

## Référencement : La règle des "3C" (1ère partie)

[Retour au sommaire de la lettre](#)

*Le référencement d'un site web est avant tout tributaire de la façon dont le site en question a été conçu. Dans ce cadre, une stratégie gagnante prendra en compte divers aspects de cette conception, dans le cadre d'un concept global intitulé "Règle des 3C". Ce mois-ci, nous examinons le premier des "C", soit la Conception : comment, dès le départ, dès la rédaction du cahier des charges, rendre son site "spider friendly" ? Réponse sous la forme d'un "mémo" et d'un "pense-bête" qui vous permettra de ne pas rater une étape importante lors de vos réflexions...*

Pour optimiser au mieux le référencement d'un site, il est devenu important, voire primordial, de le penser dès le départ pour être compatible avec les différents moteurs de recherche qui vont venir le visiter, grâce à leurs *spiders*, robots qui viennent "aspirer" les pages web et suivre les liens qu'elles contiennent.

Pour qu'un site soit parfaitement "compris" et "analysé" par les moteurs de recherche, il faut donc qu'il ait été pensé pour être compatible avec les critères d'exploration et de pertinence de ces outils. Nous allons, dans cette série de trois articles, expliciter une règle qui nous est chère, et que nous avons pu expérimenter sur de nombreux sites, celles des "3C".

Ce mois-ci, le premier "C" exploré sera celui de la **Conception initiale** du site. En d'autres termes, comment imaginer pour son site une structure qui soit totalement compatible avec son exploration par les *spiders* ? Comment mettre en place un réseau d'informations aidant ces robots à obtenir une meilleure compréhension du maillage de votre source d'information ? Comment faire pour éviter tout obstacle technologique freinant ou bloquant pour les moteurs ? Etc.

Bref, nous allons lister, sous la forme d'un "mémo", dans cette première partie, une suite de "bonnes pratiques" qui va vous mener à mettre en ligne un site qui sera 100% "*spider friendly*" dès son lancement... Sachant qu'une fois que ce travail sera fait, il restera 2 "C" à compiler les mois prochains... Le travail ne sera donc pas terminé pour autant mais vous pourrez partir sur de bonnes bases, ce qui est loin d'être négligeable... :-)

### Conception > Structure du site

Le premier des points à inspecter est la structure du site : les robots peuvent-ils aller partout de façon efficace et découvrir toutes vos pages ? Voici quelques points à vérifier :

#### Conception > Structure du site > Fichier Robots.txt

- Votre site contient-il un fichier robots.txt ?
- Son nom est-il bien orthographié ("r" minuscule, "robots" au pluriel) ?
- Est-il disponible à la racine de votre site ([www.votresite.com/robots.txt](http://www.votresite.com/robots.txt)) ?
- En cas d'utilisation de sous-domaine ([motclé.votresite.com](http://motclé.votresite.com)), chaque sous-domaine dispose-t-il de son propre fichier robots.txt ([actu.votresite.com/robots.txt](http://actu.votresite.com/robots.txt), [produits.votresite.com/robots.txt](http://produits.votresite.com/robots.txt)) ?
- Les zones "interdites au robots", si elles existent, sont-elles bien listées dans le(s) fichier(s) robots.txt ?
- Certains spiders moins importants sont-ils pris en compte ou interdits si nécessaire ou si vous avez remarqué dans vos statistiques que leur venue gêne votre serveur (trop de bande passante occupée lors du *crawl* par exemple) ?
- Votre fichier robots.txt indique-t-il l'url de votre fichier "sitemap" (voir plus loin) ? Rappelons la syntaxe (très simple) pour ce faire avec un fichier "robots.txt" basique (ici celui du site <http://www.boutique-abondance.com/robots.txt>) :

```
User-agent: *  
Disallow:  
Sitemap: http://www.boutique-abondance.com/sitemap.xml
```

Avec la ligne "User-agent: \*" on indique qu'on s'adresse à tous les spiders.

Avec "Disallow:", on n'interdit rien.

L'option "Sitemap:" indique enfin à Google, Yahoo!, Exalead, Live Search et Ask.com l'url de votre fichier Sitemap (voir <http://www.abondance.com/actu/2007-15/sitemaps.php>).

- Avez-vous vérifié la syntaxe de votre fichier robots.txt grâce à un outil disponible en ligne ?

