

# Drupal SEO pitfalls & how to avoid them

Tien technical seo tips for drupal

E-book by **Wouter De Bruycker & Brent Gees**



## SEO optimization for business impact

Advanced SEO techniques for businesses involve much more than [standard SEO optimizations](#) and high-quality content. Digital experiences that captivate your website audience depend greatly on CMS-specific optimizations.

When it comes to Drupal, SEO optimization seems simple enough at first sight. A majority of the most popular Drupal modules have been designed with SEO in mind and, as a result, they are relatively easy to optimize. However, Drupal does come with its own particular and sometimes less well-known features that can have a significant impact on your website's Google search rankings.

In this e-book, you will discover ten technical, Drupal-specific SEO tips. We share the most common SEO pitfalls in Drupal and explain their potential impact on your business. We also share our tips and solutions to help you counter these problems.

# Wouter De Bruycker

Wouter De Bruycker is an experienced SEO specialist and works as a senior digital strategist at Dropsolid. His professional background as a web developer helps him to assess the impact of development choices on SEO results. He is a frequent speaker at SEO conferences in Belgium and abroad.



# Brent Gees

Brent Gees is a Drupal architect, back-end developer and Drupal trainer with Dropsolid. He is well-recognised in the Drupal development community and is a regular speaker at Drupal community events.

## 10 Drupal SEO pitfalls

- 1 Too many redirects
- 2 Indexable internal search
- 3 Indexed test environments and test pages
- 4 Public entities
- 5 All pages should be an entity
- 6 Assets blocked by robots.txt
- 7 Module overload
- 8 Security issues
- 9 Crawling vs. indexing
- 10 Incorrect analytics data



Tip #1

Too many redirects





# Too many redirects

## Problem

Redirects are more of a curse than a blessing - at least where SEO is concerned. A build-up of otherwise harmless consecutive redirects can have a major negative effect on your website loading times and your SEO score.

The screenshot shows the 'Link Redirect Trace' tool interface. It displays a sequence of three redirects for the URL <http://aquadomo.be/>. The tool is branded with 'LinkResearchTools' and includes a 'Help' link.

Step	Status Code	Redirect To	Time (ms)	Percentage	Backlinks	LRT Power	LRT Trust	Page
1	301	<a href="http://www.aquadomo.be/">http://www.aquadomo.be/</a>	4264	52%	4	2	1	2
2	301	<a href="https://www.aquadomo.be/">https://www.aquadomo.be/</a>	2579	31%	2K	2	1	2
3	200	<a href="https://www.aquadomo.be/">https://www.aquadomo.be/</a>	1324	16%	85	3	2	6

If Googlebot (Google's web crawler) visits a web page that returns a redirect status code, it will automatically add this page to the bottom of its 'to visit' list for that particular website. This isn't much of an issue for smaller websites, but large environments will soon feel the hit of much slower indexing. The conclusion is straightforward: redirect chains are not search engine-friendly.

## Solution

Avoid going down the easy route of uncommenting the most obvious options in .htpasswd. Instead, think about the logical flows you want to create with your URLs and adjust your code accordingly to minimize redirects.

```
# Redirect to https://www if https = off, www = off
RewriteCond %{ENV:HTTPS} !on
RewriteCond %{HTTP_HOST} !^www\. [NC]
RewriteRule ^(.*)$ https://www.%{HTTP_HOST}%{REQUEST_URI} [L,R=301]

# Redirect to https://www. if https = off, www = on
RewriteCond %{ENV:HTTPS} !on
RewriteCond %{HTTP_HOST} ^www\.(.+$) [NC]
RewriteRule ^(.*)$ https://www.%{HTTP_HOST}%{REQUEST_URI} [L,R=301]

# Redirect to https://www. if https = on, www = off
RewriteCond %{ENV:HTTPS} on
RewriteCond %{HTTP_HOST} !^www\. [NC]
RewriteRule ^(.*)$ https://www.%{HTTP_HOST}%{REQUEST_URI} [L,R=301]

# the remaining case is already https = on, www = on.
```

*Example of a clean redirect set-up, without any negative SEO impact.*



Tip #2



Indexable internal  
search







# Indexable internal search

## Problem

By default, internal search result pages of a website are often indexable by search engines, unless you explicitly rule out this option in your CMS settings. This implies that Google and other search engines can index your web pages that show nothing but a list of search results and, subsequently, display these pages in their search engines. In theory, this shouldn't be much of an issue. In practice, however, pages with aggregate search results are considered by search engines to be pages with inferior content, resulting in a near-certain hit to your SEO score.

