

Regardez vos sites avec l'oeil du spider...

[Retour au sommaire de la lettre](#)

Les spiders, ou robots, des moteurs de recherche, ont une façon bien spécifique de voir vos pages web, très différente de celle que peuvent avoir les internautes humains. Voici deux astuces qui vous permettront d'adopter cette vision et de mieux comprendre comment un moteur "visualise" une page web, autant d'informations nécessaires pour mieux optimiser vos données...

Lorsque vous allez consulter un site web, vous avez bien sûr pour ces pages l'œil de l'internaute "humain" qui regarde l'écran de son ordinateur. Mais les spiders des moteurs de recherche ont, pour leur part, une vision toute autre de vos pages. Pour en savoir plus, voici deux façons astucieuses de vous mettre à la place d'un robot et de visualiser votre site sous un œil nouveau et... parfois assez surprenant...

Comment les spiders voient-ils vos pages ?

Tout d'abord, il est important de bien comprendre que les robots des moteurs n'ont qu'une vision parcellaire de vos documents. On les compare souvent, avec raison, à un utilisateur aveugle qui utiliserait un système de reconnaissance vocale pour comprendre ce que contiennent vos pages. Aussi :

- Ils lisent vos textes sans soucis.
- Ils ne savent pas reconnaître vos images (mais le texte alternatif - option "alt" des balises images - si).
- Le format Flash leur pose le plus souvent de sérieux problèmes pour en extraire le texte qui y est inséré.
- Les liens JavaScript, s'ils n'ont pas été "conçus pour", risquent de ne pas être suivis.

Bref, pour les spiders, l'essentiel est : le texte, le texte, le texte !

Le cache de Google

La première façon de "vivre dans la peau d'un spider" et de visualiser ce si important contenu textuel est d'utiliser le cache de Google. Pour ce faire, recherchez votre site sur le moteur et visualisez le résultat obtenu, comme nous l'avons fait ici avec le site Abondance.com sur Google :

Abondance : référencement et moteurs de recherche - toute l'info ...
Abondance d'infos sur le référencement et les moteurs de recherche : description des moteurs, actualité, faqs, outils d'audit, méthodologie de référencement ...
www.abondance.com/ - 59k - [En cache](#) - [Pages similaires](#) - [À noter](#)

A ce stade, cliquez sur le lien "En cache" proposé à droite de l'url. Vous obtenez cette page :

Voici la version **G o o g l e** de la page mise en cache de <http://www.abondance.com/> extraite le 8 nov 2007 09:11:58 GMT. La version « En cache » proposée par **G o o g l e** correspond à la page telle qu'elle se présentait lors de la dernière consultation effectuée par Google. Il se peut que la page ait été modifiée depuis cette date. Cliquez ici pour consulter [la page actuelle](#) (sans mises en valeur). Cette page mise en cache peut renvoyer à des images qui ne sont plus disponibles. Cliquez ici pour obtenir uniquement le [texte mis en cache](#). Pour créer un lien avec cette page ou l'inclure dans vos favoris/signets, utilisez l'adresse suivante : <http://www.google.com/search?q=cache:ICck6ZHqVgJ:www.abondance.com/+abondance&hl=fr&ct=clnk&cd=1&gl=fr>.

Google n'est ni affilié aux auteurs de cette page ni responsable de son contenu.

Les termes de recherche suivants ont été mis en valeur : **abondance**

Location / Montalbert
 Été: Jusqu'à -30% Hiver 2007/8 déjà en ligne!

Chalet morzine
 Acheter, Louer, Vendre un chalet ?
 L'Immobilier c'est sur Explorimmo

Abondance

Toute l'actualité et l'information sur le référencement et les moteurs de recherche

Vendredi 9/11/2007

Vous débutez ?	Actualité	Blog	L'actu sur votre site	Dossiers / Articles	Méthodologies	Outils de recherche
Audits	Forums / Chat	Lettres d'information	Études	FAQ	Livres	Emploi
Tribune	Humour	Liens	Offre commerciale	Boutique en ligne	Zone Abonnés	Réseau Abondance

A la une de l'actualité

Google paye pour utiliser un algorithme de clustering
 Google vient d'acheter pour 2 millions de \$ à la société américaine Invenda (anciennement E-centives) une licence pour utiliser un brevet intitulé "Search Results Using Dynamic Categorization" qui semblerait décrire un processus de clusterisation... [Mercredi 7 novembre, 9h15]

Découvrez [toute l'actu de la semaine, de 2007...](#) et aussi :

[08/11] [La Google NewsBar pour intégrer Google News sur votre site](#)
 [08/11] [Nouvelles fonctions orientées "référencement" pour l'outil de mesure d'audience 1DStat](#)

Rechercher sur Abondance :

Search

Tout **Abondance** Toute l'actu depuis 1998

Les partenaires Abondance

Vous cherchez un **prestataire de référencement** ?

