

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. Le mois dernier, nous avons décrit la meilleure façon de configurer l'installation de base afin de le rendre réactif aux moteurs de recherche ? Ce mois-ci, la seconde partie de cet article vous explique comment améliorer les performances du site, créer un maillage interne optimal entre ses pages, mais aussi comment choisir et construire un modèle de page "SEO inside", tout en finissant avec la description de deux modules indispensables au bon référencement de votre source d'information...

Dans l'article du mois dernier, nous avons fait le tour des paramétrages et des modules nécessaires pour rendre Drupal compatible SEO. Mais il reste encore du travail à faire pour :

- Optimiser les performances ;
- Créer un maillage interne optimal ;
- Choisir ou construire un modèle de page (template) compatible SEO.

Aujourd'hui, nous allons donc développer ces trois sujets, et nous terminerons en examinant un problème spécifique à la version 7 de Drupal, ainsi que les possibilités apportées par deux extensions extrêmement puissantes : Open Calais et SolR.

Mais avant d'aller plus loin, il faut s'attarder sur un point qui peut s'avérer fâcheux pour un site qui peut espérer du trafic en provenance de Google Images.



Débloquer le référencement des images dans Drupal 6

Ceux qui ont conçu le robots.txt par défaut dans Drupal 6 sont partis du principe que les webmasters ne souhaitent pas que leur contenu (y compris les images) soient "scrapé" ou copié.

Le répertoire `/sites/` est donc placé dans une ligne « `disallow:` » du fichier robots.txt. Le répertoire « images » par défaut se situe dans le répertoire `/sites/`, ce qui signifie qu'aucun crawler n'aura accès aux images, *a fortiori* le crawler spécialisé « images » de Google.

Ce problème a été corrigé dans les versions les plus récentes de Drupal 6 et Drupal 7 (versions postérieures à la 6.20), en débloquent le crawl de `/sites/`. Mais si vous disposez d'une version plus ancienne de Drupal, et si vous voulez permettre aux crawlers d'accéder à vos images, il faut ajouter les lignes suivantes au robots.txt :

```
User-agent: Googlebot-Image  
Allow: /
```

Cette modification fonctionne même si le répertoire `/sites/` est bloqué, ce qui est sans doute préférable pour sécuriser l'accès aux fichiers uploadés par le webmaster.

Optimiser les performances de Drupal

Drupal vient d'être élu (le 11 novembre dernier) par la société d'éditions « Packt Publishing » meilleur CMS pour ses performances, devant Joomla! et Plone. C'est un résultat logique si on

compare Drupal à Joomla!, beaucoup plus surprenant comparé à Plone, qui est un CMS très performant et économe en ressources.

En réalité, il ne faut pas sous-estimer les problèmes de performance que l'on peut rencontrer avec Drupal. Or un site qui n'est pas en mesure de délivrer ses pages rapidement aux robots des moteurs de recherche verra son référencement affecté, parfois de manière très sérieuse. Tout d'abord, Drupal est un système relativement lourd qui peut s'avérer surdimensionné pour certains projets comme des blogs simples ou des sites éditoriaux simples comportant peu de pages. Néanmoins, même si d'autres solutions peuvent s'avérer plus performantes, l'écart dans la vitesse de génération des pages peut s'avérer minime, surtout avec le cache activé. Ce petit handicap peut donc être corrigé et n'a en général aucun impact sur le SEO.

Le problème devient plus sérieux pour des sites complexes, qui comprennent beaucoup d'URL avec un trafic important. Dans ce contexte, la moindre erreur de paramétrage peut conduire à une dégradation des performances désastreuses. Alors qu'au contraire, un bon paramétrage transformera votre Drupal en « bête de course ».

Bien choisir ses modules

Le point faible de Drupal est la complexité de sa base de données. Générer une page nécessite l'appel de nombreuses requêtes sur des dizaines de tables. La situation empire si l'on a ajouté de nombreux modules. Tous ne sont pas optimisés, et l'ajout d'un seul module mal écrit peut conduire à des ralentissements considérables. Nous invitons le lecteur à relire les conseils donnés dans l'article précédent quant aux choix des modules.

