

Optimiser le référencement d'un site Joomla! : ce que change la version 1.6 du CMS en termes de SEO (1ère partie)

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Joomla! est un CMS utilisé pour de nombreux sites web, mais qui n'était jusqu'à maintenant, de façon native, que peu configuré par défaut pour les moteurs de recherche et le SEO. Cependant la version 1.6 de l'outil corrige plusieurs points importants et rendra la tâche des webmasters plus aisée pour optimiser leur site. Certains points, en revanche, demanderont encore l'utilisation de plugins spécialisés ou de "hacks" parfois complexes. Cet article, en trois parties, tente d'explorer les méandres du CMS Joomla! et du référencement naturel pour vous aider dans vos configurations...

La version 1.6 du CMS Joomla! (CMS = *Content Management System* : Système de gestion de Contenu en français) est sortie il y a quelques semaines, et elle intègre de nouvelles fonctionnalités destinées à améliorer le référencement des sites Joomla!...

Ce CMS, très populaire chez les webmasters en raison de sa simplicité d'installation ou d'utilisation, n'avait pas jusqu'ici une très bonne réputation chez les référenceurs. La liste des défauts que présente une installation "par défaut" de Joomla! est en effet assez longue, et l'ajout de *plugins* non optimisés pour le référencement peut souvent compliquer les choses.

Néanmoins, il est assez facile de corriger la plupart des défauts qui empoisonnent parfois la vie des référenceurs. La plupart des fonctionnalités manquantes peuvent être ajoutées sans effort grâce à l'installation de plugins complémentaires adaptés, et ceux qui veulent une plateforme parfaitement compatible avec les contraintes SEO peuvent l'obtenir à l'aide de "hacks" assez simples, pour qui connaît un peu le langage PHP.

Cet article est le premier d'une série de 3, consacrés à l'optimisation des sites réalisés à l'aide de Joomla! (voir à la fin de cet article ce qui vous attend les mois prochains...).

Avertissement : sauf mention contraire, les conseils qui suivent s'appliquent aux version 1.5x de Joomla! Les changements apportés par la version 1.6 sont indiqués de manière explicite.

Corriger les défauts d'une installation "out of the box"

Le robots.txt

Dans Joomla 1.5, le fichier robots.txt ressemble à ceci :

```
User-agent: *
Disallow: /administrator/
Disallow: /cache/
Disallow: /components/
Disallow: /images/
Disallow: /includes/
Disallow: /installation/
Disallow: /language/
Disallow: /libraries/
Disallow: /media/
```



```
Disallow: /modules/  
Disallow: /plugins/  
Disallow: /templates/  
Disallow: /tmp/  
Disallow: /xmlrpc/
```

Bizarrement, le répertoire "images" est interdit, alors que le plus souvent, un webmaster voudra laisser Google indexer ses images. On peut donc modifier simplement le robots.txt comme ceci :

```
User-agent: *  
Disallow: /administrator/  
Disallow: /cache/  
Disallow: /components/  
Disallow: /images/  
Disallow: /includes/  
Disallow: /installation/  
Disallow: /language/  
Disallow: /libraries/  
Disallow: /media/  
Disallow: /modules/  
Disallow: /plugins/  
Disallow: /templates/  
Disallow: /tmp/  
Disallow: /xmlrpc/
```

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

(si vous voulez autoriser tous les moteurs à indexer les images, il vous suffit de supprimer la ligne *Disallow: /images/*).

Ajouter des pages de la zone publique dans le robots.txt est souvent inutile et parfois même néfaste.

Attention : nous déconseillons formellement d'ajouter des lignes dans le robots.txt pour empêcher le crawl de pages appartenant à la zone "publique" d'un site Joomla!. Si vous ne souhaitez pas qu'une page soit indexée, il faut utiliser la balise meta "robots" avec l'attribut *noindex*, plutôt qu'une ligne dans un fichier robots.txt. Attention, cette balise n'est pas accessible pour toutes les pages dans Joomla 1.5!, il faut donc se servir d'un plugin SEO pour avoir accès à ce réglage.

