

Le TrustRank, bien mystérieux "indice de confiance" de Google

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Google s'est fait connaître sur le Web au travers de son célèbre PageRank, indice de popularité calculé sur la base d'une analyse quantitative et qualitative des liens pointant vers une page web. Mais, depuis quelques années, on parle de plus en plus d'un autre indice, baptisé TrustRank, et qui servirait à Google pour mesurer la confiance que l'on peut avoir dans un site web donné, sur la base de critères humains et automatiques. Qu'en est-il exactement de ce fameux TrustRank, dont le nom vient en fait d'ingénieurs de chez Yahoo! ? Info ou intox ? Enquête pour en savoir plus...

En ce début d'année, nous allons revenir sur un des mystères récurrents du web, avec cet article consacré au TrustRank. Est-ce que ce système est toujours utilisé par Google ? Quelles en seront les évolutions probables en 2009 ? Commençons notre exploration...

Définition du TrustRank

On a commencé à entendre parler de ce nouveau critère de classement au travers d'un article rédigé en mars 2004 par deux chercheurs de l'université de Stanford et intitulé "*Combating web spam with TrustRank*" (<http://www.vldb.org/conf/2004/RS15P3.PDF>). Les chercheurs (qui travaillaient pour Yahoo) proposaient de créer une liste de sites de confiance (*trusted sites*) et d'accorder une attention particulière aux liens du web provenant de ces sites, partant de l'hypothèse qu'un lien issu d'un site de confiance pointe généralement vers un autre site de confiance.

Le TrustRank, au cœur de ce nouveau système, désigne ainsi l'**indice de confiance** accordé à un site web, et ce signal se propage d'un site à l'autre de façon décroissante. Plus on est "loin" du site de confiance initial (au sens du nombre de liens donc du nombre de clics nécessaire pour y arriver), plus le TrustRank diminue.

Pour la petite histoire, Google a déposé en 2005 le nom de marque TrustRank mais il n'y aurait pas de rapport direct avec le projet développé par les chercheurs de Yahoo. Dans une vidéo publiée sur Youtube en novembre 2007, on voit notamment Matt Cutts expliquer que le TrustRank de Google n'a rien à voir avec celui décrit dans l'article de Stanford (voir <http://fr.youtube.com/watch?v=p8mUXQzwEvs>).



Néanmoins, bien qu'il n'ait jamais été officiellement reconnu par Google, le TrustRank semble être au cœur de son algorithme : pour être bien classé dans Google, il est en effet important d'obtenir des liens depuis des sites "de confiance".

Ceci est confirmé dans plusieurs explications données dans l'aide en ligne Google :

Le classement de votre site dans les résultats de recherche Google est en partie basé sur l'analyse des sites qui comportent des liens vers vos pages. La quantité, la qualité et la pertinence de ces liens sont prises en compte pour l'évaluation de votre site. Les sites proposant des liens vers vos pages peuvent fournir des informations sur l'objet de votre site, et peuvent indiquer sa qualité et sa popularité

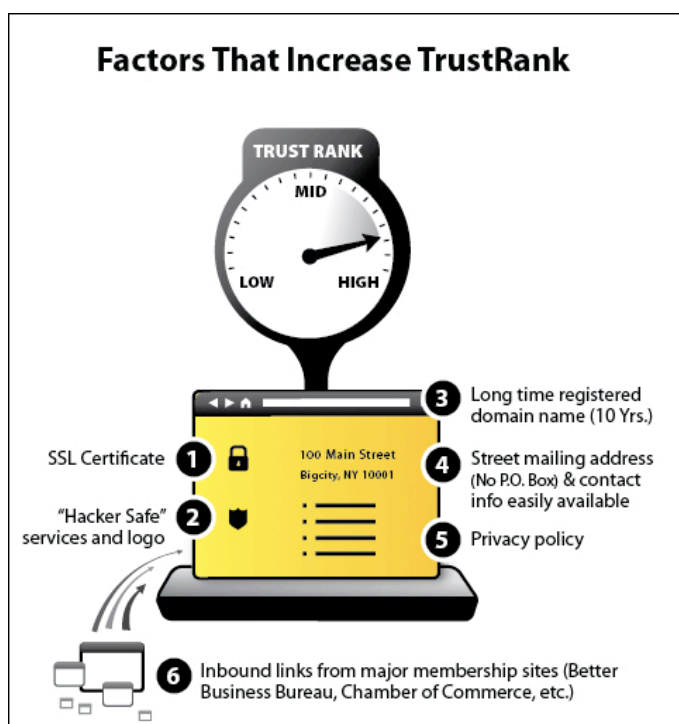
(<http://www.google.fr/support/webmasters/bin/answer.py?answer=66356&topic=15263>).

Le PageRank tient également compte de l'importance de chaque page qui "vote" et attribue une valeur supérieure aux votes émanant de pages considérées comme importantes. Les pages importantes bénéficient d'un meilleur classement PageRank et apparaissent en haut des résultats de recherche. (<http://www.google.com/corporate/tech.html>).