Consultez la liste des **partenaires Abondance^{Nouveaux}**, tous signataires de la **charte de qualité sur le référencement**

En haut de la page, dans la zone textuelle proposée par Google, vous trouverez cette phrase :

Cliquez ici pour obtenir uniquement le [texte mis en cache](#).

Cliquez sur le lien ainsi proposé. Vous obtenez alors une vision "spider friendly" de votre page, avec uniquement le texte affiché, donc ce que voient les spiders des moteurs :

Voici la version **G o o g l e** en texte seul de la page mise en cache de <http://www.abondance.com/> extraite le 8 nov 2007 09:11:58 GMT. La version « En cache » proposée par **G o o g l e** correspond à la page telle qu'elle se présentait lors de la dernière consultation effectuée par Google. Il se peut que la page ait été modifiée depuis cette date. Cliquez ici pour consulter [la page actuelle](#) (sans mises en valeur). Cliquez ici pour obtenir la [totalité de la page mise en cache](#), avec les images. Pour créer un lien avec cette page ou l'inclure dans vos favoris/signets, utilisez l'adresse suivante : <http://www.google.com/search?q=cache:ICck6ZHqVgJ:www.abondance.com/+abondance&hl=fr&gl=fr&strip=1>.

Google n'est ni affilié aux auteurs de cette page ni responsable de son contenu.

Les termes de recherche suivants ont été mis en valeur : **abondance**

[Abondance.com : Recherche d'information, référencement et promotion de sites Web](#)

Toute l'actualité et l'information sur le référencement et les moteurs de recherche

Vous débutez ?	Actualité	Blog	L'actu sur votre site	Dossiers / Articles	Méthodologies	Outils de recherche
Audits	Forums / Chat	Lettres d'information	Études	FAQ	Livres	Emploi
Tribune	Humour	Liens	Offre commerciale	Boutique en ligne	Zone Abonnés	Réseau Abondance

A la une de l'actualité

Google paye pour utiliser un algorithme de clustering
 Google vient d'acheter pour 2 millions de \$ à la société américaine Invenda (anciennement E-centives) une licence pour utiliser un brevet intitulé "Search Results Using Dynamic Categorization" qui semblerait décrire un processus de clusterisation... [Mercredi 7 novembre, 9h15]
<http://www.abondance.com/rss/rss.xml> Découvrez [toute l'actu de la semaine, de 2007...](#) et aussi :

Rechercher sur **Abondance** :

Tout **Abondance** Toute l'actu depuis 1998

Les partenaires **Abondance**

Vous cherchez un **prestataire de référencement** ?

Ce n'est clairement pas la même chose... Faites le test avec votre site et vous risquez peut-être d'être surpris... Notez également que les feuilles de syle (CSS) ne sont plus appliquées non plus, ce qui peut tout changer...

Notez, à ce sujet, que ni Yahoo! ni Live Search, ni Ask.com, autres outils proposant un système de cache en ligne, ne proposent cette version textuelle de vos pages.

Utiliser un simulateur de spider

Le petit inconvénient du fait d'utiliser le cache de Google est que ce dernier affichage, s'il donne un excellent aperçu des données prises en compte par le moteur, respecte la majeure partie des informations de mise en page, notamment pour ce qui est des tableaux et, plus globalement, de la façon donc les "blocs de texte" sont agencés dans la page web (si ces informations ne sont pas présentes dans les CSS). Or, un spider a le plus souvent une vision beaucoup plus "linéaire" des données : il lit les codes HTML de haut en bas sans réellement tenir compte de la mise en page proposée dans la fenêtre d'un navigateur...

Alors, il est possible, pour obtenir une vision encore plus réaliste de la façon dont un moteur lit vos documents, d'utiliser un "simulateur de spider" comme Spider Simulator (<http://www.spider-simulator.com/>), du réseau Abondance, qui donnera alors, dans sa page de résultats, cette vision linéaire plus proche de celle des robots :

