Search Engine and Performance Optimization of Drupal-based Websites

Jie Song, Hui Du and Qi Huang
Department of Information Management, School of Economics and Management, Beijing Jiaotong University, No.3 Shangyuan Village, Haidian District, Beijing, China
Email: 592167898@qq.com

Abstract. In view of incomplete and in-depth of research on search engine and performance optimization of Drupal-based websites, using the method of documentation, the common solutions are summarized. Their implementation methods are described and put into practice. Reference and guidance are provided for the search engine and performance optimization of Drupal-based websites.

1. Introduction
Drupal is an open source content management system based on PHP. Its modular design idea can help to develop websites efficiently. Strong global community guarantees the stable updating and growth of modules. Therefore, the number of websites based on Drupal is increasing.

Search Engine Optimization (SEO) can make the website friendly to search engines, and then get a higher ranking weight on search results. Performance optimization can not only enhance the popularity and influence of the website, but also improve the user's experience. Therefore, SEO and performance optimization are also key issues to be solved for Drupal-based websites.

In the SEO based on Drupal websites, Wang Y recommends using SEO Checklist, Pathauto, Meta tags and Apache Solr modules. In terms of performance optimization, Wang Y suggested that small and medium-sized website servers use the Boost module to statically process pages accessed by anonymous users. Zuo Q said that the performance of anonymous user access is much higher than that of registered users, the bottleneck of performance is not the database or memory, but the execution of PHP code.

2. Summary of SEO and Performance Optimization Solutions

2.1. Summary of SEO Solutions
Table 1 summarizes three common Drupal-based website SEO solutions.
Table 1. Drupal-based website SEO solutions

| Solutions                              | Descriptions                                      | Limitations                                      |
|----------------------------------------|---------------------------------------------------|--------------------------------------------------|
| Using the SEO Checklist module         | Provides a set of website SEO sub-solutions       | Some of the sub-solutions of SEO have no obvious effect. |
| Use the Real-time SEO for Drupal module| Organize website content quickly around keywords and monitor content that helps SEO in real time | It is difficult to configure. It is difficult to maintain. |
| Using the Drupal SEO Essentials module | Recommends a variety of SEO modules               | Officially not recognized; Unstable operation    |

Due to the stable operation, easy configuration and maintenance, SEO Checklist has become the most widely used solution. Table 2 summarizes the modules recommended for the partial sub-solutions of SEO Checklist with version number "7.x-4.1".

Table 2. Modules recommended for use in the SEO Checklist sub-solutions

| Modules                             | Descriptions                                                                 |
|-------------------------------------|-----------------------------------------------------------------------------|
| Meta tags                           | Improve the ranking and display of sites in search engine results by controlling meta tags. |
| Pathauto                            | Set a more reasonable URL path for the website; Avoid the Chinese characters in the URL becoming scrambled code. |
| XML sitemap                         | Help search engines crawl websites more intelligently and ensure real-time updates of results |
| Administration menu                 | Display all management links to provide direct access to any page.          |
| Meta tags quick                     | Adding meta tags to non-node entities.                                        |
| Drush                               | Used as Drupal command line shell and Unix scripting interface.              |
| Global Redirect                     | Standardize the Drupal website path.                                         |
| Redirect                            | Implementing path redirection and related control activities.                |
| Google Analytics                    | Selectively track/exclude users, roles, and pages; Monitor files downloaded by users from websites. |
| Context Keywords                    | Provide context for finding search keywords on a website.                    |
| Microdata                           | Allows users to use inline metadata to other sites and services to share content. |
| HTML Purifier                       | Help websites remove malicious code and ensure documents are compliant.      |
| SEO Compliance Checker              | Check the content of the node created or modified by the user in the search engine and feedback its compliance with the SEO rules. |
| Site verification                   | Make website verification for search engines and other services as simple as possible by adding meta tags or uploading specific files. |
| Site map                            | Indicate the orientation and connection of information resources to facilitate search engines to grasp website information. |
| Module Filter                       | Change the layout of the module list page and sort the modules alphabetically. |
| Security Review                     | Automatically check whether there is a security risk in the website.         |
| Scheduler                           | Enabling content editors to schedule nodes to be published and unpublished at a specified time in the future. |
| Read More Link                      | Transfer the "read more" link to the end of the summary.                     |