Exception : il faut éliminer les zones de type "pièges à robots"

Il existe une exception à cette règle : les rubriques de type "spider trap". Par exemple, la plupart des plugins ajoutant des fonctions "calendrier" dans Joomla! créent un nombre infini d'URL. Il s'agit notamment des composants *Events Calendar* et *ExtCalendar* (le calendrier propose des mois pour l'année 2011,2012,2143,3487,11489...).

Dans ce cas-là, il faut ajouter une ligne *disallow* ressemblant à ceci :

```
Disallow: /nom_du_composant/
```

Si l'on utilise la réécriture d'URL, il est important dans ce cas de s'assurer que les URL de ce composant commencent bien par une syntaxe permettant d'isoler les URL du plugin incriminé.

Cas particulier : les pages "pdf" et "imprimables"

Il est possible, par un réglage de l'admin, de générer automatiquement une version PDF et imprimable de la page. Evidemment, ces versions sont des "pages dupliquées".

Une solution radicale, souvent conseillée mais loin d'être parfaite, est d'ajouter une ligne *Disallow: /index2.php* dans le robots.txt. Mais on perd du pagerank interne en utilisant cette

méthode, et les URL en [index2.php](#) peuvent quand même se retrouver dans l'index (sans titre et sans description). Une autre solution consiste à hacker le template en plaçant un *nofollow* sur le lien conduisant à la version imprimable ou la version pdf. Mais là aussi on perd du jus...

En fait, la version PDF de la page est certes aussi une forme du duplicate content, mais comme elle n'est pas particulièrement gênante, on peut laisser Google crawler et indexer les PDF.

Google ayant fortement évolué, la meilleure solution aujourd'hui consiste à laisser le moteur crawler librement ces versions. Si l'indexation de la version imprimable vous gêne, le mieux est d'utiliser la balise meta robots avec l'attribut "noindex, follow". C'est le meilleur compromis (qui nécessite toutefois un hack ou l'utilisation d'un plugin SEO).

Notons que les liens vers la page "envoyer à un ami" posent un problème analogue, qui se résout de la même façon.

Les URL "search engine friendly"

Par défaut, les URL produites par Joomla! ressemblent à ça :

http://www.mondomaine.com/index.php?option=com_content&view=article&id=6&Itemid=6

Vu le nombre de paramètres et le caractère obscur de certains d'entre eux, certains moteurs auront clairement du mal à explorer efficacement un site avec de telles URL. Google, arrive néanmoins à se débrouiller avec ces syntaxes sans trop de casse, mais il est préférable d'optimiser tout cela, et de passer à une syntaxe qui contient un peu plus de mots clés dans l'URL.

Si on utilise les réglages par défaut de Joomla!, on aboutit à ce genre de syntaxes :

</jardinage/8-jardins-paysages/7-patios-d-exterieur.html>

On notera la présence d'un ID numérique identifiant l'article ou la catégorie.

Les syntaxes réécrites par défaut dans Joomla! cumulent trois défauts :

1. **La syntaxe réécrite et non réécrite coexistent sur le site**, et renvoient des codes 200 (pas de redirection). Il est donc assez facile, pourvu que les deux syntaxes viennent à être connues de Google, de se retrouver avec des pages accessibles avec deux URL.
2. Si une URL est mal recopiée, cette nouvelle syntaxe risque de **renvoyer une page avec un code 200**. En effet, le processeur d'URL de Joomla! se sert de l'ID numérique pour renvoyer sa page, le fait de modifier l'orthographe des mots clés situés à côté dans l'URL ne change rien.
3. Les possibilités de personnalisation des URL réalisées avec les options par défaut sont assez limitées.

Bref, les SEF (*Search Engine Friendly*) URL de Joomla! créent des cas de **DUST** ("*Different URL Same Text*"), et sont trop stéréotypées pour s'adapter à tous les besoins.

Heureusement, tout cela peut-être corrigé et sérieusement amélioré par l'utilisation du plugin **sh404SEF**, dont l'utilisation et le paramétrage sera décrit dans le 3e article de la série.