Google en propose un dans sa zone "Webmaster Tools" mais il en existe d'autres :

<http://tool.motoricerca.info/robots-checker.phtml>

<http://www.sxw.org.uk/computing/robots/check.html>

<http://www.searchenginepromotionhelp.com/m/robots-text-tester/robots-checker.php>

Etc.

Enfin, pour en savoir plus sur la syntaxe du fichier "robots.txt", consulter :

<http://www.abondance.com/docs/robots.html>

### **Conception > Structure du site > Fichier Sitemaps**

Les cinq moteurs de recherche principaux se sont mis d'accord sur un format de fichier, nommé "Sitemap", décrit à l'adresse <http://www.sitemaps.org/>, et qui fournit aux moteurs un "plan du site", aux formats TXT ou, mieux XML, indiquant aux *spiders*, pour chaque page de votre site :

- Son url.
- Sa priorité d'indexation (0 à 1).
- Sa fréquence de mise à jour (de "always" à "never" en passant par "hourly", "weekly", etc.).
- Sa date de dernière modification.

Ce fichier est aujourd'hui indispensable pour obtenir une bonne "indexation" quantitative de votre site. Il n'est, en revanche, pas une garantie de positionnement... Mais le fait d'avoir sur son site un fichier Sitemaps bien fait est, de façon certaine, un plus pour un meilleur "référencement" (au sens de l'indexation) de vos pages. Notons également que la présence d'un fichier Sitemaps ne vous dispense pas de créer un site web qui soit "*spider friendly*" au niveau des liens qu'il propose... Les robots des moteurs doivent également pouvoir découvrir le plus de pages possible de votre site en suivant les liens de vos pages (notamment grâce à votre plan du site pour les internautes, au format HTML, voir plus loin)...

Voici la liste des questions qu'il faut se poser au sujet de vos fichiers Sitemaps :

- Votre site propose-t-il un fichier Sitemaps aux *spiders* ? Comme pour le fichier robots.txt, si vous utilisez des sous-domaines, n'oubliez pas de proposer un fichier Sitemaps par sous-domaine.

- Le fichier Sitemaps recense-t-il de façon exhaustive toutes les pages de votre site ?

- Les indications qu'il fournit pour chacune d'elles (date de dernière modification, fréquence de mise à jour) sont-elles conformes à ce que le robot découvrira de lui-même s'il parcourt votre site en suivant les liens mis à sa disposition ? Il est important que les informations que vous délivrez dans le fichier Sitemaps soient le reflet de la réalité et soient cohérentes par rapport aux voies d'exploration de votre site par les robots...

- Avez-vous mis en place une procédure simple de gestion de votre fichier Sitemaps pour les zones dynamiques, dont les informations changent souvent (nouvelles pages, modifications des pages existantes) ? Le but sera de fournir aux moteurs les données les plus fraîches possible...

- Avez-vous indiqué l'adresse du Sitemaps dans le fichier robots.txt (voir paragraphe précédent) ?

- Avez-vous soumis votre fichier Sitemaps à Google

(<http://www.google.com/webmasters/sitemaps/>) et Yahoo! (<http://siteexplorer.search.yahoo.com/>)

au travers de leurs outils pour Webmasters ? L'outil de Google, notamment, vous fournira des statistiques de visites ainsi que des diagnostics d'erreur éventuelles sur vos fichiers, bien utiles parfois...

### **Conception > Navigation**

Lors de sa venue, votre site doit proposer une opération "portes ouvertes" aux robots des moteurs de recherche (pour les zones auxquelles ils ont accès, bien sûr...). Aussi, il est important de répondre aux questions suivantes :

- Vos liens seront-ils tous compris et suivis par les *spiders* (sous la forme `<A href="http://www.votresite.com/...">`) même s'ils sont écrits en langage JavaScript ?  
Voir l'article : *Comment rendre des liens JavaScript compatibles avec les spiders des moteurs ?* (<http://abonnes.abondance.com/archives/2004-10/>).
- Chaque page de votre site est-elle accessible en 3 clics maximum à partir de la page d'accueil, ce qui garantira une meilleure prise en compte quantitative de vos pages ?
- Les liens à l'intérieur du contenu éditorial sont-ils facilement configurables lors de la saisie du contenu (texte du lien, url de destination, saisie éventuelle d'un attribut "title", etc.) ?
- Un plan du site proposant des liens textuels simples est-il disponible ? Permet-il aux spiders d'accéder à chaque page de votre site en trois clics au plus ?
- Avez-vous prévu des pages d'erreur en cas de renvoi de code 404, 403, etc. ? Ces pages contiennent-elles des liens vers les différentes zones de votre site, autant de points d'entrée pour l'internaute mais également pour les spiders ?

