

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Dans cette lettre, nous avons déjà parlé de 2 CMS majeurs (Wordpress et Joomla!) et de leurs interactions avec le monde du SEO. Assez logiquement, nous allons traiter, dans cette série de deux articles, d'un autre CMS très utilisé sur le Web, à savoir Drupal. Comment configurer l'installation de base afin de le rendre réactif aux moteurs de recherche ? Quels modules installer ? Voici bon nombre d'informations pour vous aider à faire de votre site sous Drupal une "bête de référencement"...

Le CMS Drupal (<http://drupal.org/>) devient de plus en plus populaire, notamment pour les webmasters qui veulent monter des projets complexes combinant gestion de contenu, fonctionnalités sociales, e-commerce etc... Drupal a la réputation d'être particulièrement flexible, et de permettre un bon référencement dans les moteurs de recherche.

Nous allons voir qu'en effet, Drupal est l'un des CMS qui recèle le moins de problèmes bloquants pour le SEO. Mais il faut savoir le paramétrer correctement, choisir (ou créer) les bons templates, et savoir également choisir les bons plugins pour en tirer le meilleur parti.

Drupal est hélas un CMS un peu complexe, qui déroute souvent le débutant. La courbe d'apprentissage sur Drupal est un peu plus longue que pour un Joomla! ou un Wordpress... Il est par ailleurs très « horizontal » : on peut faire à peu près n'importe quel site web avec ce CMS. En outre, une communauté nombreuse et hyperactive développe des milliers de plugins, plus ou moins optimisés pour le SEO, étendant les possibilités de base du CMS.

Impossible, dans ces conditions, de vous guider pas à pas dans chaque paramétrage possible de Drupal : chaque implémentation recèle ses propres pièges pour le SEO. C'est d'autant plus vrai à la date de rédaction de cet article, car nous sommes en voie d'adoption par les webmasters de la version 7 de Drupal, qui est sortie récemment. La transition est très progressive, car la base de plugins pour la version 7 n'est pas encore aussi importante que pour la version 6, et la version 7 contient quelques différences majeures qui freinent les webmasters les plus conservateurs. Ces différences entre les versions 6 et 7 conduisent également à des paramétrages subtilement différents.

Nous nous bornerons, dans cet article en deux parties sur l'optimisation de Drupal, à rappeler les grands principes de l'optimisation SEO sur ce CMS.

Dans le premier article, nous allons nous intéresser au paramétrage de Drupal et de ses principaux modules, pour rendre la plateforme compatible avec les règles de base du SEO. Dans la deuxième partie, nous expliquerons comment organiser son contenu, et créer des menus optimisés pour le SEO. Nous aborderons également l'optimisation des performances, les spécificités du SEO pour la version 7 de Drupal, et nous évoquerons pour finir les « plus » avancés de Drupal pour le SEO, qui rendent ce CMS hautement performant pour les webmasters exigeants.

Drupal "Out of the box" ?

On entend assez fréquemment dire qu'un Drupal installé « de base », « out of the box », est déjà optimisé pour le SEO. C'est complètement faux. Drupal est livré sous la forme d'un noyau extrêmement allégé, qui ne comprend pas en natif nombre de fonctionnalités présentes sous Wordpress par exemple. Il existe par exemple des « core modules », fournis avec la



distribution, qu'il faut activer et paramétrer correctement. L'un des « core modules » utiles pour le SEO est par exemple « taxonomy » qui permet de « tagger » les contenus et de créer des navigations optimisées. De plus, les webmasters de la communauté Drupal considèrent en fait de manière implicite qu'un Drupal « de base » comprend également les plugins populaires indispensables (comme pathauto), ce qui explique pourquoi les adeptes de Drupal affirment que leur CMS est optimisé de base pour le SEO.

En fait, il faut en plus paramétrer correctement les modules gérant les titres, les meta-tags, CCK, les views, et beaucoup d'autres choses, avant d'avoir un CMS parfait ! Mais on peut effectivement parvenir à cela facilement avec Drupal, alors que l'on sera souvent bloqué avec d'autres CMS.

L'optimisation de Drupal passe donc par l'installation préalable d'un certain nombre de modules externes que nous allons détailler maintenant.