Conseil : les webmasters activent très souvent les options de réécriture d'URL dès l'installation du CMS. C'est souvent trop tôt. Il est certes particulièrement conseillé de changer la syntaxe des URL AVANT la mise en ligne du site, car tout changement dans ce domaine peut avoir des conséquences funestes sur le référencement, surtout en l'absence de redirections bien faites. Par contre, il est intéressant d'attendre d'avoir parfaitement stabilisé la structure de son site avant de choisir la syntaxe de ses URL. Par "structure" on entend ici notamment la

liste des différents types de pages présents dans le site, le système de menus et l'arborescence des catégories.
Donc les SEF URL se paramètrent AVANT la mise en ligne du site, mais APRES QUE l'organisation de son contenu ait été parfaitement définie.

Les redirections

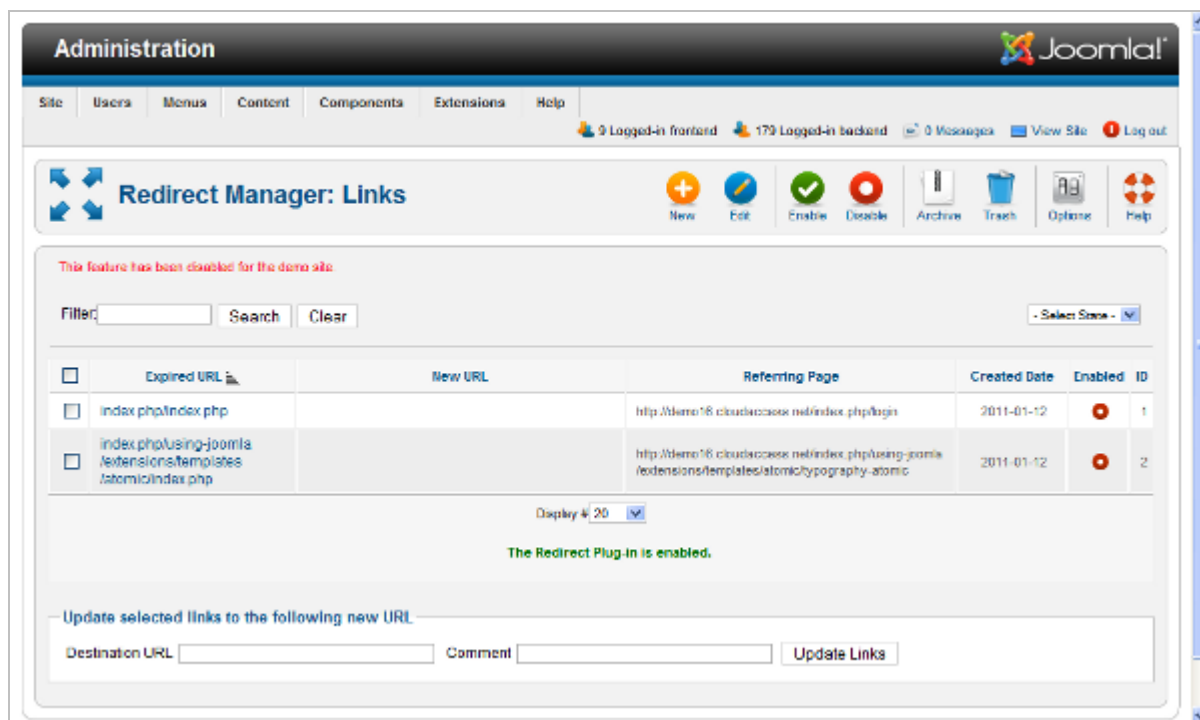
La gestion des redirections et des liens brisés est un peu indigente dans Joomla!.

Rediriger la syntaxe non réécrite vers la syntaxe réécrite par exemple n'est pas possible par des réglages dans l'admin, et typiquement un changement de titre produira des erreurs 404 lorsque les URL sont réécrites.

Pour des changements de syntaxe "globaux", les plus techniques d'entre vous peuvent se lancer dans une redirection gérée par le fichier *htaccess*. Mais une gestion URL par URL n'est pas recommandée avec cette méthode.