Mais nous rappelleront quand même ce conseil de bon sens : désactivez les modules et les blocs Drupal dont vous ne vous servez pas !

Optimiser la base de données

Si vous connaissez bien Mysql, on peut détecter la liste des requêtes lentes en consultant les logs de requêtes, et tester ces requêtes avec la commande EXPLAIN. Il n'est pas rare que le simple ajout d'un index supplémentaire dans Mysql accélère la génération de pages de manière considérable.

Il existe un module (DB Maintenance) qui peut vous aider à optimiser les tables depuis l'admin de Drupal : <http://drupal.org/node/41588>. Attention, il n'est pas compatible avec toutes les versions de php et de mysql. Mais il est possible de lancer les requêtes d'optimisation à l'aide de phpmyadmin par exemple (Exemple :*OPTIMIZE TABLE accesslog,cache,comments,node,users,watchdog;*)

Si votre site est hébergé sur un serveur dédié, il est possible d'optimiser les paramètres Mysql pour Drupal. Voici les paramètres recommandés pour Drupal 5/6 :

```
max_connections = 800
max_user_connections = 800
key_buffer = 36M
myisam_sort_buffer_size = 64M
join_buffer_size = 2M
read_buffer_size = 2M
sort_buffer_size = 3M
table_cache = 1024
thread_cache_size = 286
interactive_timeout = 25
wait_timeout = 1800
connect_timeout = 10
max_allowed_packet = 1M
max_connect_errors = 999999
query_cache_limit = 1M
query_cache_size = 16M
```

query_cache_type = 1
tmp_table_size = 16M

Activer le cache

Drupal possède un système de cache performant : bien paramétré, il permet de régler bien des problèmes, y compris lorsque la génération de pages est lente.

Attention toutefois : le cache (dynamique) ne résout pas tout pour de gros sites. En effet, il est très probable que Googlebot par exemple tombera régulièrement sur des pages expirées, et déclenchera une nouvelle génération de page et sa mise en cache. Dans ce cas, on peut avoir un site apparemment rapide pour l'utilisateur, mais toujours très lent pour Googlebot.

Néanmoins, l'utilisation du cache est toujours un plus qui soulage considérablement la base de données et le serveur web, il faut donc vérifier qu'il est bien activé.

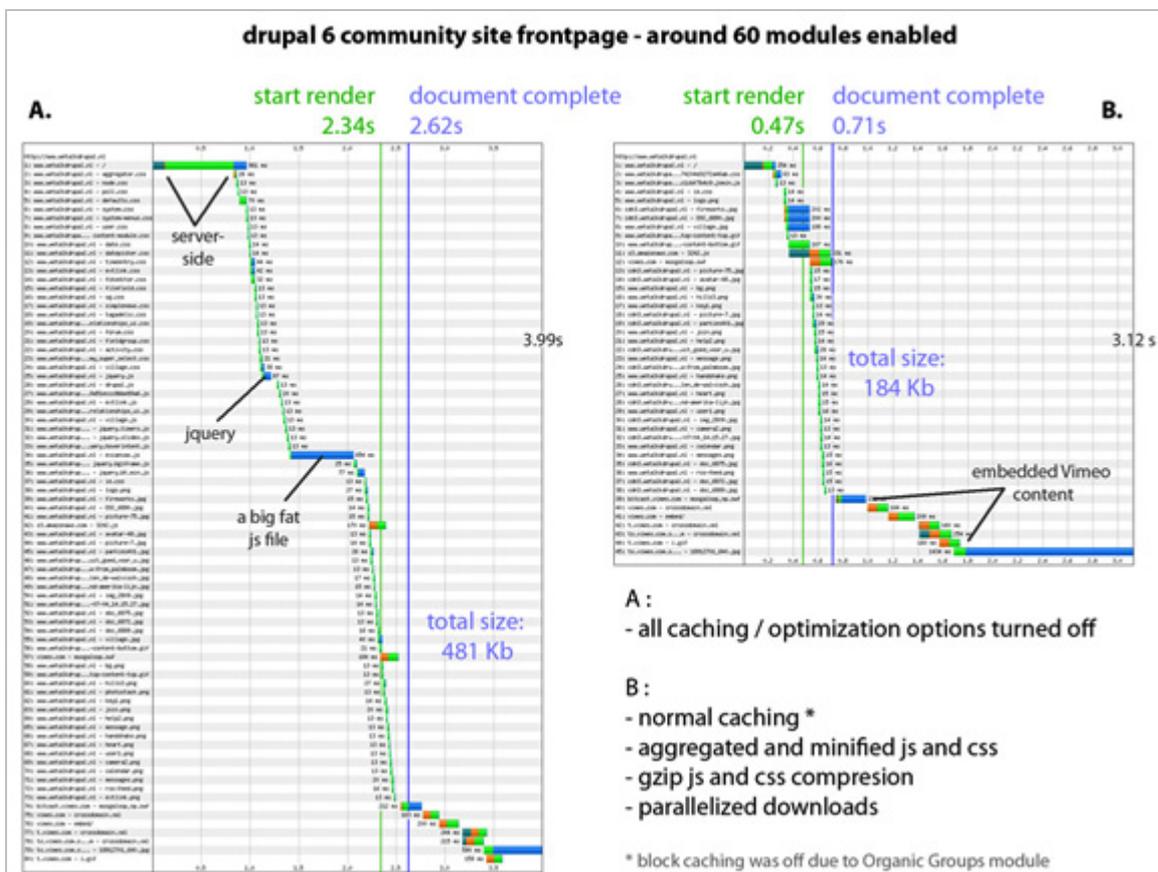
Les paramètres de cache sont accessibles dans l'admin dans la rubrique : *Site Configuration > Performance*.