## Solution

In Drupal, you can anticipate this SEO problem by installing the [metatag](#) and [metatags\\_view](#) module and submodule, respectively. They offer you the possibility to change indexation settings for your search result pages. This hides your search result pages from the view of Googlebot and other web crawlers.

<b>PAGE SETTINGS</b> Path: /search Menu: No menu Access: Unrestricted	<b>▼ ADVANCED</b> Meta tags that might not be needed by many sites. <b>Robots</b> <input type="checkbox"/> Allow search engines to index this page (assumed). <input type="checkbox"/> Allow search engines to follow links on this page (assumed). <input checked="" type="checkbox"/> Prevents search engines from indexing this page. <input checked="" type="checkbox"/> Prevents search engines from following links on this page. <input type="checkbox"/> Prevents cached copies of this page from appearing in search results. <input type="checkbox"/> Prevents descriptions from appearing in search results, and prevents page caching. <input type="checkbox"/> Blocks the <a href="#">Open Directory Project</a> description from appearing in search results. <input type="checkbox"/> Prevents Yahoo! from listing this page in the <a href="#">Yahoo! Directory</a> . <input type="checkbox"/> Prevent search engines from indexing images on this page. <input type="checkbox"/> Prevent search engines from offering to translate this page in search results. Provides search engines with specific directions for what to do when this page is indexed.
<b>HEADER</b>	
<b>FOOTER</b>	
<b>NO RESULTS BEHAVIOR</b>	
<b>PAGER</b>	



When using a node for your search pages, install the metatag module and add the field to your content type.

The screenshot shows a Drupal content editor interface. On the left, there is a 'Title' field with the value 'Search' and a 'Body' field with a rich text editor containing placeholder text. Below the body field is a 'Text format' dropdown set to 'Basic HTML'. At the bottom left, there are 'Save', 'Preview', and 'Delete' buttons. On the right, a 'Published' sidebar is visible, showing 'Last saved: 10/22/2019 - 19:06', 'Author: brentg', and a 'Revision log message' field. A red box highlights the 'METATAGS' option in the sidebar's menu.

## ▼ ADVANCED

Meta tags that might not be needed by many sites.

### Robots

- Allow search engines to index this page (assumed).
- Allow search engines to follow links on this page (assumed).
- Prevents search engines from indexing this page.
- Prevents search engines from following links on this page.
- Prevents cached copies of this page from appearing in search results.



Tip #3

Indexed test  
environments and test  
pages





# Indexed test environments and test pages

## Problem

Development and staging environments are often crawlable and indexable by search engines. Temporary content is also prone to accidental indexing by crawlers.

This usually happens as a result of configuration issues, minor mistakes or simply time pressure on the development team. Fortunately, this issue is as easy to resolve as it is to overlook.

Obviously, development and staging environments are better off without indexing, to avoid the risk of being detected by the general public or the competition.

The screenshot shows a search engine results page for the query 'inurl:dev site:glue.be'. The search results include:

