

PageRank Sculpting : état des lieux

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Imaginé par Google et ses confrères en 2005 pour lutter contre le spam dans les blogs et les forums, l'attribut 'nofollow' des liens hypertexte a été rapidement utilisé par les référenceurs pour choisir à quelles pages ils allaient transmettre - ou non - de la popularité ou PageRank. Google ayant mis en juin 2009 un frein à ces pratiques, qu'en est-il aujourd'hui et est-il encore possible malgré les filtres du moteur de recherche majeur de travailler ainsi sa popularité en jouant sur les liens internes ?...

La technique du PageRank Sculpting a largement alimenté les discussions des forums cet été, suite à une révélation fracassante de Matt Cutts sur son blog. Est-ce que le PageRank Sculpting est toujours utile ? En quoi cela consiste exactement ? Voici quelques éléments de réponse...

Quand les webmasters deviennent sculpteurs de PR !

La technique de PageRank Sculpting remonte à l'invention de l'attribut 'rel=nofollow' par Google, en janvier 2005 (<http://actu.abondance.com/2005-03/blog-spam.html>). Pour mémoire, cette directive (actuellement non reconnue par la W3C) consiste à donner l'instruction « ne pas suivre ce lien » aux moteurs de recherche, directement à l'intérieur de la balise HTML. Elle s'écrit de la façon suivante :

```
<a href="http://www.monsite.fr" rel="nofollow">Mon site</a>
```

En janvier 2005, Google annonçait sur son blog qu'il prenait désormais en compte l'instruction 'nofollow', dans l'optique de lutter contre le Spam dans les commentaires de blogs en spécifiant ceci :

« A partir de maintenant, quand Google verra l'attribut (rel="nofollow") sur les hyperliens, ces liens n'auront plus aucun crédit quand nous classerons les sites dans nos résultats de recherche » (<http://googleblog.blogspot.com/2005/01/preventing-comment-spam.html>).

En janvier 2007, le portail Wikipédia soulevait une tempête d'interrogations et de critiques lorsqu'il annonçait utiliser désormais le nofollow sur tous les liens sortants, de façon à décourager le référencement d'un site par le biais d'articles plus ou moins pertinents postés dans l'encyclopédie (<http://actu.abondance.com/2007-04/wikipedia-nofollow.php>).

Bien que toujours controversée, cette technique est aujourd'hui entrée dans les mœurs ; de nombreux portails d'information utilisent le nofollow notamment pour leurs commentaires, sans oublier les forums et les blogs.

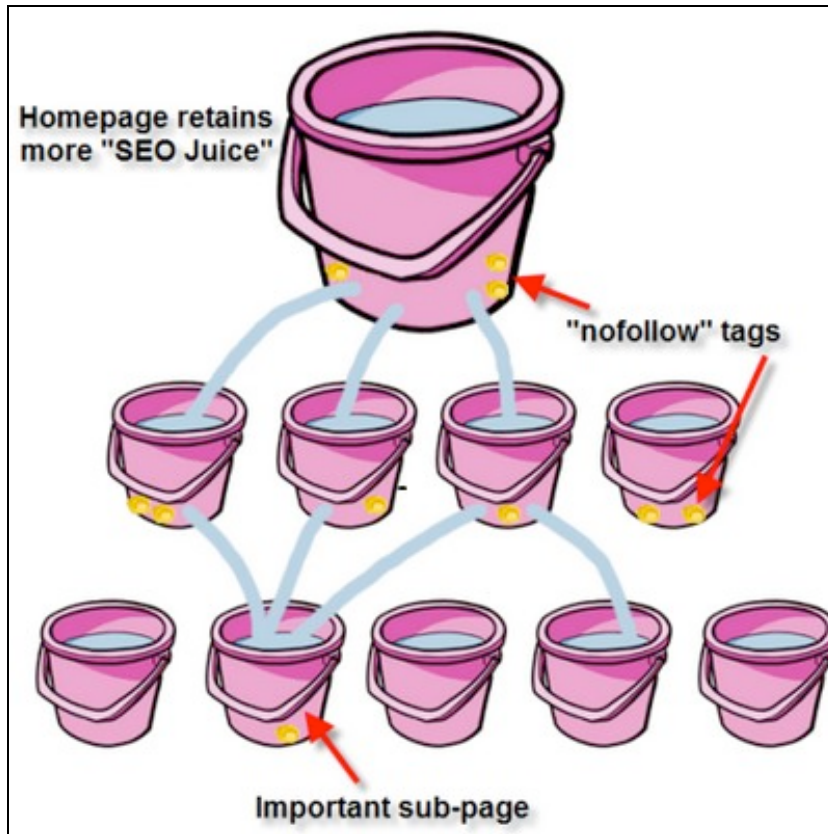
La technique du PageRank Sculpting se sert donc du nofollow et part de cette constatation très simple : si on peut maîtriser la façon dont Google prend en compte les liens sur un site, alors on peut maîtriser le PageRank.