Important : n'oubliez pas d'activer le caching des blocks également.

Autres paramètres

Pour améliorer les performances, n'oubliez pas d'activer (toujours dans la rubrique « performance » :

- Page compression ;
- Optimize CSS files ;
- Optimize JavaScript files.



Etude des temps de téléchargement des éléments d'une page (graphe en cascade). A gauche, sans optimisation activée, à droite avec le cache et les optimisations activées.

Organiser son contenu avec Drupal

Il existe de multiples façons d'organiser son contenu avec Drupal, certaines sont simples et classiques, d'autres sophistiquées. Jusqu'à la version 7, une logique d'architecture de type « blog », « sites d'actus » ou « livre » était mise en avant, mais c'est moins vrai avec la version 7 qui permet véritablement de créer les pages, les types de contenu et le maillage interne que l'on veut.

Compte tenu de la souplesse et de la sophistication offerte par Drupal dans ce domaine, il n'est pas possible de décrire toutes les approches possibles dans cet article. Nous nous bornerons donc à rappeler quelques principes, et les spécificités de Drupal dans ce domaine.

Contrairement à d'autres CMS, Drupal s'appuie assez peu sur la notion de « pages web », pour lui préférer la gestion de « nœuds » (« nodes » en anglais). Un « nœud » correspond en gros à tout ce qu'un utilisateur peut « poster » dans la base du CMS (un article, un billet de blog, un post de forum, le contenu d'un formulaire etc. Exception : les commentaires sont associés à un nœud, mais ne constituent pas des nœuds). Cette séparation entre la logique de stockage de l'information, et la manière de présenter ces infos dans une page offre une grande souplesse. En particulier, créer des pages composites rassemblant des « nœuds » de nature différentes est tout à fait possible...

Mais ce qui est très spécifique à Drupal, c'est que par défaut le CMS ne propose aucune organisation hiérarchique entre les pages, en tout cas aucune stockée dans la base de données, comme on le voit dans d'autres CMS. On peut donc créer le maillage que l'on veut depuis zéro, avec une totale liberté. C'est assez déroutant au début, mais bien pratique parfois...

Classiquement, les webmasters créent une arborescence virtuelle en utilisant les informations stockées dans la taxonomie (une base de mots clés permettant de « baliser » les contenus pour les regrouper suivant la logique voulue : hiérarchique, thématique, analogique, structurelle...).

