eMarketing
UNIVERSITY

Get full access to our training courses in
Internet Marketing
www.seo-training-course.com

DUPLICATE CONTENT ISSUES

We have already stated in our Training Course that both Internet users and search engines want fresh, unique and quality content. Nevertheless, online business practice often has to contend with repetitive Web page content, which has been created either accidentally or deliberately. That's why duplicate content has become a huge topic of discussion recently. The harm caused by duplicated content has become increasingly apparent thanks to the new filters search engines have implemented.

WHAT IS DUPLICATE CONTENT?

Duplicate content generally refers to multiple versions of the same content that exist on different pages, either within one domain or across different domains. These blocks of similar information might either completely match each other or appear similar.

Many webmasters and SEO/SEM experts often speculate about the percentage of similar pages, and try to predict the figures which may lead to penalties for carrying duplicated content. However, it is very difficult to distinguish between the percentage of pages that are absolute duplicates and that which would trigger a duplicate content filter and those pages that are slightly similar. In fact, the detection of duplicate content involves more than simple, direct comparison: when comparing two similar pages, search engines consider other factors such as site authority, link popularity, domain age and others.

TYPES OF DUPLICATE CONTENT

In its Webmasters/Site Owners Help, Google identifies the following types of **non-malicious** duplicate content:
- Discussion forums that can generate both regular and stripped-down pages targeted at mobile devices
- Store items shown or linked via multiple distinct URLs
- Printer-only versions of Web pages.

Besides, Google is already able to evaluate navigation panels, common header text, ads, footer text, and repetitive page links. These instances of duplicate content are not penalized but are ignored.

Other types of content deliberately duplicated across domains and created to manipulate search engine rankings are considered **malicious**. These may include similar landing pages created to attract more visitors to your site, subdomains, or domains with substantially duplicated content and pages with stolen content.

eMarketing University

Get full access to our training courses in
Internet Marketing
www.seo-training-course.com

In most cases, you are very unlikely to run the risk of being penalized for duplicate content if you do not create it deliberately. However, you should be armed with knowledge to ensure you do not use malicious duplicate content and accidentally trigger a search engine's filter.

HOW SEARCH ENGINES TREAT DUPLICATE CONTENT

Most webmasters have already learned that search engines do not like duplicate content. The problem is that multiple pages with the same content confuse SEs, which aim to list the most relevant, unique and original results, not clutter. Thus, in an effort to provide more varied results to their users, search engines filter websites that appear too similar to each other: except for the most relevant results; similar results are excluded.

As stated in Google's Webmasters/Site Owners Help, "Google tries hard to index and show pages with distinct information. This filtering means, for instance, that if your site has a "regular" and "printer" version of each article, and neither of these is blocked in robots.txt or with a no-index meta tag, we'll choose one of them to list."

In other words, **duplicate content filters** are algorithms designed to compare one page against another. If the filter considers two or more pages to be substantially similar, it simply keeps the most trusted one in the primary index, while moving the others to the supplemental index.

Penalties may arise when you start copying hundreds or thousands of pages from other domains or create exact replicas of existing sites. Moreover, you run a risk of being penalized in case the ratio of Unique Content vs. Borrowed Content is too low on your site.

SEARCH ENGINES' INITIATIVE WITH THE "CANONICAL TAG"

As you have just learned, search engines have a difficult task excluding duplicate results from their indexes. Duplicate content appears here and there as the result of various incidences of article publishing, blog posts or different URLs of the same site leading to the same content. Ecommerce sites often need to have several pages listing the same set of products.

The announcement from Yahoo!, Live & Google that they will be supporting a new "canonical url tag" to help webmasters and site owners eliminate self-created duplicate content in the index was the biggest change to SEO best practices since the emergence of Sitemaps.

The tag is part of the HTML header on a web page; the same section you'd find the Title attribute and Meta Description tag. In fact, this tag isn't new, but like nofollow, simply uses a new rel parameter. For example:

```
<link rel="canonical" href="http://www.example.com/product.php?item=product-name"/>
```

eMarketing UNIVERSITY

Get full access to our training courses in
Internet Marketing
www.seo-training-course.com

This tag is used in the head section of the page in order to give search engines the canonicalization suggestions. Thus a specified "canonical" page http://www.example.com/product.php?item=product-name becomes the preferred version of a set of pages with highly similar content.

Canonical tags are useful in cases where multiple URLs are pointing at the same page, but might also be used when multiple versions of a page exist. This tag will operate in a similar way to a 301 redirect for all URLs that display the page with this tag. You can use relative or absolute links, but absolute links are recommended by search engines.

Here are some more Google Webmasters/Site owners Help recommendations:

"To migrate to a completely different domain, permanent (301) redirects are more appropriate. But sometimes it's not easily possible to set up redirects. In a situation like this, **you can use the rel='canonical' link element across domains** to specify the exact URL of whichever domain is preferred for indexing."

