

Le PageRank Modeling, avenir du PageRank Sculpting ?

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

La notion de PageRank a, depuis des années, fait couler beaucoup d'encre (parfois virtuelle) dans le monde du référencement naturel. Les techniques de PageRank Sculpting, visant à mieux distribuer le PageRank entre les pages internes d'un site, ont été utilisées mais aussi décriées, avant que Google ne change les règles du jeu l'année dernière. Il nous semblait important de faire un point approfondi sur ces techniques, Philippe Yonnet, président de l'association SEO Camp, expliquant ici un nouveau concept, celui de "PageRank Modeling", plus efficace et plus respectueux, selon lui, des règles de conduite des moteurs de recherche...

Depuis quelques semaines un violent débat fait rage entre certains référenceurs américains au sujet du PageRank Sculpting (NDLR : voir également notre article de septembre 2009 à ce sujet : <http://recherche-referencement.abondance.com/2009/09/pagerank-sculpting-etat-des-lieux.html>). Certains (en particulier Michael Martinez) ont remis en cause l'intérêt de la technique, et même son bien-fondé théorique. D'autres, comme Rand Fishkin de SEOMoz, continuent de défendre cette technique contre vents et marées en expliquant qu'elle est efficace, et même indispensable pour certains sites. Ces échanges récents ne sont que la résurgence d'échanges aigre-doux entre spécialistes qui durent depuis des années autour du PageRank, et des techniques de PR sculpting. Le débat a été ranimé par une annonce qui a fait l'effet d'une bombe dans le Landerneau des référenceurs : les liens en nofollow sont pris en compte dans le calcul du PR par Google !

Mais au fond, qu'est-ce que le PR sculpting ? Quel est l'intérêt réel de cette technique ? Et est-ce que le PR sculpting est sanctionné par Google ? Nous essaierons tout au long de cet article de répondre à ces questions et d'examiner les arguments des détracteurs et des promoteurs de cette technique...

Nous verrons également que les techniques "classiques" de PR sculpting ne sont pas des solutions durables et qu'elles ne produisent pas toujours les résultats escomptés. Il existe une approche alternative destinée à améliorer la circulation et la concentration de "link juice" sur les sites. Pour décrire ces techniques, nous avons introduit voilà plusieurs mois le concept de "PR modeling". Nous découvrirons à la fin de cet article en quoi cette approche nous semble plus respectueuse des règles édictées par les moteurs et en général plus efficiente.

Mais qu'est-ce que le PageRank Sculpting ?

Le rôle des signaux s'appuyant sur des liens dans les algorithmes des moteurs de recherche

Les moteurs de recherche essaient d'utiliser toutes les informations à leur disposition pour réaliser un classement pertinent des pages en réponse à une requête donnée (des informations que l'on appelle "signaux" chez Google). Les premiers moteurs s'appuyaient essentiellement sur des critères liés au contenu des pages : il est logique en effet de classer les pages en fonction de la "proximité" entre le contenu de ces pages et la requête. Il existe d'autres "signaux" intéressants à exploiter : les informations véhiculées par les liens pointant vers les pages. Tout d'abord, les liens apportent de l'information sur la page grâce à l'*anchor text* du lien distant. Le contenu indexé comprend donc non seulement les textes de la page visible par les internautes, mais aussi l'ensemble des anchor texts des liens pointant vers cette page.

Ensuite, les liens donnent des indications intéressantes sur l'importance des pages sur le net : plus une page reçoit de liens, plus elle est importante, en sachant que l'on doit prendre en compte le fait que tous les liens ne se valent pas, car certains émanent de pages considérées elles mêmes comme importantes, et que le nombre de liens sur la page doit également être

pris en compte. Ce que nous venons de décrire, ce sont tout bonnement les principes à l'oeuvre derrière le PageRank, l'algorithme de calcul d'une note d'importance de la page sur le web introduite par Google.

On peut tirer d'autres informations des liens pointant vers une page : ces liens peuvent donner des informations sur la qualité de la page, sa thématique, sa popularité, son caractère "spammy" etc...

Ces critères "indépendants de la requête" (les notes calculées à partir des liens ne changent pas d'une requête à une autre) ne sont qu'une partie des "signaux" exploités par un moteur de recherche comme Google. Ils ne peuvent pas avoir en règle générale une pondération forte dans l'algorithme : les critères **dépendants** de la requête sont forcément les plus intéressants pour arriver à distinguer les pages pertinentes sur une requête et celles qui ne le sont pas. Les critères **indépendants** de la requête (comme le PageRank) servent par contre à mieux classer les pages qui auraient sinon des notes proches sur des critères de similarité sémantiques.

Liens internes, aussi importants que les liens externes ?

Dans l'algorithme initial du PageRank, aucun *distingo* n'était fait entre liens externes (reliant des pages de sites différents) et liens internes : tous avaient le même poids dans l'algorithme. Il est difficile de dire si les choses sont toujours aussi tranchées aujourd'hui, par contre il est clair que les signaux fournis par le linking interne restent une source d'information indispensable et qu'ils sont exploités par les moteurs de recherche.