Liste des modules SEO potentiellement utiles pour Drupal (version 6)

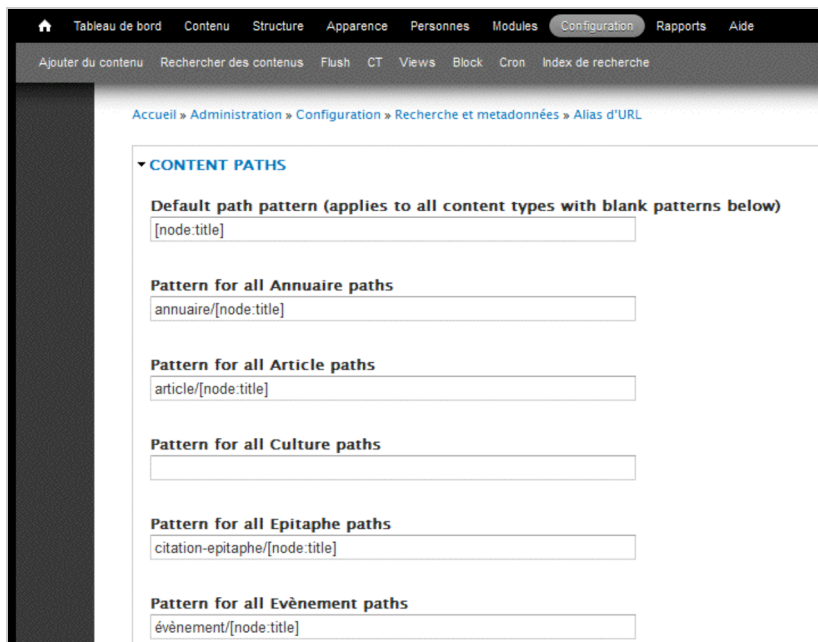
- **Page title** (http://drupal.org/project/page_title)

Page title ajoute une possibilité de personnalisation du titre sur tous les nodes.

- **Pathauto** (<http://drupal.org/project/pathauto>)

Ce module permet de créer des urls « SEO friendly » automatiquement.

Pathauto seul génère des problèmes pour le SEO par ailleurs (que règle « pathredirect » et « globalredirect » ci-après).



L'interface admin de Pathauto (version Drupal 7)

- **Pathredirect** (http://drupal.org/project/path_redirect)

Ce module permet de gérer facilement et proprement des redirections 301.

- **Globalredirect** (<http://drupal.org/project/globalredirect>)

Ce module permet de régler des problèmes gênants pour le SEO créés par les modules pathauto et les urls réécrites en général.

- **Taxonomy title** (http://drupal.org/project/taxonomy_title)

Permet de remplir une balise title pour chaque terme défini dans Taxonomy. Pour le SEO, l'outil Taxonomy est particulièrement intéressant, mais sans ce module, l'optimisation des pages créées pour chaque terme est difficile.

- **Meta tags** (alias nodewords)

<http://drupal.org/project/nodewords>

<http://drupal.org/project/metatag> (nouvelle version, pour 6.XX et 7)

Permet de personnaliser entièrement les meta tags sur les pages de Drupal. Tout bonnement indispensable.

- **XML sitemap** (<http://drupal.org/project/xmlsitemap>)

Permet de créer un fichier sitemap dynamique pour les moteurs.

- **Site verify** (pour la vérification de site GWT) : http://drupal.org/project/site_verify

Permet d'ajouter la balise de vérification de propriété exigée pour la validation de site exigée pour accéder aux données du compte Google Webmaster Tools.

- **Read More link** (http://drupal.org/project/ed_readmore)

Permet de personnaliser la gestion des liens « en savoir plus... » et de les rendre SEO friendly.

- **Google News sitemap generator** (<http://drupal.org/project/googlenews>)

Crée un sitemap optimisé pour Google News.

- **Open Calais** (<http://drupal.org/project/opencalais>)

Active la reconnaissance d'entités nommées par l'application déportée Open Calais. Permet de « reconnaître » des noms de personnes, de sociétés, de lieux et autorise derrière l'agrégation de contenus associés à ces entités nommées. Technologie sémantique avancée utilisable directement dans Drupal. Nous expliquerons comment utiliser Open-Calais pour le SEO dans la seconde partie de cet article, le mois prochain.

