Google Issues Duplicate Content Warning by @martinibuster

Today John Mueller, Google Webmaster Trends Analyst, issued a warning via Twitter about duplicate content issues he's seeing. He also clarified what doesn't qualify as duplicate content.

Google's John Mueller issues a warning about duplicate content and clarifies what is not duplicate content.

Trailing Slash on Root/Hostname

Trailing slashes on root/hostnames don't matter. What that means is that it doesn't matter if there is a forward slash at the end of your domain name or not, they both equal the same thing. That means you can accidentally have the home page linked throughout your site as www.example.com/ and www.example.com and Google won't see it as a duplicate content issue.

This is a good clarification because it means you no longer have to worry about that (non) issue.

ADVERTISEMENT

Forward Slash at End of Files are Seen as Duplicate

This is important to understand. File names with and without a forward slash can be seen as duplicate. Thus if your web page can be reached by example.com/fish and example.com/fish/, you have a duplicate content issue. If the real URL is /fish/ then

your server should be redirecting /fish to /fish/.

Different Protocols DO Matter

Here is where the duplicate content issue becomes real. John Mueller points out that Google will see the same page as a two different pages if you write the same URL with a different protocol.

For example, https://www.example.com is going to be seen as different from http://www.example.com. As long as you have 301 redirects to handle that, then you're fine. But if you don't, then Google may see that as a problem. And it could be a problem.

How a Competitor Can Confuse Google

Some servers will still serve a web page as HTTPS, even if you don't have a security certificate. Google will see that as a duplicate web page. All it takes is for a competitor to begin linking to your site with https to get Google indexing a duplicate web page.

 This is a timeout error generated by following an HTTPS URL to a non-secure website.

And some servers may not serve a non-SSL web page at all via HTTPS if no redirects are in place to handle that request. Thus, if your non-SSL site doesn't have redirects in place to handle requests for an HTTPS version, and your server's response is to deliver a "site can't be reached" error, a competitor can create links to the non-existent HTTPS version, Google may see that as a separate page, according to Google's John Mueller.

 Illustration of what constitutes duplicate content posted by John Mueller on Twitter.

According to John Mueller's illustration:

“Different protocols & hostnames do matter…”

Then he uses this example:
https://www.example.com/ is not the same as https://www.example.com/

John Mueller then illustrated more examples of duplicate content:

This URL:
https://www.example.com/

Is not the same as this URL:
https://example.com/

And this:
https://example.com/fish

Is not the same as this:
https://example.com/fish/

All of the above examples represent ways a competitor can link to your site and create what Google will see as duplicate content. Will this duplicate content hurt your rankings? Probably not. Except for site is down errors, Google is usually pretty good at figuring out that two pages are the same then combining them. Still, it’s a good practice to not confuse the search bots.

How to Protect Yourself from Duplicate Content Issues?

1. Canonical Tag
Define a canonical page for each page. This tells Google which version of your URL is the correct one. While Google is not obligated to obey the canonical, it will take it as a confirmation and a hint of which URL is the correct one to show in the search results pages.

https://support.google.com/webmasters/answer/139066?hl=en

2. Test how your server responds to requests for secure and insecure URLs
You may need to add 301 redirects to compensate for duplicate URL or site is down errors.

3. Audit your URLs
Crawl your page with Screaming Frog (paid) or XENU Link Sleuth (free) and review your URLs for any duplicates or page not found errors.

4. Investigate 404 Errors
Check your server logs, traffic analytics or Google Search Console and track down the sources of any 404 page not found erorrs. 404 errors should always be investigated.

John Mueller's clarification about what is and what is not a duplicate page to Google is important. It's good to have official clarification. And although you may find issues, I'm fairly confident that in most cases Google will figure out which page is the right page, so there isn't a need to panic. However, SEO is literally about thousands of little details and this is just one of them.

Image Credits
Featured Image: Anastasia_B/Shutterstock.com
Altered beyond recognition by Roger Montti

http://tracking.feedpress.it/link/13962/7777546