Si les moteurs prennent en compte les liens internes dans leur algorithme, c'est parce qu'il est assez fréquent de voir des sites dont seules quelques pages reçoivent des backlinks externes : il s'agit souvent de la home et des pages rubriques/catégories les plus importantes. Pour les pages profondes, l'essentiel et parfois même l'intégralité des backlinks proviennent de pages du même site. Pour arriver à doter les pages profondes de notes indépendantes de la requête, il est indispensable :

- de tenir compte du maillage interne pour savoir quelle part du potentiel de PageRank reçu par le site (link juice) doit être transmis aux pages profondes ;
- et de hiérarchiser les notes des pages en fonction de ce maillage interne.

Le maillage hypertexte entre pages du même site influence le référencement

Le problème, c'est que certaines structures de sites ont des effets inattendus sur les "notes" ainsi calculées. Par exemple, il est fréquent de voir que certaines pages du site sont bombardées de liens émanant de TOUTES les pages du site, alors qu'elles ne sont pas forcément les pages les plus importantes et encore moins les pages les plus intéressantes à positionner sur les moteurs de recherche. C'est le cas en particulier des pages liées par des liens en pied de page (par exemple "mentions légales" ou "contact"), mais aussi certaines pages "techniques" comme la page contenant le formulaire "envoyer à un ami".

On peut aussi avoir des structures extrêmement profondes (il faut un grand nombre de clics pour découvrir une partie du contenu du site). Un grand nombre de sites sont concernés, même parfois ceux qui croient appliquer des règles d'ergonomie ("tout le contenu est accessible en 3 clics depuis la home"). Il suffit d'une pagination mal conçue, ou d'une navigation par liens complexes pour créer une situation dans laquelle de nombreuses pages se retrouvent au delà de cinq ou six clics depuis la home. Or plus on s'éloigne de la *home*, plus le PageRank transmis diminue, et plus les pages sont profondes, moins elles reçoivent par ailleurs de liens internes. La combinaison de ces phénomènes réduit le PR de ces pages à la portion congrue. Or (ce point a été rappelé par Matt Cutts régulièrement), les pages sont crawlées par ordre décroissant de PR, ce qui signifie qu'une page qui ne reçoit pas une quantité suffisante de PR peut ne plus être crawlée (notons au passage qu'il existe d'autres critères pilotant le crawl et le recrawl des pages, mais le PR est l'un des plus importants).

Enfin, les menus de navigation influencent fortement la hiérarchie des notes sur ces critères. Par exemple, certaines pages de catégories peuvent être liées à toutes les pages du site, parce qu'un menu présent sur toutes les pages contient systématiquement un lien vers ces pages. La grande masse des liens internes sont donc créés par les menus de navigation, ou les fils

d'Ariane, et selon la structure de menus adoptée, la hiérarchie des pages sera différente.

Et le PageRank sculpting alors ?

Le PageRank sculpting consiste à tenter de canaliser le potentiel de PageRank (*link juice*) vers les pages importantes, en empêchant le PR de se diffuser vers les pages considérées comme peu importantes. Pour obtenir cet effet, il existe trois solutions :

- soit montrer aux moteurs une page qui ne contient pas les mêmes liens que la page vue par les internautes (mais on fait alors du cloaking, c'est-à-dire une méthode non conforme aux conditions générales des moteurs de recherche).
- soit empêcher les moteurs de découvrir certains liens : on "masque" ou on "brouille" des liens (en règle générale avec un script en javascript, voire en Flash/Actionscript). On peut également utiliser le chargement de certains contenus en Ajax ou dans des iframes.
- soit en indiquant aux moteurs de ne pas tenir compte des liens découverts.

La dernière méthode s'appuie sur l'attribut nofollow, mais nous verrons plus loin que cette technique qui était devenue très utilisée n'est probablement plus efficace et même devenue antiproductive.

Les arguments des détracteurs

La technique dite du "PR Sculpting" est souvent critiquée dans le petit monde du référencement. Les arguments utilisés sont souvent les mêmes, et il est intéressant de les analyser pour voir s'ils sont fondés ou pas.

1er argument : on ne peut pas mesurer le PR, donc c'est stupide de vouloir manipuler un critère en aveugle

Bien sûr, on peut remarquer que Google fournit avec le PageRank visible dans la barre d'outils, des valeurs grossières (la fameuse échelle de 0 à 10), et mises à jour à une fréquence tellement faible qu'elles sont inexploitable.

Mais en réalité, la formule initiale du PageRank est publique et parfaitement connue, ainsi que de nombreuses variantes de cette formule inventées au fil des ans. Cela signifie que quiconque peut calculer une note précise de PR interne, bien plus précise que celle de la "toolbar". Le fait que l'on ne sache pas si Google utilise exactement tel ou tel type de formule n'enlève pas aux notes calculées à partir des formules connues leur rôle d'indicateur de la manière dont se diffuse et s'accumule le "link juice" sur les pages.

On peut remarquer par ailleurs dans la littérature scientifique (abondante) sur le sujet du PageRank, que la plupart des améliorations proposées consistent justement à sophistiquer le rôle du coefficient d'atténuation dans la formule (transformé en "fonction d'atténuation" non fixe) sans que l'on remette en cause les principes généraux du calcul, le niveau d'atténuation étant par ailleurs contraint pour des raisons mathématiques (le calcul doit converger vers une valeur finie) et pratiques.