Ces limites sont heureusement parfaitement corrigées par le plugin **sh404sef**.

Dans la version 1.6, une gestion des redirections fait son apparition (enfin !), mais ne remplace pas complètement la gestion complète assurée par sh404sef.



Le nouveau gestionnaire de redirections de Joomla ! 1.6

La balise title

Par défaut, la balise title sur un article sera de cette forme :

"nom du site - le titre de l'article qui peut parfois être interminable avec les mots clés importants à la fin"

Pour optimiser le titre, beaucoup de référenceurs préfèrent mettre le nom du site à la fin, voire ne pas l'afficher, et limiter la longueur du titre.

Pour changer cela, on peut recourir à un hack (changer le code source) :

`/includes/Joomla.php` vers la ligne 508

en remplaçant:

```
$this->_head['title'] = $title?  
$GLOBALS['mosConfig_sitename'] . ' - ' . $title :  
$GLOBALS['mosConfig_sitename'];
```

par sa propre version...

Dans Joomla! 1.6, le problème a en partie disparu dans cette version : on peut choisir entre différentes versions de la balise title, avec ou sans le nom du site.

Par contre, si on veut remplacer le système de titre par défaut par une règle semi-automatique, un hack est hélas toujours nécessaire.

Les autres "metas"

Un certain nombre de principes adoptés dans Joomla! 1.5 pour simplifier le travail des webmasters ont des conséquences gênantes pour le référencement : beaucoup de metas utiles pour le référencement sont génériques : une seule version de la balise est disponible, reproduite sur tout le site. On peut certes passer outre et définir des metas personnalisées pour chaque article, mais les balises par défaut définies dans le menu "site" sont recopiées dans certaines pages essentielles comme les pages "catégories" ou les pages "sections".

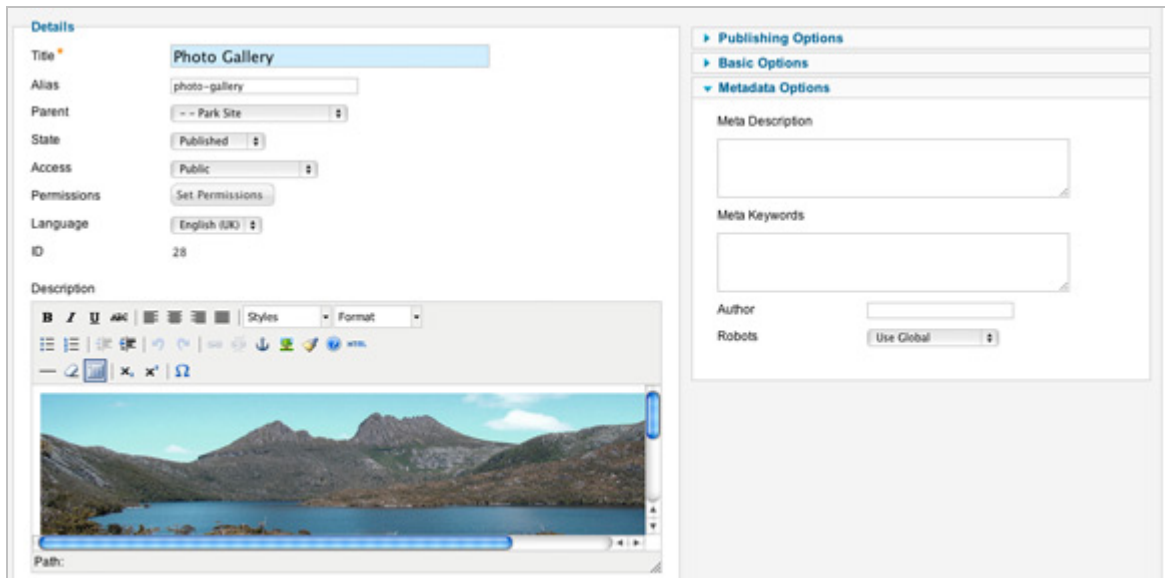
La solution passait par des hacks ou l'utilisation de plugins, mais le problème est complètement résolu dans la version 1.6.