- **Rdf** (<http://drupal.org/project/rdf>)

Active le support du format RDF pour les échanges de données.

- **More like this**

<http://drupal.org/project/morelikethis>

<http://drupal.org/project/autoload> (ce module est un prérequis pour faire tourner

morelikethis)

Ce module affiche des liens « suggérés » en regard d'un contenu donné (liens contextuels). Ces liens contextuels sont extraits des informations stockées dans les taxonomies (nœuds associés aux mêmes étiquettes) ou des informations ajoutées par Open Calais.

- **Node hierarchy** (<http://drupal.org/project/nodehierarchy>)

Ce module permet de créer des arborescences de nœuds de manière libre, ce qui peut être particulièrement utile pour créer des blocs de liens automatiques à partir des informations données par cette arborescence.

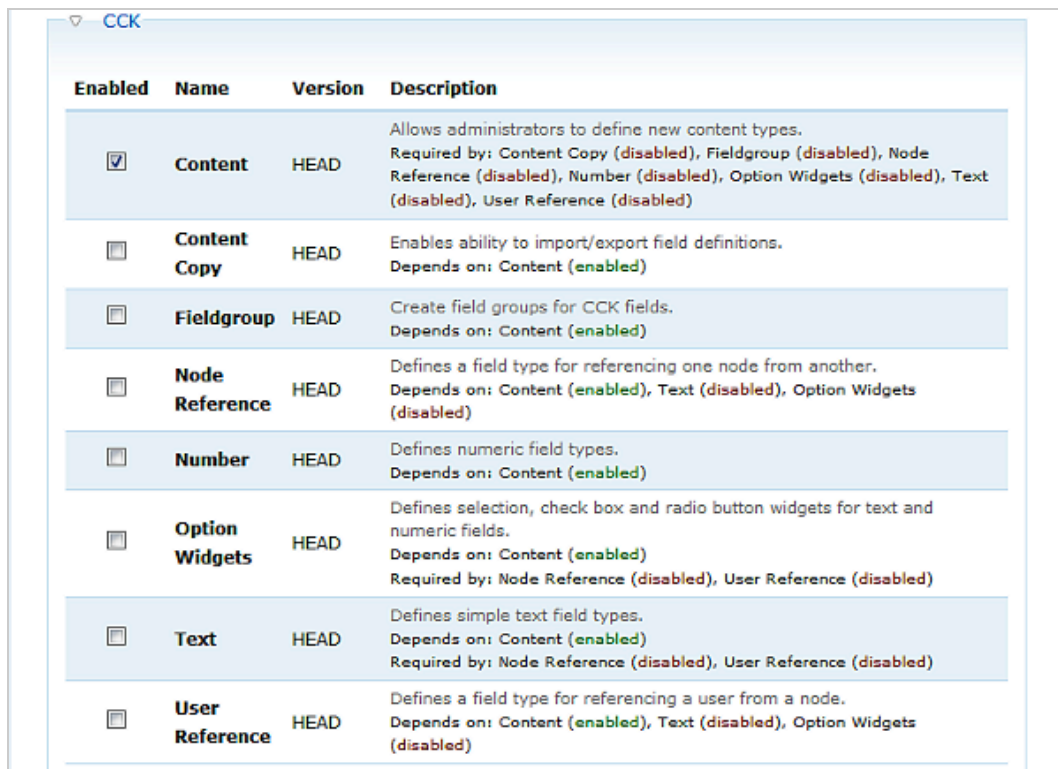
- **Alinks** (<http://drupal.org/project/alinks>)

Crée des liens automatiques sur certains termes (exemple : glossaire)

Les plug-ins utiles, mais sans impact direct sur le SEO, sans en être très éloignés...

- **cck** (<http://drupal.org/project/cck>)

Le « content construction kit » est un des modules les plus puissants pour Drupal 6. Il est tellement utile qu'il a été intégré comme élément natif dans la version 7. Il permet de rajouter des champs dans les formulaires de Drupal, et de gérer l'affichage sous formes de « vues » ce qui rend Drupal particulièrement flexible. CCK lui-même a fait l'objet de nombreuses extensions, qui facilitent le déploiement de certaines fonctionnalités supplémentaires répondant à des besoins courants.