Bref, il semble vraiment peu probable que les notes calculées par Google le soient à partir de principes ou de formules complètement différentes. La vraie part d'inconnu réside dans la manière dont un lien transmet ou non du PageRank à une autre page, et la manière dont les liens sont comptés sur les pages.

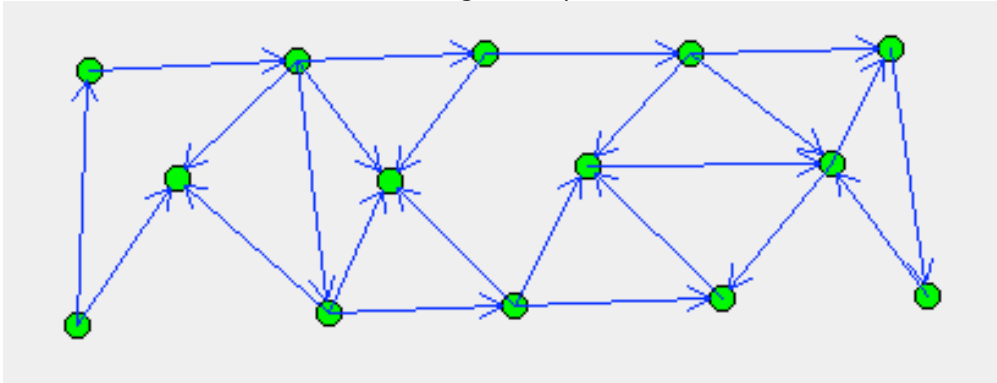
Grâce à ces formules, on peut quantifier de manière parfaitement précise et scientifique les effets que peuvent avoir le maillage interne du site sur une note de type "PageRank". Cela permet donc de créer une modélisation mathématique des effets de ce maillage. Bien sûr, il est incontestable que l'exercice a ses limites, et que la précision du calcul ne doit pas laisser penser que l'on calcule cette note de manière exacte : c'est probablement faux, puisque l'on

ne connaît pas la formule exacte utilisée par Google et la manière dont les signaux envoyés par le maillage hypertexte sont utilisés. Mais travailler avec ces indicateurs permet clairement de ne pas travailler en "aveugle".

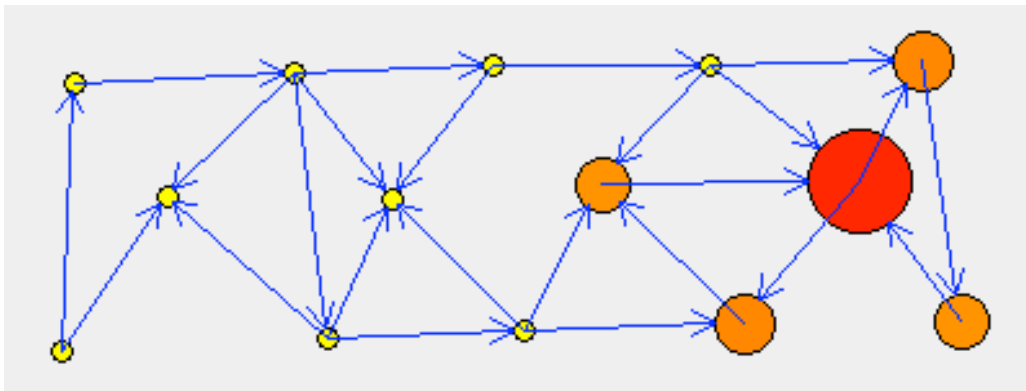
Voici un exemple de la modélisation du calcul du PageRank à l'aide de l'outil « PageRank Demo » développé par Hubert Wassner de l'ESIEA

(<http://professeurs.esiea.fr/wassner/?2007/06/03/74-l-algorithme-pagerank-comment-a-marche>)

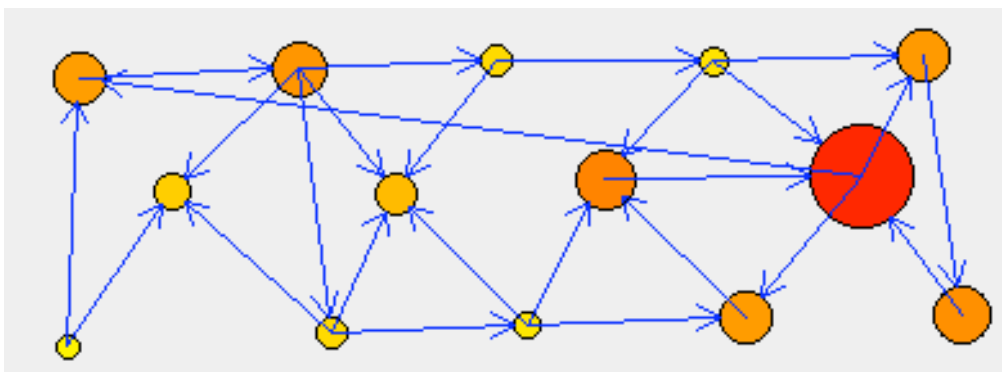
Maillage de départ



Calcul du PR (la taille du rond reflète l'importance du PR obtenu) : on voit que l'endroit où se concentre le PR n'est pas toujours intuitif