Contenu textuel de la page ⓘ

Toute l'actualité et l'information sur le référencement et les moteurs de recherche Vous débutez ? Actualité Blog L'actu sur votre site Dossiers / Articles Méthodologies Outils de recherche Audits Forums / Chat Lettres d'Information Études FAQ Livres Emploi Tribune Humour Liens Offre commerciale Boutique en ligne Zone Abonnés Réseau Abondance A la une de l'actualité Google : Vous ne viendrez plus chez nous par hasard ...
Le fabricant américain de pompes à essence Gilbarco Veeder-Root a annoncé qu'il allait équiper certaines de ses pompes d'écran tactile et du logiciel Applause media system permettant d'afficher les cartes routières de Google Maps, permettant ainsi de consulter une liste de commerces locaux... [Vendredi 9 novembre, 10h12]
Découvrez toute l'actu de la semaine, de 2007... et aussi :
[09/11] Google : Vous ne viendrez plus chez nous par hasard ... [08/11] La Google NewsBar pour intégrer Google News sur votre site [08/11] Nouvelles fonctions orientées référencement pour l'outil de mesure d'audience IDStat [08/11] Looksmart se sépare du site FindArticles [08/11] AOL rachète la régie publicitaire Quigo [07/11] Google paye pour utiliser un algorithme de clustering [07/11] Google Sketchup est disponible en version française [07/11] Microsoft veut rivaliser avec Google Mini [06/11] Chine : 75% de parts de marché pour Baldu selon le CNNIC [06/11] Yahoo! lance Kickstart, réseau social pour étudiant et prépare des API de géolocalisation A la une de la lettre Recherche & Référencement d'octobre 2007 Les générateurs de mots clés : comparatif des outils de Google, Yahoo! Search Marketing, Microsoft adCenter et Miva
La plupart des régies publicitaires de liens sponsorisés proposent aujourd'hui des outils baptisés générateurs de mots clés , très utiles pour savoir si une requête est souvent tapée - ou non - sur les moteurs de recherche. Voici un comparatif complet de cinq outils en ligne proposés par les quatre plus importantes régies (Yahoo! en proposant deux, ancienne et nouvelle version intégrée à sa plate-forme Panama)...
Autres articles récents de la lettre Recherche & Référencement :
[Octobre 2007] Référencement : La règle des 3C (3ème partie) [Octobre 2007] La commande Doctype et le référencement [Octobre 2007] Lettres accentuées et référencement (2ème partie) [Octobre 2007] Les nouveautés de Microsoft Live Search [Octobre 2007] Le spamdexing : approche juridique du problème [Octobre 2007] Les générateurs de mots clés : comparatif des outils de Google, Yahoo! Search Marketing, Microsoft adCenter et Miva [Octobre 2007] Le marché de la recherche d'information et des moteurs de recherche dans le monde [Octobre 2007] SerpArchive : un nouvel outil de veille innovant [Septembre 2007] Longue Traîne et Référencement : un concept aujourd'hui incontournable [Septembre 2007] Google traite-t-il réellement le spamdexing dans ses résultats ? [Septembre 2007] Méthodologie de recherche des mots clés de référencement [Septembre 2007] Recherche d'information : revue d'effectif des nouveaux moteurs de 2007 Sur le blog d'Abondance Les premières copies d'écran de l'outil pour webmasters de Live Search
Live Search, le moteur de recherche de Microsoft, devrait proposer la semaine prochaine, si tout va bien, un outil pour webmasters comparable aux Webmaster Tools de Google et Site Explorer de Yahoo!. Le site Search Engine Roundtable a publié cette semaine les premières copies d'écran de cet outil qui est encore actuellement en mode private bêta

Seul inconvénient : le texte est fourni "tel quel" donc assez difficile digérable et facile à lire s'il est... abondant :-). Mais les spiders n'ont-ils pas de gros estomacs ?? ;-)

Toujours est-il que cette vision est certainement la plus proche de celle d'un spider, et notamment en ce qui concerne son ordre de lecture : aussi, dans le cadre de votre référencement, il vous faudra tout particulièrement soigner les 100 premiers mots ainsi détectés : ils doivent parfaitement définir l'univers sémantique de votre page. Est-ce le cas pour votre site web ? Pas si sûr...

Voici quelques outils similaires pour effectuer vos tests :

Spider Simulator (réseau Abondance)

<http://www.spider-simulator.com/>

Search Engine Spider Simulator

<http://www.webconfs.com/search-engine-spider-simulator.php>

SEO Tools - Spider Simulator

<http://www.seoachat.com/seo-tools/spider-simulator/>

SummitTOOLS

<http://tools.summitmedia.co.uk/spider/>

Webmaster Toolkit - Search Engine Spider Simulator

<http://www.webmaster-toolkit.com/search-engine-simulator.shtml>

Testez-les pour trouver celui qui vous convient le mieux. Sachez cependant qu'ils renvoient tous à peu près la même information...

Olivier Andrieu

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

<http://abonnes.abondance.com/blogpro/2007/11/novembre-2007-regardez-vos-sites-avec.html>