Enabled	Name	Version	Description
<input checked="" type="checkbox"/>	Content	HEAD	Allows administrators to define new content types. Required by: Content Copy (disabled), Fieldgroup (disabled), Node Reference (disabled), Number (disabled), Option Widgets (disabled), Text (disabled), User Reference (disabled)
<input type="checkbox"/>	Content Copy	HEAD	Enables ability to import/export field definitions. Depends on: Content (enabled)
<input type="checkbox"/>	Fieldgroup	HEAD	Create field groups for CCK fields. Depends on: Content (enabled)
<input type="checkbox"/>	Node Reference	HEAD	Defines a field type for referencing one node from another. Depends on: Content (enabled), Text (disabled), Option Widgets (disabled)
<input type="checkbox"/>	Number	HEAD	Defines numeric field types. Depends on: Content (enabled)
<input type="checkbox"/>	Option Widgets	HEAD	Defines selection, check box and radio button widgets for text and numeric fields. Depends on: Content (enabled) Required by: Node Reference (disabled), User Reference (disabled)
<input type="checkbox"/>	Text	HEAD	Defines simple text field types. Depends on: Content (enabled) Required by: Node Reference (disabled), User Reference (disabled)
<input type="checkbox"/>	User Reference	HEAD	Defines a field type for referencing a user from a node. Depends on: Content (enabled), Text (disabled), Option Widgets (disabled)

L'interface admin du module CCK (Drupal 6)

- **Google analytics** (http://drupal.org/project/google_analytics)

On peut « tagger » le site de différentes manières, mais la présence d'un outil de web analytics est indispensable.

- **HTML Purifier** (<http://drupal.org/project/htmlpurifier>)

Permet de nettoyer le HTML produit dans les zones écrites par des utilisateurs. Utile uniquement si ce type de contenu existe sur le site.

- **Syndication** (<http://drupal.org/project/syndication>)

Crée une centralisation des flux rss générés par Drupal : attention, nécessite un paramétrage particulier pour éviter des effets de bord SEO dûs au crawl des flux par les robots.

Créer des URL optimisées avec Path Auto

Les URL de base de Drupal ne sont pas optimisées pour le référencement et ont cette forme :

- [node/123](#)
- [user/5](#)
- [taxonomy/term/1](#)

Pour créer des URL optimisées, on peut activer la création des « clean urls » dans l'interface d'administration, et créer des alias un peu plus « sexy » pour ces urls en utilisant le « core module » path, disponible dans la distribution de base.

« Path » vous permettra de générer des URL du type www.domaine.com/tondeuse-a-gazon au lieu de www.domaine.com/node/123.

Cette stratégie d'alias devient vite fastidieuse sur un gros site, et difficile à maintenir. Pour gérer la réécriture d'URL sur un site plus volumineux, on installera "Pathauto", un module beaucoup plus flexible et complet, qui vous permettra de créer des syntaxes de réécriture complexes pour chaque type de contenu.

Ex : [article/\[node:title\]](#) (notation drupal 7)

qui génère une URL de type : www.domaine.com/article/hold-up-a-la-societe-generale

Générer la bonne syntaxe d'URL pour des cas complexes, impliquant en particulier des contenus exotiques créés avec CCK, croisés avec la taxonomie, demande une certaine expérience. Il faut notamment prendre des précautions pour générer des URL uniques, et éviter de créer des URL interminables. Certaines syntaxes sont par ailleurs dangereuses (« hackables »), d'autres posent des problèmes de performance. Mais la communauté Drupal peut vous aider à trouver la bonne solution.

[nid]	Node ID
[type]	Node type
[type-name]	Node type (user-friendly version)
[title]	Node title
[title-raw]	Unfiltered node title. WARNING - raw user input.
[author-uid]	Node author's user id
[author-name]	Node author's user name
[author-name-raw]	Node author's user name. WARNING - raw user input.
[term]	Name of top taxonomy term
[term-raw]	Unfiltered name of top taxonomy term. WARNING - raw user input.

Quelques exemples de « tokens » (variables) utilisables dans les règles de génération d'URL avec pathauto.

Attention à cette phase : le choix d'une syntaxe d'URL doit être fait une bonne fois pour toutes. Si vous devez corriger ou ajuster votre syntaxe par la suite, cela peut avoir des conséquences néfastes pour votre référencement. Réfléchissez bien au schéma d'urls que vous désirez pour votre site avant de l'implémenter.