- Main Content Only - Lievegem Frisbee Team**  
lift-it-up.org › Staging › index.php › features › main-content-only  
Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when ...
- March 2018 posts - Safety and health at work - EU-OSHA**  
https://test.osha.europa.eu › blog  
Mar 8, 2018 - EU OSHA information about occupational health and safety across EU countries. Find surveys, regulations and standards, risk assessment ...
- Testing Post - Matti - STAGING SITE | Kaverina olen cool.**  
https://staging.maeniusaa.fi › testing-post-matti  
This is a sample text not using the Oxygen Builder. Lorem Ipsum is a dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's ...
- DevBeta**  
https://dev.beta.globo.com › Translate this page  
Único cometa de curto período que é regularmente visível a olho-nu da Terra e a aparecer nos céus durante uma só geração. Há 1 ano Universo. Tudo Sobre o ...
- cliik hier - S.M.A.K. - Glue**  
dev.smak.be.preview.glue.be › tentoonstelling - Translate this page  
No information is available for this page. Learn why
- Wachtwoord vergeten | Woody Official**  
https://dev-shop-woodyworld.preview.glue.be › account - Translate this page  
Woody pyjamas en homewear van 0m tot XXL voor volwassenen in jersey flanel terry polar interlock.
- Shop ons ruim aanbod badmode accessoires - Vijgeblad**  
dev-vijgeblad-be.preview.glue.be › ... › Badmode - Translate this page  
Op zoek naar stijlvolle accessoires voor aan het strand? Van jurkjes tot jumpsuits. Koop je badmode veilig online bij Vijgeblad.
- Maakt meten een school beter?**  
dev.onderwijskwaliteit.be.preview.glue.be › files › files - Translate this page  
Vakken. 3ASO 4ASO 5ASO 3TSO 4TSO 5TSO 3BSO 4BSO 5BSO TOTAAL. Aardrijkskunde. 2. 1. 3. Bedrijfseconomie. 1. 1. 2. Biologie. 3. 1. 4. Boekhouden. 1. 1.
- Emotional Supply Chains at Zabłudowicz Collection, London ...**  
dev.artlead.net.preview.glue.be › content › journal › emotional-supply-cha...  
Jul 13, 2016 - The Zabłudowicz Collection is pleased to present Emotional Supply Chains, an exhibition addressing the construction of identity in the digital ...

Example of publicly accessible test environments

## The solution

Test pages in live environments can be unpublished or removed from indexing using the [metatag module](#).

▼ **ADVANCED**

Meta tags that might not be needed by many sites.

**Robots**

- Allow search engines to index this page (assumed).
- Allow search engines to follow links on this page (assumed).
- Prevents search engines from indexing this page.
- Prevents search engines from following links on this page.

Test environments can be password protected using the `.htpasswd` file.

Sign in

**Warning**  
Your connection to this site is not private

Username

Password



Tip #4



Public entities



# Public entities

## Problem

By default, entities are publicly available through their own unique URL. For example, a team member module might generate a public node for each team member, whilst these are only used for the team overview page and the website might not need a subpage for every team member. In Drupal, the URLs for these entities look as follows:

*<https://mysite.com/node/42>, <https://mysite.com/taxonomy/term/42>, etc.*

This results in low-value pages that are fairly thin on content - not an ideal situation for indexation by Google and other search engines. Redundant pages are a waste of resources, including bandwidth, crawl budget and database storage.

## Solution

You can prevent irrelevant entities from being accessed by your website visitors. In Drupal, the [rabbit\\_hole](#) module does an excellent job at this.



## Tip #5

All pages should be an  
entity





# All pages should be an entity

## Problem

Certain types of web pages are usually generated based on other content. These pages are not an 'editable node' in the back-end of a website. Home pages and overview pages are typical examples of this. It can be difficult to make SEO-related changes to web pages that aren't editable nodes. Examples of items that are difficult to edit are meta tags and XML sitemap configuration for page inclusion.

## Solution

The solution is twofold:

- Use the Drupal layout builder to create a home page.



- Use Paragraphs in combination with the [block\\_field](#) or [overview\\_field](#) module to add blocks to the page.

**Content \*** **Header**

**Language**  
English ▾

**Title \***  
News and events

This field is used as the title for this page in your browser tab, in overviews and for SEO and sharing purposes. It is also shown on the detail page for this page. You can change what is used for SEO and sharing purposes in the metatag tab on the right hand side of this form. You can also change what is shown on the detail page by setting a different title below (this does not work when the node is set as the homepage)

Dark background (all languages)  
When checked, the page will have a grey background instead of white.