Calcul du PR après changement dans la structure des liens



Un seul lien a été ajouté (à vous de trouver lequel) : la valeur des PageRanks a beaucoup changé

2e argument : si on peut estimer une note de PR interne, par contre on ne connaît pas le PR transmis par les backlinks externes, et donc on ne peut pas calculer le vrai PR des pages

C'est parfaitement exact, mais ce n'est pas un problème totalement bloquant. Dans une logique de PR sculpting, on cherche le maillage interne optimal. La manière dont le PR se diffuse et s'accumule ne dépend pas du PR transmis depuis l'extérieur du site. Donc connaître un indicateur du PR interne suffit pour étayer la plupart des raisonnements. Par contre, ne pas connaître le PR transmis depuis l'extérieur a deux inconvénients majeurs :

- Tout d'abord, si un fort PR est transmis à une page profonde, et pour créer un maillage optimal, il faudra tenir compte de la nécessité de diffuser ce fort apport de PR dans d'autres pages du site : la structure ainsi créée peut donc être très différente de la structure que l'on aurait créé en raisonnant uniquement sur le PageRank Interne. Mais d'une part ce genre de situation est rare, l'essentiel du PR parvient souvent au sommet de l'arborescence du site.

- Ensuite les valeurs données par la toolbar permettent de se rendre compte que le modèle élaboré à partir du calcul du PR interne doit être revu et corrigé. Les corrections apportées sont forcément grossières, mais sur un plan opérationnel, elles permettent malgré tout d'optimiser le maillage de manière efficace à défaut d'être parfaite.

3e argument : jamais personne n'a démontré que cela marchait

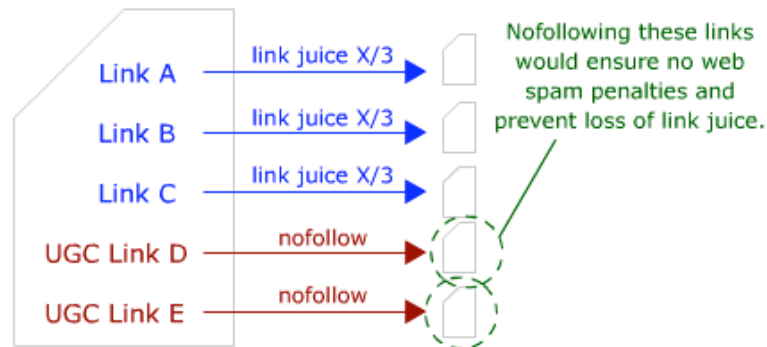
Souvent on fait remarquer l'absence de démonstration scientifique de l'efficacité du PR sculpting. Les promoteurs de la méthode ont tendance à rétorquer que "*personne n'a démontré que cela ne marchait pas non plus*". En fait, la méthode est forcément empirique, et ne peut pas se faire en "conditions de laboratoire". A chaque fois que l'on essaie de voir si l'amélioration du maillage interne a permis d'améliorer un référencement, il sera toujours difficile de séparer les effets de la méthode des effets d'un changement dans l'index ou dans l'algorithme de Google, ou d'un changement dans d'autres signaux sur le même site. Il serait possible de montrer, en faisant des statistiques sur un grand nombre de sites et en prenant beaucoup de précautions pour que les chiffres aient du sens, que la méthode fonctionne. Mais là on se heurte à un autre obstacle : les référenceurs ne prennent pas le temps de faire ce genre d'études, et ont tendance, sauf un nombre restreint d'entre eux, à ne pas partager leurs informations sur ce qui marche ou ne marche pas.

En fait, ceux qui croient dans l'efficacité de la méthode s'appuient sur des considérations théoriques, mais qui peuvent être prises (en fait c'en est un) pour un syllogisme : "*Google utilise un critère de type PageRank dans son algorithme. Donc le maillage hypertexte interne sur un site joue un rôle dans les positions des pages dans les résultats de Google. Si mon maillage actuel semble ne pas être optimal pour transmettre le maximum de PR aux pages importantes, le modifier aura un effet positif sur le référencement*". Evidemment, on est loin du raisonnement scientifique. Certains référenceurs font confiance à la méthode parce qu'ils ont l'expérience d'améliorations du référencement obtenues après avoir fait du PR sculpting. Mais d'autres se font aussi l'écho d'expériences inverses.

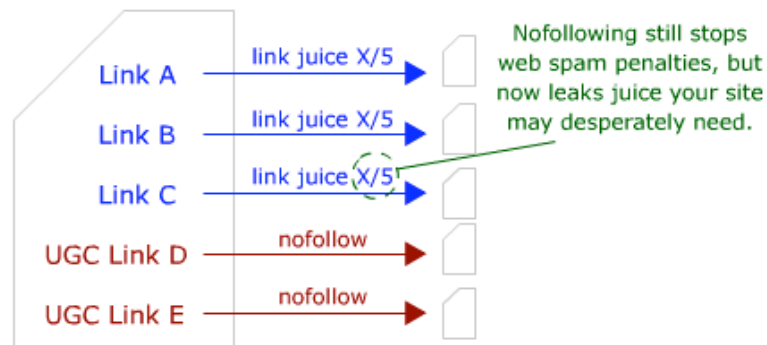
4e argument : les outils utilisés ne marchent pas

- S'agissant de l'utilisation du `rel='nofollow'`, la messe est dite : Matt Cutts a révélé en juin dernier que les liens contenant un attribut `rel='nofollow'` sont comptés comme liens sortants dans l'algorithme utilisé par Google. Dans ce contexte, l'utilisation de l'attribut pour le PR sculpting perd tout intérêt : au lieu de concentrer le link juice sur les liens normaux, cette méthode fait perdre du PR.