Drupal et les URL dupliquées (DUST)

Drupal, comme de nombreux CMS, accepte à la fois la syntaxe par défaut (<http://www.domaine.com/node/123>), et la syntaxe réécrite (<http://www.domaine.com/fiche-produit/tondeuse-a-gazon>). Les deux URL renvoient exactement la même page, créant des cas de « DUST » (*Different URL Same Text*), ce qui n'est pas recommandé pour le SEO.

Heureusement, deux modules permettent d'éliminer le problème. Le premier, « **Global Redirect** », permet d'effectuer une redirection 301 depuis la syntaxe non réécrite (« [node/123](#) ») vers la syntaxe réécrite avec pathauto (« [produit/tondeuse-a-gazon](#) »).

Le second, « **Path Redirect** », permet de gérer les urls générées au cas par cas avec « Path », ou de traiter des vieilles syntaxes ou des syntaxes erronées figurant dans des backlinks.

Optimiser le fichier robots.txt et le .htaccess

Les fichiers robots.txt et .htaccess fournis avec la distribution ne sont pas optimisés et doivent être modifiés pour éviter des problèmes avec les crawlers des moteurs de recherche.

Les modifications à apporter au robots.txt

Le fichier par défaut fourni avec Drupal serait presque parfait s'il n'y avait pas une particularité de fonctionnement dans Drupal qui produit un effet de bord inattendu.

Un coup d'œil rapide au robots.txt d'un site Drupal montre que les pages d'administration du site sont interdites aux robots. Ce qui respecte parfaitement les « règles de l'art » en la matière.

```
[...]  
# Paths (clean URLs)  
Disallow: /admin/ <- les pages admin sont interdites aux crawlers  
Disallow: /comment/reply/  
Disallow: /filter/tips/  
Disallow: /node/add/  
[...]
```

Le problème, c'est que par défaut, www.mondomaine.com/admin/ pointe vers la même page que www.mondomaine.com/admin (sans le dernier slash) dans Drupal. **Euh, et alors ?? :-)**

Ce problème du « trailing slash » peut être résolu partiellement en paramétrant « global redirect » pour rediriger une syntaxe vers l'autre, mais la solution la plus sûre est de modifier légèrement le robots.txt pour empêcher la version sans slash d'être crawlée.

Il suffit de rajouter des lignes Disallow sans le trailing slash (solution qui marche avec tous les moteurs).

```
[...]  
# Paths (clean URLs)  
Disallow: /admin/  
Disallow: /comment/reply/  
Disallow: /filter/tips/  
Disallow: /node/add/  
#lignes ajoutées (sans le trailing slash)  
Disallow: /admin  
Disallow: /comment/reply  
Disallow: /filter/tips  
Disallow: /node/add  
[...]
```

Le fichier .htaccess

Le fichier .htaccess de base est correct, mais il est prévu pour une installation par défaut. Certains paramètres intéressants pour le SEO sont bien présents, mais les lignes correspondantes sont commentées : il faut penser à les décommenter pour obtenir le comportement souhaité.

Il s'agit en particulier de paramétrer le serveur web pour qu'il effectue une redirection 301 lorsque que quelqu'un appelle une URL en <http://domaine.com> vers <http://www.domaine.com>. Ces lignes figurent à la fin du fichier .htaccess :

```
# If your site can be accessed both with and without the 'www.' prefix, you  
# can use one of the following settings to redirect users to your preferred  
# URL, either WITH or WITHOUT the 'www.' prefix. Choose ONLY one option:  
#  
# To redirect all users to access the site WITH the 'www.' prefix,  
# (http://example.com/... will be redirected to http://www.example.com/...)  
# adapt and uncomment the following:  
# RewriteCond %{HTTP_HOST} ^example\.com$ [NC]  
# RewriteRule ^(.*)$ http://www.example.com/$1 [L,R=301]  
#  
# To redirect all users to access the site WITHOUT the 'www.' prefix,  
# (http://www.example.com/... will be redirected to http://example.com/...)  
# uncomment and adapt the following:  
# RewriteCond %{HTTP_HOST} ^www\.example\.com$ [NC]  
# RewriteRule ^(.*)$ http://example.com/$1 [L,R=301]
```

Si vous souhaitez mettre en place la redirection du <http://xxxxx> vers le <http://www.xxxxx>, il faut décommenter (enlever le # en début de ligne) les deux premières lignes rouges. Si vous voulez faire l'inverse (rediriger le <http://www.xxxxx> vers le <http://xxxxx>), il faut décommenter les deux dernières lignes.