Les meta tags dans la version 1.6 peuvent être paramétrés à différents niveaux :

- Comme dans la version 1.5, ils peuvent être définis dans la configuration globale, ou dans les paramètres de chaque article. Les paramètres définis pour un article prévalent sur les paramètres définis pour l'ensemble du site.
- Dans la version 1.6, les meta-tags peuvent aussi être paramétrés au niveau des éléments de menu, ou des catégories. Des champs pour les balises meta ont fait également leur apparition dans les catégories liées aux composants (Les composants *banner ad*, *contact*, *newsfeed*, et *weblinks* par exemple).
- Dans cette nouvelle version, les balises meta n'ont pas de valeur par défaut. Cela signifie que lors de l'installation du CMS, aucune balise meta n'est définie. Même pas la fameuse description "*Joomla - the dynamic portal engine and content management system*", que l'on retrouve sur beaucoup de pages de sites faits avec ce CMS.

Dans la version 1.6, une hiérarchie de priorités apparaît pour la prise en compte des balises définies pour différents éléments du site. Cette hiérarchie suit les règles habituelles pour les paramètres Joomla.

- Si aucune balise meta n'est définie, à quelque niveau que ce soit, les paramètres globaux sont utilisés.
- Les balises définies dans les menus ont la priorité sur les paramètres globaux.
- Les balises définies pour les éléments "catégories" ont la priorité sur les paramètres de menu.
- Les balises définies pour un article donné ont priorité sur les paramètres définis pour la catégorie dont dépend l'article.



Dans l'admin de Joomla ! 1.6 apparait pour la première fois des champs pour paramétrer les metas dans les catégories. Un champ pour la meta robots a également été ajouté.

Même si ces changements corrigent un défaut majeur de Joomla!, ce système de metas ne fonctionne pas avec des règles réutilisant des variables. Créer des balises différentes pour chacune des pages (ce qui est chaudement recommandé) nécessite donc d'entrer toutes les données à la main ! Pour un gros site, cela devient vite une contrainte intenable ! La seule solution consiste dans ce cas à "hacker" le code et à remplacer la fonction qui génère les balises avec une fonction de son choix. C'est radical, mais cela signifie toucher au coeur du CMS, ce qui n'est jamais une bonne idée sur le long terme...

Quelques ajouts utiles dans le fichier .htaccess

Redirection 301 de /home.html vers l'URL canonique / (http://www.domaine.com/home.html => http://www.domaine.com) :

```
Redirect 301 /home.html http://www.example.com
```

Redirection 301 de /index.php :

```
RewriteCond %{THE_REQUEST} ^[A-Z]{3,9}\ /index\.php\ HTTP/  
RewriteRule ^index\.php$ http://www.example.com/ [R=301,L]
```

Redirection 301 des URL sans www vers la version avec www :

```
RewriteCond %{HTTP_HOST} ^example\.com$ [NC]  
RewriteRule ^(.*)$ http://www.example.com/$1 [R=301,L]
```

Attention : ceci doit être adapté à la configuration de votre serveur. Par exemple il sera souvent nécessaire, pour éviter d'avoir des erreurs 500 avec ces ajouts dans le fichier .htaccess, de rajouter une directive :

```
RewriteBase /nomdussrepertoire
```

(si votre installation n'est pas dans le répertoire par défaut)

et de désactiver "Options +FollowSymLinks" en ajoutant un # au début de la ligne.

Dans ce premier article, nous avons listé les principaux défauts d'une installation de Joomla! "out of the box" et indiqué comment les corriger. Le second article sera consacré à l'optimisation des templates et à la résolution des problèmes de performance, et de structure

(partiellement résolu avec la version 1.6). Le dernier article sera consacré à l'utilisation du plugin sh404Sef, un module additionnel qui dote Joomla! de 90% des fonctionnalités avancées pour le SEO qui manquent au CMS. A très bientôt !

Philippe Yonnet, *Global SEO Strategist, WEB DMUK (Londres) – Easyroommate / Vivastreet*

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://blog-abonnes.abondance.com/2011/02/optimiser-le-referencement-dun-site.html>