2.2. Summary of Performance Optimization Solutions
Table 3 summarizes nine common Drupal-based Website performance optimization solutions.
Table 3. Drupal-based website performance optimization solutions

| Solutions                  | Descriptions                                                                 | Limitations                                                                 |
|----------------------------|------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| Using the Boost module     | Store cached pages as files, without requiring the server to execute any code and data queries | Caching static pages only for anonymous users; cannot be used in conjunction with the "Cache pages for anonymous users" solution |
| Using the Views Content Cache module | Caching based on time or content monitors the data in the cached view and dynamically adjusts the maximum cache duration | Only pages generated by Views can be cached |
| Setting up the Panels module cache | The user selects whether to cache the parts of the page, and can cache the static part in the dynamic page. | Only content in Panels can be cached |
| Caching pages for anonymous users | Cache static pages and load them directly when accessed by anonymous users | Cache exists in the database, calling cached pages will occupy database connections and Intranet traffic; only for anonymous users |
| Block Caching              | Caching blocks in pages                                                      | Anonymous users cannot see the changes in the block within the "Minimum cache lifetime" unless all caches are cleared after the modification |
| Setting bandwidth optimization | Effective for both anonymous and registered users, reducing the size and number of external files requested for page loading | The result of modifying a CSS or JS file may not take effect at once, unless you empty the cache manually. |
| Using the JavaScript Aggregator module | Merge multiple JS files and cache them                                        | Drupal 7 and later versions are not supported |
| Use Views caching          | Reduce the time for Views to execute database queries and generate pages based on time caching | Even if the data is updated in the cache lifetime, new views can only be published later. |
| Turn off unnecessary modules | Including non-business related modules and modules used in development or testing | The effect is not obvious when fewer modules. |

3. Implementation of SEO and Performance Optimization Solutions

3.1. Installation and Enablation of Modules
Visit Drupal's official website (www.drupal.org) to download the applicable module installation package and decompress it to "/sites/all/modules" in the Drupal installation directory. Select the module you want to enable via "Manage -> Module" and click "Save configuration" button.

3.2. Settings of SEO Solutions
The modules that need special settings in the SEO Checklist include the Meta tags module, the Pathauto module, and the XML Sitemap module.

3.2.1. Settings of the Meta tags module. On the module settings page, set the meta tags for the "Page title", "Description" and "Keywords" of the page type "Global: Front page".

3.2.2. Settings of the Pathauto module. In the module settings page, set the alias pattern of content type. Where "Default path pattern" applies to all content types without alias pattern. If it is set to
"node/[node:nid]", Drupal's path form is used by default.

3.2.3. Settings of the XML Sitemap module. Go to the "SEARCH ENGINES" tab via the XML Sitemap settings page, select the search engine that submits the sitemap and the minimum commit period, then save it. Set the "XML Sitemap" option for the content type that needs to add the sitemap link to "Included", then restart the Apache service.

3.3. Settings of Performance Optimizing Solutions
As the first six items listed in Table 3 are currently used in a high volume and have a good effect, the following will explain their settings separately.

3.3.1. Settings of the Boost module
- Open the "Clean URLs" service via "Settings -> Search and Metadata".
- Go to the Boost settings page, select “All pages except those listed”, and enter “user*” in the text box below. Enable caching of html and set the cache period.
- Go to the “.HTACCESS” page and set “Server’s URL or Name”, “Document Path”, “ETag Settings” and “Boost Tags”. After saving, click the “.htaccess Generation” button to copy the code generated in the text box. And insert the copied code in the “.htaccess” file according to the page prompt, and then save.
- Insert the following code in the robots.txt file in the root directory of the website and save it.

Disallow:/boost_stats.php
# Boost
Disallow:/boost_stats.php
Disallow:/CHANGEOLOG.txt

3.3.2. Settings of the Views Content Cache module
- Enter the "Views" editing page, set the "Caching" in the "Advanced" option to "Content-based".
- Click on the "Content cache" hyperlink to choose which events will update this View's cache, set "Maximum and Minimum Cache Cycle" and save.
- Repeat the above settings for all Views that need to be cached.