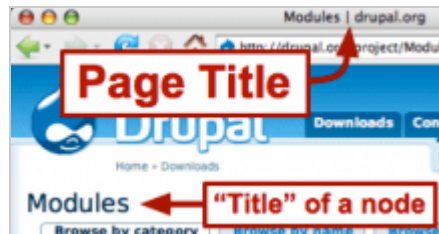
Le problème des balises meta et des Titles

Par défaut, Drupal ne gère pas les balises title et les meta tags, en tout cas pas au niveau de granularité souhaitable pour le SEO, c'est-à-dire un contenu différent pour ces balises dans chaque page html !

Dans Drupal 7, il est possible de rajouter soi-même les champs nécessaires dans les types de contenu et de les insérer dans un template au bon endroit.

Dans Drupal 6, la solution la plus courante est d'installer deux modules :

- « Title » tout d'abord, qui permet de générer un <title> différent pour chaque page de contenu. Il s'interface avec le module Tokens, ce qui permet de générer des règles sophistiquées pour la création des <title>.



Title existe pour les versions 6 et 7 de Drupal

- « Metatags » ensuite, qui permet de créer des metas optimisées pour chaque page. Ce module, qui s'appelait autrefois « nodewords » est en cours de développement pour Drupal 7, mais on trouve une version alpha pour D7 qui fonctionne à peu près normalement.

AVERTISSEMENT : attention à ne pas installer n'importe quel plugin !

L'une des tentations les plus courantes chez les débutants, c'est de se jeter sur tous les plugins du catalogue et d'installer tous les gadgets possibles et imaginables. Cette boulimie de modules peut avoir rapidement des effets désastreux.

Elle va rapidement transformer votre interface d'administration en tableau de bord de Boeing et créer des interactions complexes entre modules qui vous empêcheront de déboguer un problème causé par l'un des plugins.

Certains plugins ont le défaut de créer des champs et des tables dans la base de données de Drupal, et d'alourdir celles-ci. Les requêtes les plus courantes sur la base de données peuvent devenir d'une lenteur intolérable à cause d'un seul plugin.

D'une manière générale, l'accumulation de plugins nuit aux performances. Il n'est pas rare de déclencher « l'écran blanc de la mort » après avoir installé trop de plugins. Cette « page blanche », qui laisse parfois apparaître un message du type « *error : maximum execution time exceeded* » apparaît quand le temps de calcul du script php dépasse le temps d'exécution maximum alloué dans le fichier php.ini.

Il faut savoir rester raisonnable, et ne pas hésiter à désinstaller les modules peu utiles s'il le faut. Installer tous les modules cités dans la liste des plugins recommandés au début de l'article peut suffire à diminuer sérieusement les performances sur un serveur disposant de faibles ressources.

Drupal peut facilement résister à des charges très importantes, et délivrer des pages avec des performances excellentes, mais il faut le paramétrer correctement pour cela. Nous verrons comment dans le prochain article.

Nous verrons également comment organiser son contenu pour en tirer le meilleur parti pour le SEO, et comment optimiser les templates Drupal. Après avoir survolé les nouveautés de Drupal 7 et leurs conséquences pour le SEO, nous expliquerons comment utiliser les modules Open Calais et SOLR pour créer des pages utiles pour le SEO avec Drupal.

Dans le prochain article (novembre 2011), nous vous parlerons des points suivants :

- Optimiser les performances de Drupal : le cache et les paramètres ad hoc ;
- Organiser son contenu avec Drupal ;
- Optimiser le template Drupal pour le SEO ;
- Les spécificités de Drupal 7 ;
- Le plus offerts par Drupal : Open Calais et SOLR.

Philippe YONNET, *Directeur SEO international, Twenga.*

Réagissez à cet article sur le blog des abonnés d'Abondance :

<http://blog-abonnes.abondance.com/2011/10/le-referencement-sur-le-cms-drupal-1ere.html>