

Duplicate Content & Canonical URL Algorithm Update October 2009

Google have made it very clear they are penalizing websites with duplicate content. SearchForecast has detailed various ways for dealing with duplicate content on your website.

Recognize duplicate content on your website.

The first and most important step is to recognize duplicate content on your website. A simple way to do this is to take a unique text snippet from a page and to search for it, limiting the results to pages from your own website by using a Site:URL query via Google. Multiple results for the same content show duplication you can investigate. For instance, entering “coffee site:www.peets.com” into Google will show all the pages that coffee appears on

Determine Preferred URLs.

Before fixing duplicate content issues, it is important to determine your preferred URL structure. So you must choose ONE URL structure NOT have multiple URLs. For instance, all URLs should be consistent like www.yourcompany.com not company.com or www.company.com/html, or www.company.com/php

Be consistent within your website.

Once you've chosen your preferred URLs, make sure to use them in all possible locations within your website

Apply 301 Permanent Redirects .

If you can, redirect duplicate URLs to your preferred URLs using a 301 response code. This helps users and search engines find your preferred URLs should they visit the duplicate URLs. If your site is available on several domain names, pick one and use the 301 redirect appropriately from the others, making sure to forward to the right specific page, not just the root of the domain.

Implement [the rel="canonical" link element](#).

Where 301 redirects are not possible, the rel="canonical" link element can give Google a better understanding of the website preferred URLs. The use of this link element is also supported by major search engines including Yahoo!, Bing and ASK.

Use the URL parameter handling tool in Google Webmaster Tools where possible

If some or all of the website's duplicate content comes from URLs with query parameters, this tool can help you to notify us of important and irrelevant parameters within your URLs.

Don't Block Access with the robots.txt file

Google now recommends not blocking access to duplicate content with a robots.txt file. Instead, use the rel="canonical" link element, the URL parameter handling tool, or 301 redirects. If access to duplicate content is entirely blocked, search engines effectively have to treat those URLs as separate, unique pages since they cannot know that they're actually just different URLs for the same content.

A better solution is to allow them to be crawled, but clearly mark them as duplicate using one of our recommended methods. If you allow Google to crawl these URLs, Googlebot will learn rules to identify duplicates just by looking at the URL and should largely avoid unnecessary recrawls in any case.

In cases where duplicate content still leads to us crawling too much of your website, the crawl rate setting in Webmaster Tools can be adjusted.

For further information, contact:

Richard Andrus, VP Client Services
San Francisco, USA
Email: richard.andrus@searchforecast.com
Telephone: 925.858.5663

Tania Johnston
Regional Manager
Sydney, Australia
Telephone: +61 0407 078 537
Email: tania.johnston@searchforecast.com

Disclaimer

This report is provided for your general assistance and information only. SearchForecast does not make any representations or warranties (express or implied) as to the accuracy or currency of the information contained in the materials nor does it accept liability whatsoever for any reliance placed on this material, including any liability in negligence for reliance on any information in these materials or any products, services or information which may be provided by the companies and organizations referred to.