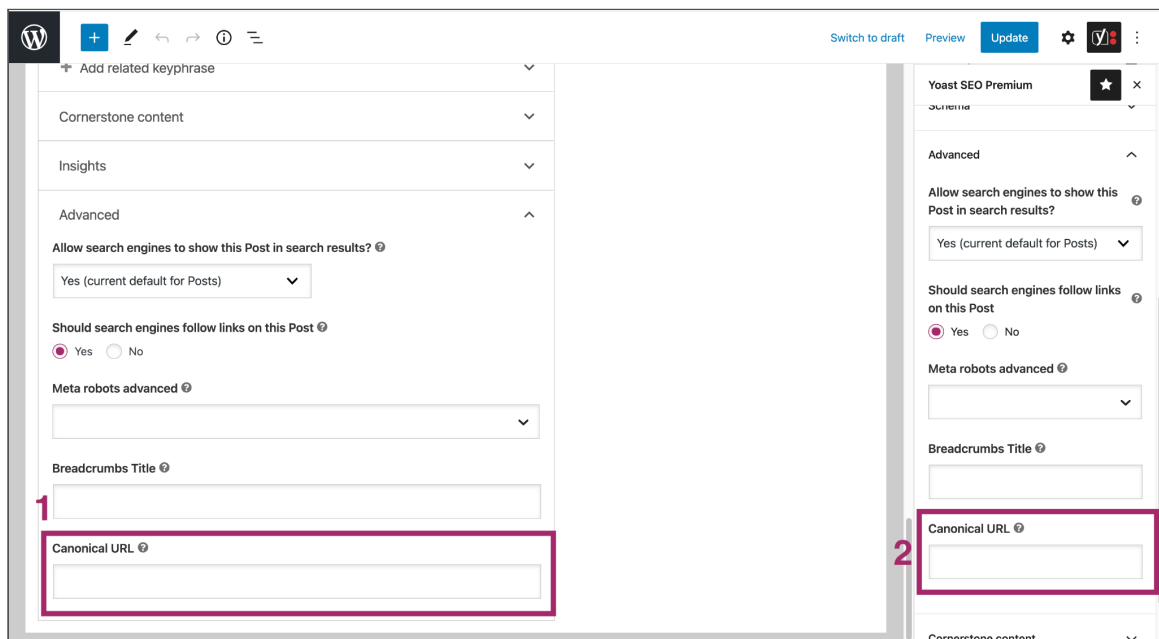


Image 2: Setting a canonical URL in Yoast SEO (outlined)

What if you don't use WordPress? Well, for example, **Shopify automatically generates canonical tags** for product pages with no collection or variant selected.

For example, if you have the following URL :

`https://example.com/collections/product/xy-?variant=123456789`

19

Shopify will create a canonical_url output for :

`https://example.com/product/xy`

Want to learn more about this topic? Read the article [rel=canonical: the ultimate guide](#).

Linking back to the original content

If you can't do any of the above, possibly because you don't control the <head> section of the site your content appears on, **adding a link back to the original article on top of or below the article is always a good idea**. You might want to do this in your RSS feed by adding a link back to the article in it. Some scrapers will filter that link out, but others might leave it in. If Google encounters several links pointing to your original article, it will figure out soon enough that that's the actual canonical version.

Yoast SEO allows you to automatically add content to your RSS feed. To learn more about it read: [How to use RSS variables in Yoast SEO](#).

Write original content

Finally, if you've discovered that a lot of websites use the same product description as you do, there is one powerful tool you can use. It's **your writing skills**! Google is focused on quality. It is always on the lookout for the best possible piece of content that fits the user's intent best. Your goal should not be to make a quick buck but to leave a lasting impression. So, make your content original and of high quality.



Advanced: Some more ways to avoid duplicate content

Are you technically savvy and willing to dig a bit deeper to solve duplicate content issues? Go through the list below, and you might find some useful tips.

- Are there Session IDs in your URLs?
These can often just be disabled in your system's settings.
- Have you got duplicate printer-friendly pages?
These are completely unnecessary: you should just use a [print style sheet](#).
- Are you using comment pagination in WordPress?
You should just disable this feature (under settings » discussion) on 99% of sites.

- Are your parameters in a different order?
Tell your programmer to build a script to always put parameters in the same order (this is often referred to as a URL factory).
 - Are there tracking link issues?
In most cases, you can use hash tag-based campaign tracking instead of parameter-based campaign tracking.
 - Have you got WWW vs. non-WWW issues?
Pick one and stick with it by [redirecting the one to the other](#). You can also set a preference in Google Webmaster Tools, but you'll have to claim both versions of the domain name.
-

Key takeaways

- Duplicate content is **content that lives in several locations, i.e. URLs.**
- Duplicate content can **harm your rankings.**
- You can **find duplicate content on your site, or across the web by using Google, or by using duplicate content finder tools.**
- You can **solve duplicate content issues** by:
 - merging and redirecting duplicate content pages;
 - using rel=canonical;
 - linking back to the main URL;
 - writing original content.