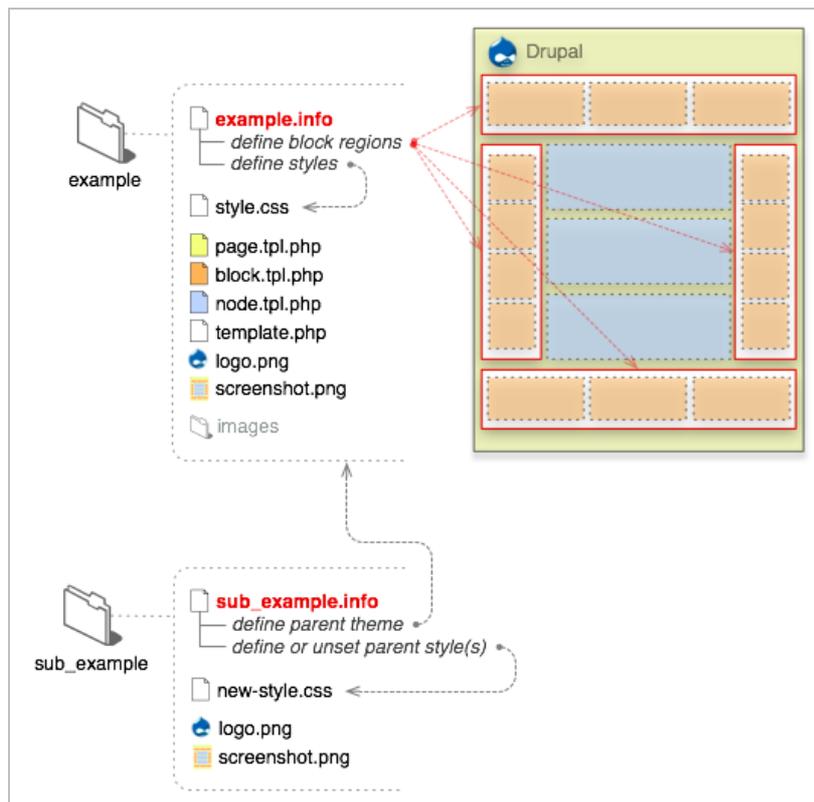
Une fois cette organisation créée, on peut développer une logique de navigation par liens en s'appuyant sur :

- Les menus paramétrables ;
- Des « blocs » qui contiennent des liens de navigation (navigation vers les « siblings », les nœuds apparentés, ou des « nœuds similaires) .

Mais cette méthode n'est pas la seule possible. En fait, Drupal autorise toutes les fantaisies en matière de maillage interne, et permet de créer des règles automatiques sophistiquées pour gérer les blocs de navigation. Ce qui représente un réel avantage pour faire du SEO avancé sur un site comportant un grand nombre de pages, en évitant la solution habituelle consistant à « hacker » le CMS.

Optimiser un template Drupal pour le SEO

Les modèles de pages (*templates*) de Drupal sont entièrement modifiables et paramétrables : il est donc possible de faire générer au CMS exactement le code HTML que l'on souhaite. Avoir un code parfaitement optimisé pour le SEO est donc parfaitement possible avec Drupal.



La structure d'un template dans Drupal 6.

Personnaliser un template s'avère assez simple à l'usage, mais demande un apprentissage préalable le temps de se familiariser avec la logique du système de modèles. Le système de sous-thèmes et de sous-templates permet de gérer facilement des variantes avec une économie remarquable de lignes de code.

Hélas, beaucoup de webmasters partent d'un « thème » existant. Or ces thèmes sont rarement optimisés entièrement pour le SEO. On se méfiera en particulier de ceux qui se proclament « optimisés pour le SEO » ou « compatibles SEO ». Sauf exception, TOUS les templates gratuits ou commerciaux appellent des modifications parfois minimes, parfois très importantes, avant de pouvoir être considérés comme parfaitement compatibles.

Il est donc éminemment recommandé d'apprendre à modifier un template pour optimiser son code parfaitement. Au passage, cela permet aussi d'accéder à toute la puissance et la souplesse de Drupal dans la personnalisation de la restitution des nœuds, et donc du contenu du site web.