Quelques informations ont filtré ici et là sur la façon dont Google pourrait déterminer qu'un site est un site de confiance. Citons :

- Les données Whois,
- La durée d'enregistrement du nom de domaine,
- La politique relative à la vie privée,
- Les informations de contact,
- L'affichage de l'adresse postale de l'entreprise sur le site,
- La taille du site (en nombre de pages),
- Le trafic,
- Le niveau de sécurisation du site,
- D'éventuelles certifications apportées par des organismes officiels,
- L'ancienneté des liens acquis,
- Une note fournie par un être humain, chez Google, chargé de recenser un certain nombre de "sites incontournables" dans certains domaines thématiques (ce qui expliquerait, par exemple, l'omniprésence de Wikipedia sur le moteur de recherche).
- Etc.

La liste peut être longue. Ceci peut être résumé dans ce schéma proposé par le site <http://searchengineoptimization.elliance.com> :



Un brevet, déposé par Google, parle également d'une notion de "TrustRank" pour son site Google News, indiquant ces critères pour l'établissement d'un "indice de confiance" dans une source d'informations sur l'actualité :

1. le nombre d'articles produit par la source
2. la longueur moyenne des articles
3. la "couverture" de la source
4. la réactivité de la source ("*breaking score*")
5. un indice d'utilisation (en nombre de clics sur cette source)
6. une opinion humaine sur la source
7. une statistique extérieure d'audience telle que Media Metrix ou Nielsen Netratings
8. la taille de l'équipe
9. le nombre de bureaux ou agences différentes de la source
10. le nombre d'entités nommées originales citées par la source (personnes, organisations, lieux)
11. l'étendue (breadth), le nombre de sujets couverts par la source
12. la diversité internationale
13. le style de rédaction, en termes d'orthographe, de grammaire, etc.

Sources : Brevet "*Systems and methods for improving the ranking of news articles*" déposé par Google, blog "*Technologies du Langage*" <http://aixtal.blogspot.com/>

Conséquences du TrustRank

Le TrustRank, tout comme le PageRank, privilégie la qualité des liens plutôt que leur quantité... Les annuaires, fermes de liens, et autres techniques utilisant des liens artificiels ne favorisent plus le classement dans les moteurs de recherche : seuls les liens issus de sites de confiance comptent.

Ceci a particulièrement été visible lors du grand ménage effectué par Google fin 2007, lorsque de nombreux sites bien connus ont vu leur PageRank chuter de façon vertigineuse à cause la présence de liens jugés comme artificiels par Google.

On trouve des explications assez claires dans le guide Google sur les systèmes de liens (<http://www.google.fr/support/webmasters/bin/answer.py?answer=66356&topic=15263>) :

Le classement de votre site dans les résultats de recherche Google est en partie basé sur l'analyse des sites qui comportent des liens vers vos pages. La quantité, la qualité et la pertinence de ces liens sont prises en compte pour l'évaluation de votre site. Les sites proposant des liens vers vos pages peuvent fournir des informations sur l'objet de votre site, et peuvent indiquer sa qualité et sa popularité. Toutefois, certains webmasters participent à des systèmes d'échange de liens et conçoivent des pages partenaires utilisées exclusivement à des fins de "cross-linking", sans tenir compte de la qualité des liens, des sources et de l'impact à long terme que cela aura sur leur site. Cette pratique constitue une infraction aux conseils aux webmasters et peut avoir des conséquences négatives sur le classement de votre site dans les résultats de recherche.

Une autre conséquence sur le classement des résultats de recherche a été la fin du Google Bombing, début 2007. Pour mémoire, le Google Bombing consistait à créer un nombre massif de liens basés sur le même *anchor text*, de façon à faire positionner un site web de façon amusante ou décalée, voire sous la forme d'un "châtiment numérique". On peut citer par exemple le cas de la requête "[Nicolas Sarkozy](#)" qui renvoie encore aujourd'hui sur le site du film Iznogoud (en deuxième page de résultats) ou le très célèbre "[miserable failure](#)" pour George Bush (et bien d'autres).

Google a communiqué sur son blog officiel à propos de la fin du Google Bombing (voir <http://googlewebmastercentral.blogspot.com/2007/01/quick-word-about-googlebombs.html>) : l'amélioration de l'algorithme analysant la structure des liens a permis de supprimer un grand nombre de résultats "fantaisistes". Google a longtemps hésité à supprimer le Google Bombing,

du fait qu'il s'agissait d'une forme de vote par les internautes, et qu'il n'était pas question de censurer le web.

Il est donc certain que les liens issus de sites de confiance comptent de plus en plus dans Google et que la notion de TrustRank est plus que jamais d'actualité.

Le TrustRank sous toutes ses formes

Comme nous l'avons vu au début de cet article, le terme de TrustRank a été originellement développé par Yahoo et c'est un principe qui est en réalité utilisé par de nombreux moteurs. La notion de "site de confiance" est en effet à la base de nombreux algorithmes, et c'est une méthode de base pour identifier les sites les plus intéressants à présenter dans les résultats des moteurs de recherche.