Old Best Practice for Links in UGC on Websites



New Risks Associated with Allowing Links in Any Form of UGC



Graphique publié sur SEOMoz résumant l'impact de la prise en compte des liens en nofollow sur le calcul du PR

- Le cloaking est une méthode qui expose à des sanctions de la part de Google. On comprend les réticences de certains à utiliser cette méthode. Il faut remarquer que cette solution est néanmoins adoptée sur certains sites à très forte audience, sans que Google (Matt Cutts a été interrogé sur certains cas précis) ne les condamne de manière claire (?).

- L'utilisation d'iframes reste possible (pour les liens dans le pied de page par exemple). Mais elle n'est pas utilisable partout, et son efficacité est jugée limitée.

- Il reste la solution des liens "javascriptés". Souvent les utilisateurs de cette méthode sont déçus parce qu'ils s'aperçoivent que Google arrive à découvrir les liens dans les javascripts. En fait on sait que depuis plusieurs années Google parvient sans difficulté à "parser" les javascripts et découvrir les liens "en clair" au sein des programmes. Cette capacité à lire le contenu des javascripts et même dans une certaine mesure à exécuter le javascript fait peur à certains webmasters qui se disent qu'il n'est donc pas possible de "masquer" les liens avec un javascript. En fait, si Google a fait des progrès dans l'analyse des javascripts, il est encore incapable de comprendre des scripts parfois simples. Pour "masquer" un lien en javascript, il suffit donc de créer un programme suffisamment complexe pour "bloquer" la découverte du lien par Google.

5e argument : les bénéfices espérés sont faibles et la méthode est sans intérêt

Si on pose comme hypothèse que les signaux provenant des liens ont un poids peu important dans l'algorithme, il est clair que les méthodes de PR sculpting peuvent sembler être du temps perdu. Il est donc assez logique que les référenceurs qui ne croient pas ou plus que le PR joue un rôle clé dans l'algorithme critiquent l'intérêt de la méthode. A l'inverse ceux qui pensent

que le PR a toujours beaucoup d'importance auront tendance à utiliser plus volontiers le PR sculpting.

6e argument : cela ne marche que pour Google

Il est clair que dans la plupart des cas, le PR sculpting est utilisé en ciblant exclusivement l'algorithme de Google. Mais il est difficile de reprocher à des référenceurs de se focaliser sur le moteur de recherche leader, surtout lorsqu'il est en situation de quasi monopole comme en France. Mais en réalité, des moteurs comme Bing et Yahoo ont régulièrement expliqué qu'ils utilisaient eux aussi les informations tirées des liens dans leur algorithme. Donc le fait que le PageRank soit un algorithme propre à Google ne permet pas de statuer sur l'inefficacité du PR sculpting pour les autres moteurs.

Qu'en pense Google ?

Le moins que l'on puisse dire, c'est que la position de Google sur le PR sculpting est plutôt ambiguë. Matt Cutts a été régulièrement interrogé sur le sujet, et ses réponses ont finalement peu évolué dans le temps. Pour résumer sa position, il a affirmé à plusieurs reprises que s'il était à la place des webmasters, en règle générale il ne dépenserait pas du temps et de l'énergie à sculpter du PR. Ces déclarations ont conforté les détracteurs du PR sculpting dans l'idée que la méthode n'avait pas d'intérêt, et que Matt Cutts en plus déconseillait la méthode.

Mais il faut remarquer qu'il citait souvent en même temps des approches qui, selon lui, étaient plus productives et qui malgré tout consistaient à améliorer le maillage des liens internes d'un site. Ces mêmes déclarations peuvent donc également être vues comme une reconnaissance du bien fondé de l'intérêt d'optimiser la structure des liens au sein d'un site.

Matt Cutts à rappelé cette position dans une interview récente donnée à Eric Enge (datée du 14 mars 2010 : <http://www.stonetemple.com/articles/interview-matt-cutts-012510.shtml>) :

Eric Enge: Et à propos du PageRank Sculpting ? Est-ce que les éditeurs doivent envisager d'utiliser des redirections de liens codées en javascript, ou utiliser des liens placés dans des iframes ?

Matt Cutts: Mon avis là dessus reste en gros le même que mon avis sur les idées à l'origine du PR sculpting. Nous avons déjà évoqué auparavant le fait que le PR sculpting n'était pas la méthode la plus efficace pour guider Googlebot dans sa découverte d'un site, et nous avons également précisé que le PageRank Sculpting n'était pas la meilleure façon d'utiliser son temps, parce que ce temps pouvait être consacré à obtenir plus de liens externes ou à créer de meilleurs contenus pour votre site.



