SEO Secrets Revealed:
Why Duplicate Content Isn't Your Enemy

\\ What is duplicate content?

The term "duplicate content" refers to a piece of content that is an exact replica or almost identical to an existing piece of content, often with slight variations such as specific keywords like brand names, product names, or locations.

When analysing for duplicate content, there are two types to look out for - internal and external. Internal duplicate content happens when the same content appears on multiple pages of a website, creating competition among those pages. External duplicate content happens when the same content is on different websites within a group, causing competition between the sites.

Changing a few words won't always prevent a page from being seen as duplicate content. This can indirectly harm your search ranking and waste your crawl budget if incorrect pages show up in search results.

A common misconception is that Google will penalise your website for having duplicate content. This is not the case in most circumstances unless content has been stolen from another website and passed off as your own.

Google rewards unique and interesting content that is of high quality and solves users' queries. In terms of duplicate content, Google's stance is as follows:

Google

"Duplicate content on a site is not grounds for action on that site unless it appears that the intent of the duplicate content is to be deceptive and manipulate search engine results. If your site suffers from duplicate content issues, and you don't follow the advice listed above, we do a good job of choosing a version of the content to show in our search results."

However, duplicate pages can cause problems like indexing issues, crawlability, and overall performance issues.

\\ The effect duplicate content can have on your website

Duplicate content does not incur a direct penalty, except when copied from another website. However, it can still affect Google Search results in other ways.

The main duplicate content issues that you will come across include:

· The wrong version of pages showing in the SERPs; for example, you've duplicated the UK page for the US version of your site, but the UK version is ranking for US searches.

· Key pages aren't performing as you expect in SERPs or experiencing indexing issues.

· Confusing crawlers and wasting crawl budget on pages you don't want to index.

· Constant fluctuations or decreases in your core site metrics, such as traffic or rank positions.

· Sending search engines confusing prioritisation signals which could cause other unexpected actions.

There is no clear guide from Google as to what content they will prioritise and deprioritise; however, they advise webmasters and content creators to create pages primarily for users and not solely for search engines.

Creating high-quality content that your users want to read and that will help them will naturally increase the chances of your pages ranking better, rather than solely focusing on getting your pages to rank highly without considering the user.

It is important to regularly check your website for content duplication and minimise the risk by having a clear architecture, performing regular maintenance, and conducting technical audits to identify and fix any issues that arise as soon as possible.

\\ How can you prevent content duplication?

The best way to ensure content duplication isn't an issue is by having fresh writers create unique high-quality content for each of the pages on your website. However, this isn't always practical if you are running a large website or have an international presence. If this is the case, other methods and strategies can be utilised to prevent the creation of duplicate content on your website and to prevent other sites from copying your content. Some of these include:

Canonical Tags – When combating duplicate content, this could be the most crucial component to incorporate on your website or across various websites.

The rel=canonical element is a snippet of HTML code that provides Google with a clear signal that the publisher owns a piece of content, even when the content can be found elsewhere. These tags are used to tell Google which version of a page is the 'main version'. These tags can be used for mobile and desktop versions of a page, multiple location targeting pages, or print vs web versions of content.

There are two types of canonical tags: those that point to another page and those that are self-referencing, meaning they recognize themselves as the master version. Referencing canonicals is important in reducing duplicate content issues and is essential good practice for sites to follow.

Duplicate URLs – URL elements can cause duplication problems on a website due to how search engines perceive them. Therefore, if there are no other directives or instructions for the search engine to follow, a different URL will always be interpreted as a different page.

As a result of this, the lack of clarity or wrong signalling can cause fluctuations or decreases in core site metrics, such as traffic and rank positions, if not addressed. It is most common for duplicate versions of URLs to occur by having both HTTP and HTTPS versions of pages, www. and non-www., and pages that include trailing slashes and those that don't.

Concerning the use of www. versus non-www and trailing slash versus non-trailing slashes, it's crucial to determine the version used on your website and maintain this version across all pages to prevent potential duplication.

Furthermore, it's advisable to establish 301 redirects to guide users to the version of the page that should be indexed, thereby eliminating the possibility of duplication. For instance, example.com should redirect to www.example.com.

Finally, HTTP URLs pose a security threat whereas the HTTPS versions don't as they employ SSL encryption, thereby ensuring the page's security.

Meta Tagging – A more technical approach to dealing with duplicated content is using Meta robots and configuring them to send certain signals to search engines from your specified pages.

Meta robots are tags that can be used to exclude a certain page or pages from being indexed by Google if you would prefer not to have them show up in the search results.

You can simply add a 'no index' meta robot tag to the HTML code of the page which will effectively tell Google not to show this page in SERPs.

This is the preferred method as opposed to the Robots.txt method, due to the more granular blocking of a particular page or file, whereas a Robots.txt is often utilized on a much larger scale.

It is worth mentioning that meta robots can be used for many reasons, not just duplicate content; however, Google will understand the directive and should exclude the selected duplicate pages from SERPs.

Redirects – Using redirects can be a useful way of reducing duplicate content. Any pages that have been duplicated from another can be redirected back to the main version of the page.

This option is ideal if there are pages on your website with high volumes of traffic or link juice that are duplicated from another page, so that the page then benefits from it.

When using the redirect method for duplication, the main thing to remember is always to redirect to the higher-performing page to limit the impact on your site's performance, and where possible, use a 301 permanent redirect.

Old Page　　**301 Redirect** →　　**New Page**

Taxonomy – It is a good idea to start by looking at your site, mapping out the pages that have been crawled, and ensuring each has been assigned a unique H1 and focus keyword. Then organizing your content into topic clusters can help develop a thoughtful strategy for your website content going forward, minimizing the chances of pages being duplicated.

International SEO (Search Engine Optimization) Hreflang – This last point indirectly affects duplicate content and is only relevant for international websites. It is important to implement correct hreflang tags if you are trying to target your website internationally.

Hreflang tags are HTML code snippets that provide Google/search engines with a signal as to which location the page is intended for and where it should be shown in the SERPs.

For example, if your main site were in the UK and then you wanted to target the US, you would have specific hreflang implemented on your US pages to provide Google/search engines with guidance that the US page is intended for a US audience and to rank that page within the US.

If hreflang isn't implemented, this can often flag up these pages as duplicate content or you will see the wrong pages being served in SERPs for the various locations you are targeting.

The best way to ensure no issues with duplicate content or the wrong page being served is by creating unique content that is specific to each location using its local wording for each page created.

\\ Duplicate content – the key takeaways

The best way to avoid duplicates is by creating unique content for all your pages on your website; however, as we have discussed, this is not always practical or achievable. As such, utilizing a combination of techniques to stay on top of any duplication issues will put you in good stead to minimize any headaches duplicated pages give you and provide your website with the best chances of ranking the pages you want it to.

Molokini Marketing is a full-service marketing agency based in Worthing, West Sussex (UK).

Bringing ideas, insight, experience, energy and passion to every marketing project, our team of 12+ designers, writers, SEO specialists, coders, marketeers, thinkers and creators is big enough to deliver but small enough to care.
Contact us today for all of your industrial strength marketing needs.