Le moteur Ask.com utilise par exemple un "ExpertRank", qui mesure la popularité d'un site vis à vis de pages "expertes" (c'est à dire de sites de confiance) :

L'algorithme ExpertRank du moteur Ask assure la pertinence des résultats de recherche en identifiant les sites les plus fiables et les plus respectés sur le Web. Avec la technologie de recherche Ask, il ne s'agit pas d'être le plus grand : il s'agit d'être le meilleur. Notre algorithme ExpertRank ne s'arrête pas à la popularité des liens (c'est-à-dire au classement des pages en fonction du nombre brut de liens pointant vers une page particulière) pour mesurer la popularité des pages dites expertes sur un sujet de recherche donné. À cet effet, on parle de popularité thématique. L'identification des sujets (également nommés "clusters"), des pages expertes sur ces sujets et de la popularité des millions de pages les plus fiables en la matière – à l'instant précis où vous lancez votre recherche – demande de nombreuses analyses supplémentaires non pratiquées par les autres moteurs de recherche. Résultat : une pertinence inégalée proposant souvent une touche rédactionnelle unique par rapport aux autres moteurs de recherche. (<http://sp.fr.ask.com/fr/docs/about/asksearch.shtml>).

Autre outil de classement, un "BrowseRank" a été développé en juillet 2008 par les chercheurs de Microsoft. Cette fois-ci, la notion de popularité ne se base pas sur la qualité des liens entrants mais plutôt sur le comportement des visiteurs : temps passé sur le site, nombre de liens cliqués, nombre de visiteurs... (voir <http://actu.abondance.com/2008/07/browse-rank-larme-de-microsoft-pour.html>). Pour obtenir ces données stratégiques, Microsoft utilise la barre d'outils Live Search (et on imagine bien que Google fait de même avec sa propre barre d'outils).

Au final, le principe de "site de confiance" semble bien être utilisé par Microsoft mais repose cette fois sur les internautes eux-mêmes plutôt que sur l'environnement web.

Le TrustRank en 2009

Le TrustRank (s'il existe) est sans doute désormais une combinaison de nombreux facteurs. Cela fait longtemps que les moteurs de recherche ne se basent plus sur seulement le nombre de liens pointant vers un site mais prennent aussi en compte la qualité du contenu, l'origine des liens, le comportement des utilisateurs...



Le PageRank n'est pas le seul critère de notation...

(source : <http://blaugh.com/2007/06/11/link-popularity-vs-pagerank-vs-yoda/>)

La fin d'année 2008 a été également marquée par l'arrivée d'un élément important pour le référencement : l'apparition des résultats personnalisés dans Google, avec le système baptisé "SearchWiki" (voir <http://googleblog.blogspot.com/2008/11/searchwiki-make-search-your-own.html>).

Concrètement, ce système permet à chaque utilisateur possédant un compte Google de supprimer, enrichir ou commenter les résultats de recherche. Il est donc fortement probable que Google va tenir compte de l'avis des internautes pour le classement des résultats de recherche (*ceci a d'ailleurs été confirmé en partie par Marissa Mayer, voir notre article à ce sujet le mois dernier dans cette lettre*). En conséquence la notation TrustRank pour Google pourrait bien s'enrichir de la directive suivante : "un site est un site de confiance s'il a reçu beaucoup de votes de la part des utilisateurs".



Le vote des internautes : nouveau critère en 2009 ?

Ceci compliquera certainement le travail des webmasters : les techniques de triche seront de moins en moins fructueuses, pour faire de son site un site de confiance, il faudra désormais plaire aux internautes (même si on peut penser qu'ici aussi, la triche aura hélas sa place...)...

En 2009, un site de confiance ne sera pas forcément un site ayant été validé par d'autres sites de confiance mais sera aussi un site ayant obtenu un grand nombre de visiteurs et de votes grâce un buzz ou un marketing viral efficace sur le web.

L'autre évolution du web qui se profile depuis plusieurs années, la recherche universelle, risque également de changer la donne. En effet la notion de TrustRank risque de prendre en compte non seulement un site web, mais ses "produits dérivés" tels que les images, les vidéos, les documents...

Il est certain qu'assurer un positionnement sur différents médias permettra de profiter de la recherche universelle et il est possible qu'un site possédant par exemple beaucoup de vidéos indexées dans Youtube verra son TrustRank revu à la hausse. Une bonne stratégie sera alors de positionner des documents multimédias dans des sites de confiance, ceux-ci assurant ensuite un vote de qualité vers le site web principal.

Des portails comme Flickr par exemple, permettront sans doute d'augmenter son TrustRank. En effet il s'agit d'un site où les images peuvent être notées et commentées par les internautes (ce qui ajoute le critère de vote humain vu précédemment) et où le contenu est soumis à modération (voir à ce sujet les règles communautaires de Flickr : <http://www.flickr.com/guidelines.gne>).

C'est donc l'aspect humain qu'il faudra privilégier dans l'avenir : réaliser du contenu à destination des internautes, plutôt que penser son site pour les moteurs de recherche. Avec les nouvelles règles du TrustRank, un site en *full Flash* pourra par exemple avoir une chance plus importante de ressortir dans les moteurs...

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/2009/01/le-trustrank-bien-mystrieux-indice-de.html>