Le PageRank Sculpting consiste à prendre le PageRank que vous avez déjà et à essayer de le guider vers d'autres pages en espérant qu'elles seront plus efficaces, mais il y'a des façons plus efficaces d'y parvenir. Si vous avez un produit qui a d'excellents taux de conversion et vous permet d'avoir des marges fantastiques, vous pouvez le placer directement à la racine de votre site et au milieu de la page. Une grande quantité de PageRank se diffusera grâce à ce lien vers cette page produit spécifique.

Travailler l'architecture du site, améliorer la manière dont les liens et la structure apparaissent sur une page pour diriger un maximum de gens vers les produits que vous voulez qu'ils voient, est en réalité une meilleure approche que d'essayer de faire du sculpting individuel sur des liens. Si vous pouvez faire en sorte que votre architecture de site concentre le PageRank sur les pages les plus importantes, ou celles qui génèrent les

plus fortes marges, c'est une bien meilleure méthode pour sculpter directement le PageRank que d'essayer d'utiliser une iFrame ou un javascript encodé.

Je pense que si vous pouvez réaliser une architecture de site correcte du premier coup, ensuite vous aurez moins de travail, voire même aucun besoin même d'envisager du PR Sculpting.

Juste pour aller plus loin que ça et pour être parfaitement clair, les gens peuvent faire ce qu'ils veulent sur leurs propres sites, mais selon mon expérience, le PageRank sculpting n'a pas forcément été la meilleure utilisation que l'on puisse faire de son temps disponible."

Les limites du PR Sculpting

On pourrait penser que les conseils de Matt Cutts sont dictés par la seule volonté de dissuader les webmasters d'utiliser une technique considérée comme "spammy". En réalité, ses conseils sont judicieux, et un référenceur qui a bien compris les mécanismes à l'oeuvre derrière la diffusion du link juice et une certaine expérience du PR sculpting donnerait exactement les mêmes.

Le problème de la volumétrie

Beaucoup de webmasters pensent faire du PR sculpting efficace, en cachant quelques liens sur quelques pages à un spider (en particulier les liens vers les pages mentions légales, contacts etc...). Dans la pratique, si cela fait à coup sûr baisser le PageRank interne de ces pages, cela ne fait pas toujours augmenter le PR d'autres pages dans des proportions sensibles. Tout dépend du nombre de liens par pages, et du nombre de liens entrants par page. Si les pages contiennent en général beaucoup de liens, le PR récupéré ainsi est faible, et l'impact en terme de référencement peut être nul. Même chose si vous avez un menu qui crée autant de liens entrants que de pages du site pour 300 pages de catégories : le PR récupéré sur les quelques pages "techniques" situées en pied de page va être faible comparé au PR diffusé aux 300 pages. Bref dans bien des cas, ces quelques liens "brouillés" auront peu d'effet.

Par ailleurs, faire du PR sculpting sur des petits sites n'est pas toujours non plus une bonne idée. En théorie, ce sont sur ces sites que les effets du "masquage" de liens en footer peut-être le plus efficace. Mais il est souvent encore plus efficace de travailler à développer fortement son contenu et son linking externe.

Supprimer des liens peut faire baisser le potentiel de PR

La manière dont le PageRank se diffuse et s'accumule dans un site n'est pas du tout intuitive. Certains webmasters font donc parfois de graves erreurs de raisonnement et en fait se tirent une balle dans le pied en utilisant ces techniques sans en comprendre les effets réels. Tant que le volume de potentiel de PR est faible, tout va bien. Mais quand on commence à "masquer" un grand nombre de liens, la perte peut être sévère.

En général, les erreurs de raisonnement proviennent du fait que les webmasters raisonnent trop en termes de "fuites" locales du PageRank. En supprimant des liens, on peut parvenir à faire monter localement le PageRank de quelques pages, mais souvent au détriment du potentiel de PageRank du site (c'est à dire la somme de tous les PageRanks de toutes les pages du site). Bref, le gain local peut faire perdre du PR à un grand nombre de pages, et ce n'est pas toujours l'effet recherché.

Ce genre de dégâts collatéraux s'était multiplié avec l'adoption de l'utilisation de l'attribut `rel=nofollow` en tant qu'outil de PR sculpting. En effet, cette solution était facile à implémenter, et a contribué fortement au développement du PR sculpting. Le problème c'est que Google s'est vite aperçu qu'un grand nombre de structures de sites étaient rendues totalement impossibles à comprendre à cause de ces "nofollow" et que l'adoption progressive de la technique conduisait dans bien des cas à l'effet inverse de celui recherché par les webmasters.

La décision de changer la manière dont les liens en "nofollow" étaient prises en compte doit donc être interprétées à la fois comme une solution qui résout un problème technique pour Google, mais aussi comme un coup d'arrêt à une technique contreproductive lorsqu'elle est mal comprise. Le problème pouvait aller jusqu'à interdire à Google la découverte de pages auxquelles les internautes avaient accès !

Souvent le PR sculpting appliqué à un grand nombre de liens sera inefficace voire dangereux

Dans bon nombre de cas, c'est la structure du site prise dans sa globalité qui pose problème. Soit parce que la structure va laisser s'accumuler le PR dans les pages de catégories au détriment de pages plus profondes. Soit parce que la structure est "invertébrée" et donne de l'importance à des pages sans intérêt. Dans ce cas, vouloir traiter le problème en enlevant des liens est quasiment toujours contreproductif : si on enlève peu de liens, les effets sont rarement mesurables. Si on enlève un volume de liens conforme à l'étendue du problème, la perte globale de "link juice" sera telle que l'on observera aussi bien des améliorations locales du crawl ou du positionnement que des effets contraires sur d'autres pages.