Search engines do hope this tag will help to regulate and facilitate the duplicate content question. If this tag can't be implemented they'll keep using algorithms designed to compare one page against another to determine the canonical.

HOW TO AVOID DUPLICATE CONTENT

1. Do not steal content from other sites.
First, stolen content adds no value to your site. Besides, this deceitful practice may harm your search engine rankings and result in your site being banned for malicious duplicate content and copyright infringement. To avoid these problems, be sure to create unique and original content. If possible, do your best to follow your site's overall theme – this will help you become an authoritative Web resource within your niche.

2. Check if you accidentally use duplicate content on your site.
When the site grows large, things may get out of control and you may find yourself repeating things you've already said. To avoid this problem, check the pages across your site for duplicate content. If you find any, alter the pages to make each of them as unique as possible.

3. Check if anybody is scraping your content.
There are a number of free duplicate content checkers available online. For example, you can use the Copyscape tool to search for copies of your page on the Internet.

If you discover someone is stealing content from you, you can report abuse to the search engines: file a DMCA infringement request with Google, Yahoo!, and MSN. The other step is to

file a legal suit (or threaten such) against the website in question, as it can take months before your DMCA motions go into effect.

4. If you use articles on your site, check if they have been distributed over the Internet.
Again, use a duplicate content checker and check if your articles are posted anywhere except your site. If you do find copies of your article on the Web, take steps to avoid duplicate content issues. First, make changes to each reposted article to make them unique. Also, ensure that your articles are relevant to your overall site theme, as large websites that aggregate a lot of content on the same topic may outrank your site in terms of your keywords and key phrases.

5. If you publish the same post on the blog syndication website and on your blog, include a link back to your own site.
This will do if your own site has already gained good link popularity, and you run no risk of being outranked by the blog syndication website.

If you feel your site is smaller and "weaker" in terms of SE positioning, make your article look different in two different locations. To accomplish this, you can rewrite the article, or use blog widgets to make the overall page look and sound different. For example, WordPress, Diachronics, and many other blog platforms allow you to add various widgets that further alter the page content that can be indexed. Each page of your personal blog may contain more customizable information such as menu systems and categories, along with static text; again, alter the overall indexable content of the site.

6. If you run several websites with identical content and promote them through different channels (e.g., one with the help of PPC campaigns and another - through organic SE rankings) **be sure to tell Google which domain you prefer to be indexed.** To do so, go to Google Webmaster Tools and include the preferred version of your URLs in your sitemap file.

7. If you have multiple URLs on the same domain that point to the same content, consider blocking duplicate pages from indexing. For example, if you don't want search engines to index the printer versions of your site's articles, disallow those directories or make use of regular expressions in your robots.txt file.

8. If you participate in an affiliate program, add value to your affiliate site. Rewrite product descriptions instead of using ready-made templates offered by the original merchant.

Also, create additional pieces of unique and relevant content (tips, news, how-to articles, etc.) to make your site more useful and attractive to users and search engines.

9. If you target multiple countries, consider creating different content for each domain.

eMarketing University

Get full access to our training courses in
Internet Marketing
www.seo-training-course.com

If your site is translated into different languages and you run several domains for each site copy, in most cases search engines will consider these domains as sites with different content. However if you run several domains with identical content in English to market in the US, Canada, Great Britain, and Australia, do your best to alter content on each site to make these four sites feature different content.

SUMMARY

1. Create unique and original content, instead of copying content from other sites.
2. Use duplicate content checkers to see if anyone scrapes content from your site.
3. If your article or blog post has been distributed over the Internet, take steps to prevent duplicate content issues.
4. Eliminate duplicate pages on your site, or at least block them from being indexed by the search engines.
5. If you participate in an affiliate program, make sure your site content is unique.

START FROM CONTENT

The first and the most perfect way is to write unique and really valuable content. You'll be gratified when some webmaster links to you just because you provide useful or interesting information. A high-quality and regularly updated site will naturally be linked to.

OLD BUT EFFECTIVE METHODS

LINKS FROM DIRECTORIES

SUMMARY

1. Inbound non-reciprocal links can raise your rank noticeably. Major tactics are as follows: display unique valuable content on your site that attracts others in your

Get full access to our training courses in
Internet Marketing
www.seo-training-course.com

industry to link to you unilaterally, write and submit articles and press releases, get links from related sites, create viral link-worthy content and distribute it on the Web using social media and other channels for spreading the word.

2. Although link exchange is considered an outdated method of raising your link popularity, you can use it in combination with other link building methods. Note however that you should avoid link exchange with link farms.

3. Creating video files, tools and widgets that people want to link their websites to is a powerful and effective method to get links and bring more traffic to your site.

4. Don't participate in manipulative link schemes that are prohibited by the search engines' guidelines. Inbound links or cross change links with a site about a topic other than yours can be seen by SEs as a Black hat technique and may be penalized at any time.