[Show row weights](#)

---

**PARAGRAPHS \***

Overview Edit ▾

**Paragraph type**  
Image ▾

[Add another Paragraph](#) to Paragraphs



## Tip #6

Assets blocked by  
robots.txt



# Assets blocked by robots.txt

## Problem

Certain website assets, including favicons and images, may be located inside a folder that is hidden from view of web crawlers through the *robots.txt* file. This way, search engines are unable to view and analyse web pages in the way a human website visitor would.

```
# Directories
Disallow: /includes/
Disallow: /misc/
Disallow: /modules/
Disallow: /profiles/
Disallow: /scripts/
Disallow: /themes/
```

## Page partially loaded

Not all page resources could be loaded. This can affect how Google sees and understands your page. Fix availability problems for any resources that can affect how Google understands your page.

*Left, the robots.txt that prevents certain assets from view. Right, the resulting error notification in Google's Search Console.*

## Solution

Make sure Google has access to the full web page that you would like to be indexed. This can be done by adding all relevant website assets (including images, icons and favicons) in a publicly accessible folder that isn't blocked by your robots.txt file. In addition, you should keep an eye on Google Search Console for any notifications on blocked resources.



**Tip #7**

**Module overload**




# Module overload

## Problem

**Modules** are a blessing to work with and they solve a variety of problems. Nevertheless, there are limits to any module-reliant CMS, including Drupal. An excess number of modules can slow down your website significantly by downloading multiple resources to your visitors' web browsers. A sluggish website is detrimental to your SEO score and has a negative impact on your visitors' user experience.

## Solution

Prevention is better than cure. Avoid modules that are too complex or not strictly necessary on your environment. If you are using a module as a shortcut or quick fix, looking into alternative solutions might be a valuable alternative.



In addition, you should check your website for modules that are no longer used on your website. You can verify this using the [unused\\_modules](#) tool.



Tip #8

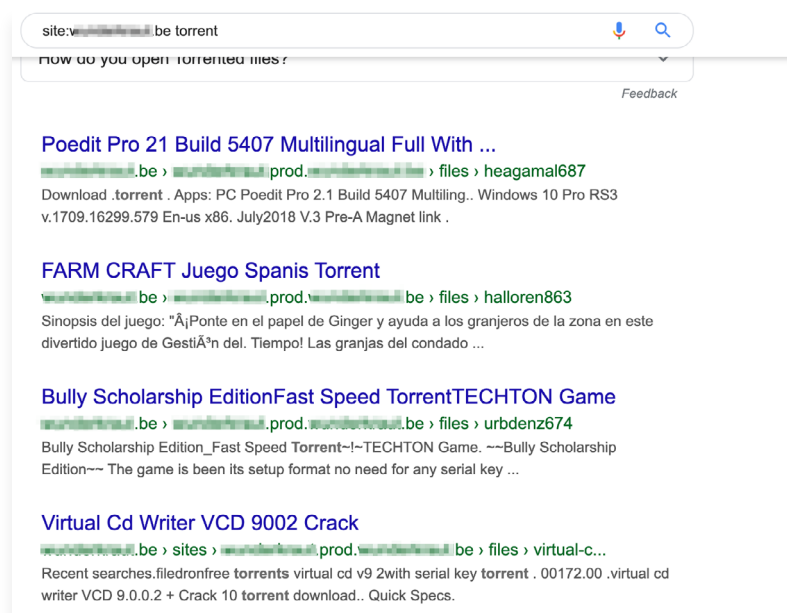
Security issues



# Security issues

## Problem

Security issues can have a profound impact on your website's SEO score. A classic example is the public file upload issue. Allowing public file uploads can result in lower organic traffic when set up incorrectly. If files can be uploaded to your website without any form of verification or CAPTCHA, this could result in thousands of files being uploaded by spammers. If these files are subsequently indexed by Google, your SERP ranking will decrease.



*Example of a legitimate website containing links to illegal torrent files. This is undoubtedly the work of spammers, who are always on the lookout to exploit security vulnerabilities.*

If the spam on your website catches Google's attention, you might become subject to a so-called manual action by a Google employee. The website page in question (or even your entire website) could get downgraded in its search engine's result rankings. This usually leads to a decrease in organic search traffic, with serious consequences for your website and your business.



## Solution

Use a CAPTCHA module to exclude spam from your public file uploads. [Recaptcha](#) and [simple\\_recaptcha](#) are amongst the most popular options for Drupal. You could also consider placing private files outside of Drupal's web root folder.