En clair, le PR sculpting n'est pas la panacée universelle : il ne résout pas tous les problèmes de structure des sites. Il peut juste améliorer certains problèmes locaux et mineurs. Ce n'est pas non plus un "placebo" qui au pire serait inefficace : un PR sculpting mal fait peut nuire forcément à votre référencement.

Du PR Sculpting au PR Modeling

Cela ne veut pas dire que cela ne sert à rien d'améliorer la manière dont le "link juice" s'accumule et se diffuse au sein des pages d'un site. Bien au contraire, travailler ces points a potentiellement des effets sur la manière dont Google cawle (ou ignore) vos pages, et sur les positions acquises par vos pages. Mais cela veut dire que la meilleure manière de le faire n'est pas de "cacher" des liens que voient les internautes aux spiders, mais de changer radicalement la structure des liens qui relient les pages entre elles.

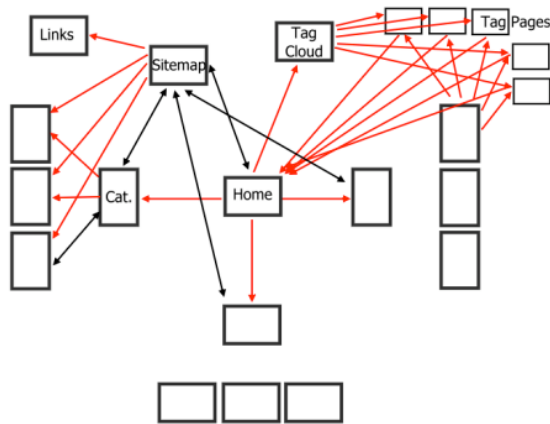
Pour qualifier les méthodes alternatives de refonte de la structure hypertexte des sites, j'ai choisi au cours de l'été 2007 d'utiliser le terme "PR modeling". Un terme qui se traduit par "modelage de PR". On est toujours dans la parabole du sculpteur. Sauf que la sculpture consiste à enlever de la matière avec un ciseau à bois ou un burin. Alors que le modelage d'une statue en argile consiste à prendre un bloc de matière et à changer sa forme sans avoir besoin de retirer de la matière.

Non, M. Jourdain, vous ne faites pas du PR modeling sans le savoir

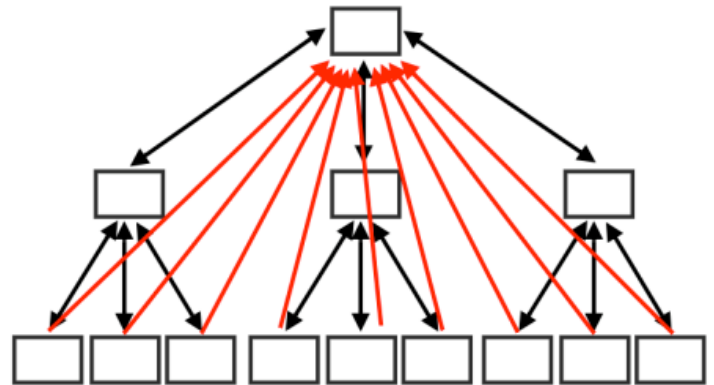
Dans le PR modeling, on ne change pas la liste des liens que l'on peut découvrir sur un site. Ce que l'on modifie c'est l'emplacement de ces liens, et la liste des liens figurant sur chaque template. On me rétorque souvent que créer une structure optimale est l'objectif de tout webmaster sensibilisé aux problèmes de SEO, et que les webmasters ou les référenceurs n'ont pas attendu que j'invente le terme pour faire du "PR modeling". Sauf qu'une fois de plus, comme les mécanismes à l'oeuvre derrière les calculs de PageRank ne sont pas toujours compris, il faut une véritable méthodologie pour améliorer cette fameuse structure, pas juste des intuitions : la plupart des sites ont des structures de liens assez loin de l'optimum. Pour s'en rendre compte, il faut disposer d'un outil capable de calculer le PageRank interne sur un grand nombre de pages. Hélas, rares sont les webmasters qui disposent ou ont développé un tel outil.

Par ailleurs, la structure optimale pour votre site n'est pas forcément celle du voisin. Pour un site d'annonces, les pages importantes pour le référencement sont plus les catégories d'annonces que les annonces elles-mêmes. Il faudra donc concentrer du PR sur les pages de catégorie, et moins sur les sites pages annonces. Sur un site éditorial, les priorités changent un peu : la catégorie "news" sera moins importante que les articles ou dossiers d'articles rassemblés dans cette catégorie. Or la plupart des sites, en raison des menus de navigation,

ont tendance à concentrer fortement le PR sur le haut de l'arborescence. Les pages profondes, mal irriguées en "link juice", risquent de ne plus être crawlées par Google, qui ne les estime pas suffisamment importantes. Bref, dans bien des cas, les structures "classiques" sont optimales pour certains types de sites précis, mais en règle générale, il faut s'éloigner de cette norme. L'une des méthodes les plus connues pour faire cela est le "siloining", c'est à dire la création d'un système de navigation privilégiant une classification par "univers" (de produits, de contenus) pour améliorer le référencement de blogs ou de forums. Le "siloining" est à ranger clairement dans les techniques de PR modeling.