En effet, le PageRank est en partie (mais pas seulement) basé sur le transfert de popularité d'une page vers une autre, qui se fait par le biais d'un lien hypertexte. ce transfert de popularité répond au doux nom de « Link Juice » ou « PR Juice », autrement dit, jus de lien ou de PageRank. On sait également que le transfert de « PR Juice » est divisé par le nombre de liens sortants.

Dans un article, WebRankInfo compare tout ceci à de l'eau salée s'échappant d'une bouteille (le sel représentant le PR). Plus il y a de trous, moins il y a de liquide qui s'échappe par chaque trou. Si une page possède 100 liens pointant vers des pages externes, chacune des pages cibles ne recevra qu'une petite portion du précieux PR Juice. (Pour plus d'infos,

consulter <http://www.webrankinfo.com/dossiers/strategies-de-liens/pagerank-sculpting-nofollow>)

L'attribut rel=nofollow peut s'imaginer comme une rustine qui se place devant chaque trou, et qui permet le passage du liquide mais pas du sel. Un lien nofollow ne transfère pas le PR Juice, tout en permettant la navigation par les internautes.



Principe du PageRank Sculpting expliqué sur <http://www.evisibility.com/blog/no-follow-tag/>

Pour un webmaster un peu bricoleur, il peut donc être intéressant de décider sur chaque page quels seront les liens qui transmettront la popularité et le PageRank. Ceci est d'ailleurs suggéré à demi mot par Google dans son aide en ligne :

Priorité d'exploration : les robots des moteurs de recherche ne peuvent pas se connecter ou s'inscrire à votre forum en tant que membre. Il n'y a donc aucune raison d'inviter Googlebot à suivre des liens du type "inscrivez-vous ici" ou "se connecter". L'utilisation de l'attribut nofollow sur ces liens permet à Googlebot d'explorer d'autres pages que vous souhaiteriez voir dans l'index Google

(<http://www.google.com/support/webmasters/bin/answer.py?answer=96569>).

Le PageRank Sculpting consiste donc à bloquer ou non les liens sortants depuis une page, de façon à influencer sur le PageRank des pages ciblées. Les liens vers des pages peu intéressantes seront bloqués par un nofollow, de façon à ne pas perdre inutilement de PR Juice.

En d'autres termes, imaginons qu'une page A contienne, dans un premier temps, 3 liens, dont aucun n'est en "nofollow". Chaque page liée reçoit donc 1/3 du « PR Juice », de la transmission de popularité de A.

Le fait de mettre, dans A, un lien en nofollow fait que, en première approche, les deux autres pages se partagent cette « capacité de vote » et reçoivent donc du « PR Juice » chacune au lieu d'un tiers auparavant. Pas négligeable...

Les limites du PageRank Sculpting

Des comportements différents selon les moteurs

Toute technique de référencement possède ses limites et le PageRank Sculpting n'y fait pas exception. Tout d'abord il convient de faire remarquer que si l'attribut `rel=nofollow` est bien pris en compte par Google (il en est l'inventeur après tout), ce n'est pas le cas de tous les moteurs.

Quand il trouve un `nofollow`, Google ne suit pas le lien et ne prend pas en compte l'*anchor text* du lien (notion de réputation). Confronté à un `nofollow`, Yahoo a un comportement légèrement différent :

Le robot d'indexation Yahoo! Slurp comprend l'attribut `rel="nofollow"` appliqué aux liens hypertextes. Les webmasters et blogueurs peuvent l'appliquer à n'importe quel lien de leur page pour indiquer qu'il n'est pas vérifié. Yahoo! Slurp peut suivre un lien marqué "nofollow" afin d'en découvrir le contenu, mais ce lien ne sera pas considéré comme un lien approuvé et ne sera pas pris en compte dans le calcul de l'indice de popularité de la page-cible (<http://help.yahoo.com/l/fr/yahoo/search/indexing/nofollow.html>).

Dans ce contexte, Yahoo peut donc indexer les pages ciblées, même s'il ne tient pas compte du lien dans le calcul de la popularité.

Quant au moteur de recherche Ask.com, ce serait encore pire : il ne tient pas du tout compte de la directive `nofollow`. Les liens seront donc suivis, indexés, et compteront dans le calcul de popularité.