Les nouveautés de Drupal 7 pour le SEO

La version 7 de Drupal apporte de nombreux changements et améliorations au cœur de Drupal, notamment en intégrant des outils développés auparavant sous forme de modules séparés (comme le fameux module CCK).

D'un point de vue SEO, la version 7 est par contre un non-événement, car elle ne corrige pas les défauts relevés dans le premier article, et n'apporte pas d'avancées majeures. Par contre, la puissance et la souplesse de Drupal sont toujours là.

Le passage à la version 7 se fait donc de manière quasi neutre pour le référencement. Il faut néanmoins signaler un problème de mise à jour de module qui agace les webmasters habitués à la puissance et la simplicité d'utilisation du module metatag cité dans le premier article.

L'absence (temporaire ?) d'un module « meta TAG » !

La plupart des webmasters s'appuyaient sur le module nodewords en version 5, metatags en version 6 pour corriger l'absence de la gestion des champs metas (et title) par défaut dans Drupal.

Ces champs ne sont toujours pas gérés dans Drupal 7, or le module metatag compatible avec la version 7 n'est toujours pas finalisé à l'heure où nous écrivons cet article. Une version « bêta » existe, mais elle comporte encore de nombreux bugs.

Il existe un module alternatif (« metags quick »), qui permet d'ajouter les champs aux principaux nœuds, et de saisir ses metas à la main, nœud par nœud, mais cette solution est impraticable sur un gros site.

C'est un très gros manque pour les webmasters un peu novices qui sont soucieux de leur SEO.

Pour les plus aguerris, il existe néanmoins une solution, assez simple :

- Ajouter les champs nécessaires en modifiant chaque type de contenu ;
- Utiliser le module « rules » intégré dans les modules de base de la version 7 pour déclencher la fabrication d'une balise meta à l'aide d'une règle automatique, déclenchée lors de l'ajout ou la modification d'un nœud ;
- Modifier les templates pour qu'ils affichent bien le contenu des champs « metas » ajoutés.

Les "plus SEO" offerts par Drupal : Open Calais & SOLR

Nous ne pouvons pas évoquer le référencement des sites reposant sur le CMS Drupal sans évoquer deux outils complémentaires, très facile à interfacier avec Drupal, et qui s'avèrent extrêmement utiles pour le SEO. Nous voulons parler d'Open Calais et de SolR.

Open Calais : la reconnaissance d'entités nommées dans un plugin Drupal !

Open Calais est une initiative de la société Thomson Reuters, destiné à favoriser le développement des outils sémantiques sur les sites grand public. Il s'agit d'un service en ligne (en mode SaaS), qui permet de baliser automatiquement les entités nommées dans un contenu non structuré (les noms de lieux, de personnes, les raisons sociales de société etc...), et qui renvoie des résultats au format RDF.

Pour en savoir plus, il suffit de se rendre sur le site de l'initiative :

<http://www.opencalais.com/>

A quoi cela sert-il ? Imaginez que vous développez un site sur les jeux vidéos. Vous avez envie de développer une page par jeu vidéo, et vous voulez aussi faire apparaître sur cette page un lien vers vos autres articles qui évoquent ce jeu.

C'est facile à faire à la main quand on a peu de pages : il suffit de rajouter un tag issu de la taxonomie à la main, pour relier la page « jeu vidéo » et les articles associés. Mais imaginez que vous deviez faire cela sur un site de 100000 pages, et sur des dizaines de thématiques à la fois : la tâche devient vite incommensurable.

Mais Open Calais peut vous aider, et peut même vous permettre de repérer dans les flux que vous reprenez, des nouveaux noms de jeu vidéo qui viennent de faire leur apparition sur le net !

Voici un exemple de texte (les illustrations sont issues de cet article :

<http://www.lullabot.com/articles/introduction-calais>).