## Tip #9

Crawling vs. indexing



# Crawling vs. indexing

## Problem

Contrary to popular belief, blocking a page (or folder) using the robots.txt file or through adding a noindex directive are two distinctly different things.

In short, robots.txt instructions impact crawling, whilst noindex directives (using a meta tag) impact indexing. This implies that Google and other search engines can occasionally stumble upon a link on a particular website that refers to a page that is blocked by robots.txt, and still index it. The result will most likely show up in search results as a snippet that doesn't contain any title or description, since Google will not be able to read it. Check the real-life example below:

The code comments show that “strange homepage URLs” were the reason behind the Drupal.org dev team decision to block Google from crawling that particular page. This proved ineffective, however.

## Solution

The rule is simple:

- **Robots.txt** instructions set rules for **crawling**.
- **Noindex** instructions set rules for **indexing**.

If you use the robots.txt and noindex instructions correctly, avoiding indexing issues should come easy.



## Tip #10

Incorrect analytics data

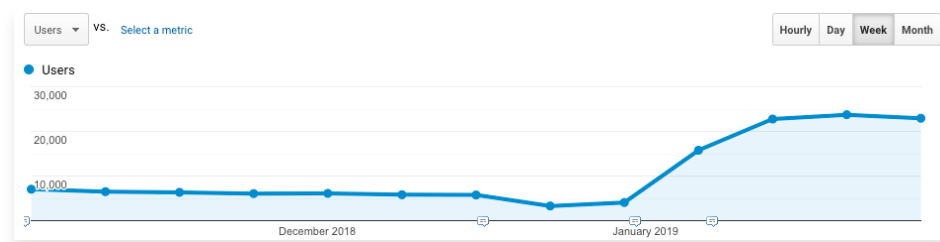


# Incorrect analytics data

## Problem

Uncontaminated data are crucial for all SEO-related decision-making. Sudden traffic spikes and drops in your Google Analytics dashboard might indicate technical issues. Always be ready to investigate this further before you start making important business decisions based on flawed data and incorrect assumptions.

An example:



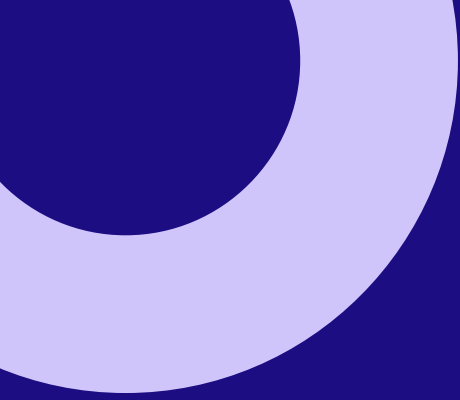
*An unexpected page view traffic spike, not triggered by any marketing effort.*

Further analysis of the case quickly revealed that something was amiss. The number of unique visitors had spiked, but all site visits had resulted from direct traffic. This was a clear warning sign: in general, traffic spikes are the result of search queries or external referral traffic.

It quickly transpired that the suspicious traffic was caused by a cookie compliance module update. After the update, each pageview would start a new session as long as website visitors didn't click the 'accept' button in the cookie pop-up notification.

## Solution

A simple configuration change in the Google Analytics settings proved sufficient to correct the error that had been caused by this otherwise harmless module update. In general, it is good practice to check all analytics settings on a regular basis. This will allow you to detect problems at an early stage and keep your data clean.



# Conclusion



# Conclusion

## Competitive SEO strategy

Using the above strategies, you can take on a majority of Drupal's SEO pitfalls. Anticipation is the key: in-depth analysis and smart configuration of settings can avoid the bulk of technical SEO-related problems further down the road. You should also bear in mind that every business case and website is unique, which implies that your SEO strategy should be bespoke as well. [A comprehensive SEO audit](#) will be able to kickstart your website optimization track. This way, you can generate added value and competitive advantage for your business.

Our Dropsolid SEO experts are available to help your business grow.

[Get in touch.](#)

### Useful links

- [SEO webinar](#)
- [SEO services by Dropsolid](#)