En conclusion, il est illusoire d'utiliser `rel=nofollow` pour empêcher l'indexation d'une page. De plus, imaginons que Google trouve sur le web beaucoup de liens pointant vers une page. Il accordera automatiquement de la popularité à cette page, quelle que soit la stratégie mise en œuvre à l'intérieur du site en lui-même par le biais des attributs `nofollow` sur les liens internes.

Google n'aime pas la sculpture

En juin 2009, Matt Cutts lançait une véritable bombe dans le monde du web, en expliquant sur son blog que Google tenait quand même compte des liens bloqués par `nofollow`. (Voir <http://actu.abondance.com/2009/06/google-ne-tiendrait-plus-compte-de.html> et <http://www.mattcutts.com/blog/pagerank-sculpting/>).

Au début, si une page possédait plusieurs liens en `nofollow`, le PageRank n'était transmis que par les autres liens sortants (principe des membranes filtrant le sel dans les trous de bouteille). Les liens `nofollow` n'était pas pris en compte dans le calcul du PageRank global transmis.

Or selon Matt Cutts, une mise à jour de l'algorithme Google comptabilise désormais les liens `nofollow` pour diviser le PageRank transmis par la page.

Si l'on revient à notre exemple avec les bouteilles, tout se passe comme si les membranes, non seulement empêchaient le sel de passer par les trous mais en fixaient une partie. Un lien `nofollow` retient une partie du PageRank de la page, ce qui diminue le PR Juice transmis par les autres liens de la page. Utiliser des `nofollow` revient donc à gâcher une quantité précieuse de PR Juice.

La règle est donc désormais différente : si on reprend notre exemple précédent avec la page A qui a 3 liens sortants et qu'un lien est en `nofollow`, les deux autres pages ne recevront toujours qu'un tiers du « PR Juice » de A au lieu de la moitié, un tiers étant donc perdu...

Selon Matt Cutts, ce système a été mis en place car certains webmasters « abusent », en bloquant trop de pages pertinentes sur leur site. Le pire, c'est qu'il ajoute que ce filtre a été mis en place par Google depuis un an, ce qui signifie que de nombreux webmasters « sculpteraient » leur PageRank depuis des mois en vain...

Concernant son propre blog, Matt Cutts déclare d'ailleurs qu'il n'utilise pratiquement pas le `nofollow` :

Je laisse vraiment le PageRank s'écouler librement sur mon site, et je vous recommanderais de faire de même. Je n'ai pas ajouté de nofollow sur mes pages de catégorie ou d'archive. Le seul endroit où j'ai délibérément ajouté un nofollow est sur le lien vers mon flux, parce que je ne trouve pas super utile d'avoir des flux RSS/atom dans les résultats des moteurs de recherche. Même pour ça ce n'est pas strictement nécessaire, parce que Google et les autres moteurs de recherche font un bon boulot en distinguant les liens des pages web classiques.

Remarque : Matt Cutts parle ici des liens insérés dans les articles et les menus : les liens présents dans les commentaires du blog sont par contre désactivés à l'aide d'un attribut `rel="external nofollow"`.

La fin du PageRank Sculpting ?

Il est clair que Google, au travers de cette action, a souhaité mettre un terme à une utilisation détournée de l'attribut nofollow. A l'origine, ce système n'avait été pensé que pour lutter contre le Spam, pas pour optimiser le transfert de PageRank à l'intérieur d'un site.

Faut-il pour autant abandonner le PageRank Sculpting ? Pas forcément, car il existe d'autres solutions pour contrôler le linking interne.

La balise Meta Robots

La balise meta robots est prise en compte à 100% par tous les moteurs de recherche. Si vous ne souhaitez pas qu'une page soit indexée, il suffit d'utiliser l'instruction « noindex » de cette balise. Voilà une façon simple de bloquer le référencement d'une page peu importante, telle qu'un formulaire de contact ou une page de mentions légales.

L'utilisation de noindex a aussi l'avantage de continuer à permettre à l'internaute d'accéder aux pages par l'intermédiaire des liens internes. Ceci dit, le fait de bloquer l'indexation d'une page n'empêchera les liens vers elle d'être pris en compte par les moteurs dans le calcul du PageRank et du « PR Juice » fourni...

Le fichier Robots.txt

Autre mesure radicale pour empêcher l'indexation d'une page et sa prise en compte dans le calcul du PageRank, l'utilisation du fichier robots.txt permet d'indiquer aux moteurs quelles pages doivent être ignorées lors du processus d'indexation.