### **Conception > Obstacles technologiques éventuels**

- Si votre site propose des animations Flash, avez-vous fait le nécessaire (zone "noembed", pages HTML complémentaires, etc.) pour le référencer au mieux ?
- Si votre site contient du JavaScript, l'avez-vous rendu compatible avec les moteurs de recherche, notamment au niveau des liens affichés (voir précédemment) ?
- Si votre site est réalisé en frames (ce qui peut paraître étrange car rares sont aujourd'hui les sites web qui utilisent encore cette technologie), avez-vous fait ce qu'il fallait pour pallier les inconvénients de ce type de système ?

Bref, si votre site présente des "obstacles technologiques" (cookies, identifiants de session, etc.) au sens des moteurs de recherche, n'oubliez pas de lire et relire cet article :  
*Comment rendre visible sur les moteurs un "site à problèmes" ?*  
(<http://abonnes.abondance.com/archives/2006-07/>)

### **Conception > Intitulés des urls**

- Si vos urls contiennent des "?" et des "&", avez-vous mis en place des techniques de réécriture d'urls ?
- Vos urls proposent-elles des mots clés descriptif du contenu des pages ([www.votre-site.fr/epicerie/condiments/exotiques/sel-guyane.html](http://www.votre-site.fr/epicerie/condiments/exotiques/sel-guyane.html)) ? Et notamment le nom de domaine et le nom de fichier (premier et dernier intitulé de l'url) ?
- Si vous désirez que vos pages soient indexées par Google News (<http://news.google.fr/>), vos intitulés d'urls contiennent-ils 3 chiffres au minimum, condition demandée par Google pour leur indexation dans son moteur d'actualités ?
- Les redirections éventuelles d'une page vers une autre sont-elles réalisées à l'aide de codes 301 ?
- Votre site utilise-t-il des identifiants de sessions qui apparaissent dans l'url ? Ceci peut poser de nombreux problèmes à certains moteurs. Voir précédemment dans cet article.

### **Conception > Pages web**

Enfin, vos pages web, dans leur structure, sont-elles compatibles avec les moteurs ?

- Par exemple, ont-elles été conçues pour être monolingue ? Une langue donnée doit être affichée majoritairement dans une page web donnée pour être "compréhensible" par les robots.

- De même, les pages doivent être conçues pour être le plus "monothème" possible. Peu importe leur longueur (dans ce cas, c'est plus la densité des mots clés qui jouera pour leur pertinence), mais il est important que chaque page cerne un point donné, quitte à multiplier les pages sur votre site. En d'autres termes, si vous avez 10 produits à présenter, n'hésitez pas une seconde : créez une page de présentation pour chaque produit et surtout pas une longue page qui les présente tous sur un même document. Vous devez être capable de décrire le contenu d'une page en 10 mots au maximum (ce sont les termes qui constitueront son titre).

### **Conception > Rubriquage**

- Avez-vous pensé les différentes rubriques du site en fonction de ce que les internautes recherchent et pas en fonction de la base de données (comme cela arrive parfois) ?

- N'oubliez pas de donner des noms explicites à chaque zone thématique de votre site. Ces noms explicites seront ensuite repris dans les intitulés de liens, primordiaux aujourd'hui pour un référencement...

### **Conclusion**

Tous les points visés dans cet article touchent à la conception du site web. Ils sont capitaux car ils représentent des étapes essentielles, fondamentales, sur lesquelles ils sera très difficile, voire impossible de revenir en arrière lorsque le site sera en ligne. Et Dieu sait si, par le passé, ce type d'erreur a été commis, nécessitant parfois une remise en question complète du site pour le rendre compatible avec les moteurs. Autant penser à ces points dès le départ, cela facilite bien les choses et permet d'éviter des surprises désagréables...

Avez-vous suivi, pas à pas, les différentes étapes que nous avons listées dans cet article, et ce dès la conception initiale de votre site ? Oui ? Dans ce cas, vous ne devriez pas avoir pas de soucis majeurs pour la suite des opérations et les deux prochains "C" à explorer... Rendez-vous le mois prochain pour la suite de cette série de trois articles...