Structure d'un blog avant « siloining »



...et après.

Quelques outils pour faire du PR modeling

1. L'Ajax

Pour le moment, les spiders ne savent pas crawler les contenus chargés en Ajax. Il est donc possible de charger une partie des liens à la demande de l'internaute, en utilisant la technologie Ajax. Cela permet par exemple de créer des menus dans lesquels les liens présents au départ sur la page sont les liens liés à l'univers et au contexte de la page. Si l'internaute cherche à changer d'univers ou de zone du site, une série de liens complémentaire sera chargée. Ces techniques relèvent du PR modeling dès lors que les liens chargés en Ajax peuvent par ailleurs être découverts par les spiders et à peu près dans l'ordre de priorité suggéré par la navigation visible sur la *home*. Mais cela évite de créer des pages comportant systématiquement des centaines de liens.

2. Remplacement de plusieurs pages par une page contenant des ancres

Si on place "contact", "mentions légales", et "qui sommes nous", sur une même page, les liens "[lesite.html#contact](#)", "[lesite.html#mentionslegales](#)" etc... seront tous considérés comme un seul lien vers "[lesite.html](#)". On conviendra que la méthode a ses limites...

3. L'iframe

Si on place les blocs de navigation importants dans une iframe, cela diminue le nombre de liens sortants sur les pages considérées et permet de transmettre plus de PR aux autres liens. La technique ne permet pas d'aller très loin dans les changements structurels, car l'augmentation du PR transmis est très locale.

4. Papier + crayon + outil de web analytics + réflexion sur l'ergonomie

D'une manière générale, la meilleure approche reste la création d'une navigation par liens entre les pages se concentrant sur les zones vers lesquelles on veut faire aller l'internaute en priorité, et en faisant disparaître les liens peu cliqués des templates.

Bibliographie

Pour approfondir les aspects mathématiques du PageRank :

Inside PageRank

Monica Bianchini, Marco Gori, and Franco Scarselli. Inside PageRank. ACM Transactions on Internet Technology, 5(1), 2005

PDF réservé aux abonnés à la revue ACM

Deeper Inside PageRank

Amy N. Langville, Carl D. Meyer, Department of Mathematics,
N. Carolina State University,

http://meyer.math.ncsu.edu/Meyer/PS_Files/DeeperInsidePR.pdf

Video de Matt Cutts :

What are your views on PageRank Sculpting ?

<http://www.youtube.com/watch?v=R4IE4WLPLZQ>

23 avril 2009

Transcript : <http://andybeard.eu/1865/pagerank-sculpting-dead.html>

Billet de Matt Cutts sur son blog

<http://www.mattcutts.com/blog/pagerank-sculpting/>

15 juin 2009

Article de Vanessa Fox sur SearchEngineLand (à propos de la prise en compte du javascript par Google)

Google I/O: New Advances In The Searchability of JavaScript and Flash, But Is It Enough?

<http://searchengineland.com/google-io-new-advances-in-the-searchability-of-javascript-and-flash-but-is-it-enough-19881>

Rand Fishkin sur SEOmoz

6 Juillet 2009

Link Consolidation: The New PageRank Sculpting

<http://www.seomoz.org/blog/link-consolidation-the-new-pagerank-sculpting>

29 août 2007

Matt Cutts on Nofollow, Links-Per-Page and the Value of Directories

<http://www.seomoz.org/blog/questions-answers-with-googles-spam-guru>

14 août 2008

PageRank Sculpting: Parsing the Value and Potential Benefits of Sculpting PR with Nofollow

<http://www.seomoz.org/blog/pagerank-sculpting-parsing-the-value-and-potential-benefits-of-sculpting-pr-with-nofollow>

Michael Martinez, <http://www.seo-theory.com/>

21 Janvier 2010

<http://www.seo-theory.com/2010/01/21/why-your-seo-science-lacks-credibility/>

29 Décembre 2009

A comment on SEOmoz' latest bogus PageRank sculpting test

<http://www.seo-theory.com/2009/12/29/a-comment-on-seomoz-latest-bogus-pagerank-sculpting-test/>

13 janvier 2010

Is Amazon.com cloaking : Serving Google different content

<http://www.seroundtable.com/archives/021504.html>

14 mars 2010

Matt Cutts Interviewed by Eric Enge

<http://www.stonetemple.com/articles/interview-matt-cutts-012510.shtml>

PageRank Sculpting - SMX East 2008 - A graphical tutorial to how NoIndex, NoFollow, and Robots.txt work

<http://www.stonetemple.com/presentations/smx-east-2008-pagerank-sculpting.ppt>

Philippe Yonnet, *Directeur Technique @Position* (<http://www.aposition.com>) et *président de l'association SEO Camp* (<http://www.seo-camp.org/>)

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

<http://blog-abonnes.abondance.com/2010/04/le-pagerank-modeling-avenir-du-pagerank.html>