Home > Create content

Create Story

Title: *

Nintendo DS

Menu settings

Split summary at cursor

Body:

The Nintendo DS (Nintendō Dīesu ,sometimes abbreviated to DS or NDS?) is a dual-screen handheld game console developed and manufactured by Nintendo. It was released in 2004 in Canada, the United States, and Japan. The console features a clamshell design, similar to the Game Boy Advance SP, with two LCD screens inside—with the bottom one being a touchscreen. The Nintendo DS also features a built-in microphone and supports wireless IEEE 802.11 (Wi-Fi) standards,[4] allowing players to interact with each other within short range (10–30 m, depending on conditions) or online with the Nintendo Wi-Fi Connection service, which launched later in the console's lifespan. This was the first Nintendo console to be released in North America prior to Japan.

The system's code name was Nitro,[5] and this can be seen in the model number that appear on the unit (NTR-001).[6] The console's name officially refers to "Developers' System", in reference to developers of new game designs; the system was meant to inspire, and "Dual Screen" the system's

Une fois ce contenu publié, un nouvel onglet est généré par Open Calais, qui comprend tous les tags identifiés par le service SaaS :

Save

Calais Document Category:
Technology
Calais Suggestions: Technology

Company:
Nintendo, Nintendo of America
Calais Suggestions: Nintendo Nintendo of America

Continent:
Europe, North America
Calais Suggestions: North America Europe

Country:
Australia, Canada, Japan, United States
Calais Suggestions: Japan United States Canada Australia

Events Facts:
Company Technology
Calais Suggestions: Company Technology

Technology:
Wi-Fi
Calais Suggestions: Wi-Fi

Anniversary:

City:

Dans ce cas précis, le document est correctement associé à la thématique « technologie », associé à la société Nintendo, et associé également à des localisations comme l'Australie, le Japon, les USA etc...

Evidemment, une fois ces balises triées, cela permet :

- Soit d'identifier de nouveaux noms de société, et de créer la page ad hoc associée immédiatement ;

- Et d'enrichir en même temps les pages existantes, comme la rubrique « technologie », ou la page consacrée aux articles concernant le Canada ou le Japon.

Open Calais est donc un outil permettant de doper la productivité du webmaster et de l'aider à créer de nouvelles pages pertinentes à partir de son contenu existant (pages « mot clé », ou pages hub rassemblant tous les contenus associables à une thématique). Cet outil est également très utile pour générer des suggestions d'articles similaires.

L'implémentation d'Open Calais dans Drupal est rendue extrêmement simple grâce à ce module :

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

Solr et Drupal

Lorsque l'on travaille sur un gros site exploitant des contenus structurés ou non, il est souvent utile de pouvoir créer des pages utilisant le moteur de recherche interne du site pour générer des pages qui correspondent à un filtre donné : exemple « marque + modèle » ou « service + ville », ou même à une recherche *full text* donnée si un champ structuré manque à l'appel.

Dans la plupart des cas, les moteurs internes des CMS présentent de nombreuses limitations. Parmi les plus courantes on peut citer :

- Une recherche *full text* peu performante ;
- Une recherche limitée à certains types de contenus ;
- Une gestion des « facettes » malaisée ;
- Un manque de pertinence des résultats.

Drupal offre, grâce à plusieurs modules performants, une passerelle avec les meilleurs outils de recherche interne du marché. Nous ne parlerons ici que de la passerelle avec SolR, car ce moteur de recherche représente à l'heure actuelle le *ne plus ultra* des fonctionnalités offertes par les moteurs open source, et son intégration dans Drupal est particulièrement complète et réussie.

SolR est un moteur de recherche libre, qui s'appuie sur la bibliothèque Lucène (une interface Drupal/Lucène existe également par ailleurs). Il est maintenu par la fondation Apache. Attention : Lucène est une bibliothèque Java : le moteur de recherche doit donc tourner sur un serveur équipé classiquement de Tomcat, ce qui évidemment n'est pas toujours possible sur un mutualisé.

