

Resolving content duplication issues

Last Modified on 22/08/2023 11:50 am IST

Avoiding potential content duplication is crucial for e-commerce websites due to several important reasons:

1. **SEO Ranking:** Search engines penalize websites with duplicate content, as it can confuse their algorithms and result in lower rankings. This affects the visibility of your e-commerce site in search engine results pages (SERPs), reducing the chances of attracting organic traffic.
2. **Search Engine Trust:** Duplicate content can lead search engines to question the credibility of your website. This can result in a lower level of trust, which may further impact your rankings and the overall reputation of your e-commerce brand.
3. **User Experience:** Duplicate content can confuse visitors, leading to a poor user experience. Customers might land on different pages with the same content, making it difficult for them to find the specific information or products they are seeking. This frustration can drive potential customers away and reduce conversion rates.
4. **Wasted Crawling Budget:** Search engine bots have a limited crawl budget, meaning they can only crawl a certain number of pages on your site within a given timeframe. If duplicate content takes up a significant portion of this budget, important pages might not get crawled and indexed, impacting your overall SEO strategy.
5. **Canonicalization Issues:** Duplicate content can lead to canonicalization problems, where search engines struggle to identify the most relevant page to display in search results. This can result in the wrong page being ranked or displayed, leading to inaccurate user expectations.
6. **Backlink Dilution:** If different versions of your content are scattered across various URLs, backlinks and social shares might also get spread thin, diluting the overall authority and impact of those links.
7. **Algorithmic Penalties:** In some cases, excessive duplicate content might trigger algorithmic penalties from search engines, further diminishing your site's visibility in search results.
8. **Competitive Disadvantage:** E-commerce is a highly competitive field, and sites with unique, high-quality content tend to outperform those with duplicated or low-quality content. Avoiding content duplication gives you a competitive edge by providing users with valuable and differentiated information.

To ensure your e-commerce website's success, it's crucial to implement proper content management strategies, including using canonical tags, avoiding content syndication issues, and regularly monitoring your site for duplicate content. This will help improve SEO, enhance user experience, and boost your brand's credibility in the digital landscape.

Here are some real steps you can take on your webstore to avoid content duplication:

Types and Examples of Duplicate Content Pages

In this post, we will consider two types of pages with duplicate content:

- Any page that is queried using **Google Ads** parameters;
- Pages that include product listing such as department and category pages.

In the first case, a link can look like this:

```
http://www.yourstore.com/dept?gclid=CKTf7smRu7sCFcEnpQodSDYAVA
```

- This link is augmented by the **Google Click ID parameter** which originates from **Google Ads links**.

Another version of the same page can be viewed at this URL:

```
http://www.yourstore.com/dept
```

- In this situation, the search engine will not know which version should be included in search results as previously stated.
- We will call this type of page with duplicate content a **hard duplication**.
- In the second case, let us suppose without loss of generality that in a given department pages, the products are listed in several pages, with each assigned a number, in the following form:

```
http://www.yourstore.com/dept?page=1
```

- Another example of hard duplication occur with persistent-filtered search pages.
- Clearly, we need to tell to the search that those pages are in fact **related** to each other so that they can appropriated displayed in the search results.
- We will refer to this kind of duplication as **soft duplication**.

WebSell's Solution:

- The central part of our technique to mitigate the problem relies on including some appropriate HTML tags in the header of each page in your store with potential duplication issues.
- These are HTML tags are precomputed and stored in NitroScript variables.
- Include those NitroScript variables in your header template so that those tags will be included in the final page.

In filtered search pages, you can specify the **canonical** URL by including the following code snippet in your header template:

```
{if (pageproperty['pageid'] eq 'filtered')}
  {ifThereAre pfscanonical}
  {forEach pfscanonical}
    <link rel="canonical" href=""/>
    <meta property="og:url" content=""/>
  {endForEach}
  {endifThereAre}
{endif}
```

To specify the canonical URL for pages requested by Google Ads, you should include this code in your header template:

```
{if (pageproperty['crawled_parameters_canonical_url'])}  
  <link rel="canonical" href=""/>  
  <meta property="og:url" content=""/>  
{endif}
```

Finally, addressing soft duplication is done by including this code:

```
{forEach linkdata}  
  {if ((linkdata['rel'] ne 'canonical') || (pageproperty['crawled_parameters_canonical_url'] eq ''))}  
    <link rel="{linkdata['rel']}" href=""/>  
  {endif}  
{endForEach}
```

To see those fixes in actions, if, for instance, you open a listing page, you should notice in the head of the resulting HTML that these tags were added:

```
<link rel="canonical" href=""/>
```

If the page is potentially vulnerable to soft duplication issues, then you will notice even more tags in the form of:

```
<link rel="next" href=""/>
```

These links are dynamically computed and are therefore not the same for a given pair of pages.