Cette mesure est encore plus efficace que la précédente puisque les moteurs sont avertis de l'interdiction avant même d'entrer sur le site ; dans le cas d'une balise meta robots, les moteurs doivent lire le contenu de la page avant de suivre les instructions données dans la balise, ce qui peut être à l'origine de quelques problèmes. Dans certains cas, la page peut continuer à être présentée dans les résultats de recherche.

Les liens JavaScript

Un substitut de l'utilisation du lien nofollow est l'utilisation d'un lien JavaScript, utilisable par les internautes mais pas par les moteurs de recherche. De cette façon vous serez sûr que le lien ne sera pas suivi et ne sera pas pris en compte pour le calcul de la notoriété.

Les syntaxes suivantes (mais il y a en d'autres) peuvent servir d'alternative à un nofollow :

```
<a href=# onClick="document.location='http://www.monsite.fr'">Mon site</a>
```

```
<a href="javascript:void(0)"  
onClick="window.open('http://www.monsite.fr')">Mon site</a>
```

Le problème est que Google travaille sur le fait d'interpréter les événements onClick sur une page web (et pas seulement sur les balises de lien). Ceci a été annoncé officiellement en juin 2009, lors d'une conférence où Google annonçait également d'importantes avancées en matière d'interprétation du Flash (voir <http://actu.abondance.com/2009/06/google-sur-le-point-de-reconnaitre-les.html>).

Pour le moment il est impossible de dire si les liens onClick seront à l'avenir comptabilisés dans la transmission du PageRank. Ce qui est certain c'est que plusieurs webmasters ont vu indexer des pages qui n'étaient accessibles que par des liens JavaScript.

Le JavaScript ne serait donc pas la solution ultime pour contrôler le linking interne sur un site. De plus il peut être gênant en termes d'accessibilité, car tous les internautes ne peuvent pas utiliser le JavaScript.

On peut également imaginer que des liens « indésirables » en termes de transmission de « PR Juice », on fasse appel à des technique d'obfuscation (<http://fr.wikipedia.org/wiki/Catégorie:Obfuscation>) ou tout simplement que l'on mette ces liens dans des iframes, etc. On peut en fait imaginer de nombreuses façons de « cacher » des liens à un moteur de recherche tout en les laissant opérationnel pour les internautes...

La stratégie de linking

La solution la plus efficace est peut-être ailleurs... Il s'agit d'une recette aussi vieille que le référencement et Matt Cutts en parle dans une vidéo intitulée « What are your views on PageRank sculpting? » (Voir <http://www.youtube.com/watch?v=R4IE4WLPLZQ>).



Pour optimiser la transmission du PageRank, il suffit simplement de choisir quels liens seront placés sur les pages à fort PageRank. Pour qu'une page reçoive un maximum de PR Juice, il faut qu'elle soit liée depuis la homepage ! C'est tout simple et cela consiste, en quelque sorte, à mettre en place une stratégie de « PR Sculpting naturel »...

Cette technique n'est pas nouvelle et s'apparente davantage à une bonne stratégie de linking interne sur un site web. Eviter de mettre trop de liens depuis la Homepage, cibler les pages stratégiques, utiliser de bons *Anchor texts*, tout ceci participe à la construction d'un *linking* efficace et constitue une forme de PageRank Sculpting, plus traditionnel.

Conclusion

Le PageRank Sculpting a donc encore de beaux jours devant lui, même s'il faut abandonner l'idée d'utiliser l'attribut nofollow, à moins que Google change d'avis d'ici quelques temps. En effet, avec sa nouvelle façon de prendre en compte cet attribut, on peut par exemple faire baisser de façon drastique le « PR Juice » transmis par une page en ajoutant de nombreux commentaires contenant des liens sur celle-ci. Ces liens ne seront pas suivis par les spiders mais seront pris en compte dans la calcul du PR transmis et ils feront donc baisser la part transmise aux « vrais » liens de la page. Diabolique comme dégât collatéral...

La panoplie des techniques et d'outils proposés aux webmasters est vaste, et il existe certainement de nombreuses astuces pour optimiser la transmission du PageRank, sans contrevenir aux nombreux interdits édictés par les moteurs de recherche. En définitive, si les nouvelles techniques sont intéressantes, il ne faut pas oublier les bonnes vieilles recettes, sur lesquelles est basé tout référencement digne de ce nom...

Guillaume Thavaud, cellule de veille (veille@brioude-internet.fr) de la société Brioude-Internet Référencement (http://www.referencement-2000.com/)

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

<http://abonnes.abondance.com/blogpro/2009/09/pagerank-sculpting-etat-des-lieux.html>