The screenshot shows a search results page for 'Digital cameras'. The page is annotated with several callouts explaining faceted navigation features:

- Manufacturer is a facet, a way of categorizing the results**: Points to the 'Manufacturer' filter section.
- Canon, Sony, and Nikon are constraints, or facet values**: Points to the list of manufacturers (Canon USA (5), Sony (2), Nikon (2), Olympus (6), Pentax (2)).
- The facet count or constraint count shows how many results match each value**: Points to the counts next to each manufacturer name.
- The breadcrumb trail shows what constraints have already been applied and allows for their removal**: Points to the breadcrumb trail at the bottom left.
- Regular search results list**: Points to the main list of search results at the bottom right.

The search results are refined by the following filters:

- Manufacturer**: Canon USA (5), Sony (2), Nikon (2), Olympus (6), Pentax (2)
- Resolution**: 6 megapixels (3), 8 megapixels and up (14)
- Zoom range**: 3X to 4X (11), 8X to 12X (1)
- More**: LCD size, Image stabilizer, Flash memory, Still image format, Maximum ISO

The breadcrumb trail shows: you selected: \$400 - \$500, SLR, remove all.

The search results list shows 17 results. The first result is: Canon EOS Rebel XS (silver, with 18-55mm lens) priced at \$459 to \$699 at 15 stores.

Un exemple typique de navigation « facettée » sur une page de résultats du site Cnet. Chaque facette constitue un « filtre », ici sur le fabricant, la résolution de l'appareil, le niveau de zoom... Les couples valeur x filtre, baptisées contraintes, permettent de sélectionner des résultats plus précis. Le moteur renvoie des compteurs qui permettent également de savoir combien de résultats vont être renvoyés. Ces fonctionnalités sont natives dans l'implémentation SolR sur Drupal.

La puissance et la souplesse du moteur SolR permet donc de créer un grand nombre de nouvelles pages. La plupart des pages correspondant à des facettes, ou à une recherche « full text », n'ont aucun intérêt pour le SEO. C'est d'ailleurs une mauvaise idée de laisser un crawler de moteur de recherche accéder à une navigation en « facettes ».

Par contre, il est très intéressant d'ouvrir certaines de ces pages au crawl lorsqu'elles sont susceptibles de correspondre à des recherches fréquentes d'internautes.

Notez bien que les pages présentées au moteur ne doivent pas être de simples pages de résultats sorties « brut de décoffrage » des moteurs. Les moteurs de recherche comme Google ont une règle : ne pas indexer de résultats d'autres moteurs de recherche. De plus, un utilisateur qui tombe, après une recherche dans Google, sur une liste de résultats mal présentée, risque de ne pas rester sur votre site.

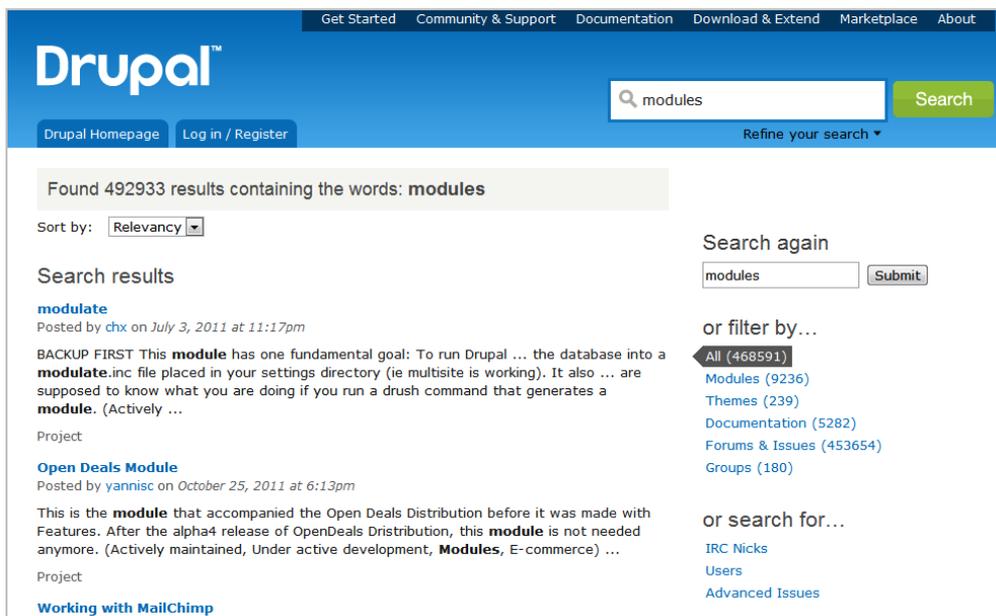
L'objectif pour ces nouvelles pages ouvertes à l'indexation consiste donc à utiliser les résultats du moteur pour rassembler les contenus à afficher pour une facette ou un mot clé donné. Cela peut s'avérer une tâche impossible avec d'autres CMS, mais avec Drupal, cela demande juste :