3.3.3. Settings of Panels cache
- Click on the "Content" page of Panels and click "Change" in the "Caching" column to enter the cache settings page.
- Set "Method" to "Simple cache" and proceed to the next step.
- Set the "Lifetime" and "Granularity" and save.
- Repeat the above settings for all Panels that need to be cached.

3.3.4. Anonymous user and block caching and bandwidth optimization settings
- Go to the Core Settings page via "Settings -> Development -> Performance" and select whether to enable "Cache pages for anonymous users". If the Boost module is used, you must choose not to enable it
- Enable block caching, and select "Minimum cache lifetime" and "Expiration of cached pages".
- Select Enable "Aggregate and compress CSS files" and "Aggregate JavaScript files" and save.

4. Application of SEO and Performance Optimization Solutions
The Chinese Culture English Portal (portal.chineseculture.biz) is a website devoted to the international dissemination of Chinese culture, which is based on Drupal 7. After realizing the required functions, according to the foregoing, the website is optimized by SEO and performance.
4.1. Installation and Enablation of Modules
The modules installed and enabled include: SEO Checklist module, Meta tags module, Pathauto module, XML Sitemap module, Administration menu module, Boost module and Views Content Cache module.

4.2. Application of SEO Solutions
- Set "Description" in "Global: Front page" to "[site: slogan]", "Page title" to "[site: name]", "Keywords" to "Chinese Culture" and so on. After setting, the browser label on the front page is shown in Figure 1.

![Figure 1 Browser Label on Home Page](image)

- Set the alias pattern for different content types according to Table 4. After accessing a content of content type "C_C News for BBS", the displayed URL is as shown in Figure 2.

| Content Type                | Alias pattern          |
|-----------------------------|------------------------|
| Default path pattern        | content/[node:title]   |
| Pattern for all Advertisement paths | advertisement/[node:nid] |
| Pattern for all Announcement paths | announcement/[node:nid] |
| Pattern for all Article paths | article/[node:nid]     |
| Pattern for all Basic page paths | basicpage/[node:nid]   |

![Figure 2. URL of the "C_C News for BBS" content type](image)

- Keep the XML Sitemap with the default settings and set the "XML Sitemap" option to "Included" in all content types. The XML sitemap engines module is enabled, and the sitemap is submitted to Google, with a minimum commit period of 12 hours.

4.3. Application of Performance Optimization Solutions
- On the Boost settings page, set the maximum html cache lifetime to 1 hour and the minimum cache lifetime to 0 seconds. Keep the default settings on the “.HTACCESS” page.
- Set the Views such as BBS_BC and Blog BC.
- Set up the Panels such as BBS and Blog.
- Due to the use of the Boost module, "Cache pages for anonymous users" is not enabled.

5. Conclusion
In order to solve the problem of incomplete and in-depth research on search engine and performance optimization based on Drupal-based websites, this paper summarizes the common solutions, gives the implementation methods, and puts them into practice. Other SEO and performance optimization solutions based on Drupal and the evaluation of the optimization effect will be the next problem that needs to be solved.

6. References
[1] Ma J 2007 Open Source Content Management System Drupal *Journal of Modern Information* 02 50-52
[2] Zhai Y 2015 Exploring the Application of Open Source System Drupal in Information Service of University Library Information and research of legal literature Z1 51-55

[3] Zhou X 2016 Research on the Construction of Chinese Reconstructed Rare Books Display Platform Based on Drupal Journal of Library Science 08 101-114

[4] Liao S 2008 Talking about the optimization of SEO Proc.Conf on 2008 New Media Technology of Internet (Shen Zhen) pp 418-431

[5] Shang S and Yin X 2016 Performance Optimization and Service Improvement of Public Archives Website Lantai World 09 57-61

[6] Wang Y 2015 How does Drupal help colleges and universities improve their website experience China Education Network 05 64-68

[7] Zuo Q 2008 Drupal's Performance Problem-- PHP Meditation No. 6 Programmer 11 91-92+8

[8] http://www.drupal.org. [2019-03-14]