- De créer un ou plusieurs types de contenu adapté ;
- De créer un template spécifique ;
- Et de combiner plusieurs blocs personnalisés.

SolR dispose par ailleurs de nombreuses fonctionnalités avancées, qui en font un moteur de recherche très puissant et adaptable à toutes les configurations de site. Il est notamment possible de régler finement les poids utilisés par l'algorithme de classement, pour améliorer la pertinence des résultats retournés.

Le module d'interface avec Apache SolR se trouve à cette adresse :

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



The screenshot shows the Drupal.org search interface. At the top, there is a navigation bar with links for 'Get Started', 'Community & Support', 'Documentation', 'Download & Extend', 'Marketplace', and 'About'. Below this is the Drupal logo and a search bar containing the word 'modules'. The search results page displays 'Found 492933 results containing the words: modules'. It includes a 'Sort by' dropdown set to 'Relevancy', a 'Search again' section with a 'Submit' button, and a 'filter by...' section with various categories and counts: All (468591), Modules (9236), Themes (239), Documentation (5282), Forums & Issues (453654), and Groups (180). The main content area shows search results for 'modulate' and 'Open Deals Module', each with a brief description and a 'Project' link.

Le site Drupal.org utilise SolR comme outil de recherche. La navigation à facettes typique de SolR apparaît dans la colonne de droite.

Drupal : un CMS puissant, POTENTIELLEMENT COMPATIBLE SEO, mais à déconseiller aux webmasters paresseux

La passerelle avec Open Calais ou avec SolR ne sont que deux exemples des outils puissants auxquels Drupal donne facilement accès. Nous aurions également pu citer le panier Übertart , qui permet de créer des sites e-commerce complets et que l'on peut rendre parfaitement compatible SEO (ce qui est rarement observé avec les applications ecommerce).

Clairement, il n'existe aucun obstacle natif dans Drupal qui empêche d'optimiser complètement un site. Les défauts originels sont corrigibles moyennant quelques ajouts de modules et quelques changements de paramètres.

Ce CMS est suffisamment ouvert, puissant et flexible pour permettre de coller aux exigences des webmasters les plus avancés en SEO. Le plus remarquable, c'est qu'avec une bonne connaissance de Drupal, on peut s'en sortir sans « hacker » le coeur (sans modifier le code du CMS).

Par contre, la courbe d'apprentissage de Drupal est plus longue que pour un Wordpress ou un Joomla !. C'est un CMS complexe, atypique par certains aspects, et qui fait moins de choix à la place du webmaster.

Cette flexibilité à un impact fort sur les performances des sites faits avec Drupal d'un point de vue SEO : on peut réaliser un site parfaitement optimisé, mais aussi un site lent, mal fichu, générant un code HTML inadapté !

Drupal est donc un CMS parfait pour un webmaster qui cherche un outil puissant et flexible, et qui est soucieux de son référencement. Par contre, il faut le déconseiller aux webmasters qui cherchent des solutions toutes faites : il faut au moins savoir régler, maintenir et préparer sa machine, comme le ferait un pilote d'avion. Et connaître tous les détails du tableau de bord...

Pour terminer, voici l'adresse de la principale porte d'entrée vers cette communauté : <http://drupalfr.org/>

Bon travail et bon référencement !

Philippe YONNET, *Directeur SEO international, Twenga.*

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://blog-abonnes.abondance.com/2011/11/le-referencement-sur-le-cms-drupal-2eme.html>