

SiteLinks : comment Google les choisit-il ?

[Retour au sommaire de la lettre](#)

Vous avez certainement remarqué, en dessous du premier résultat fourni par Google, quatre liens qui apparaissent parfois vers des zones internes du site en question. Ces informations, baptisées SiteLinks par le moteur de recherche, intriguent beaucoup de webmasters. Pourquoi certains sites en bénéficient-ils et pas d'autres ? Comment Google les choisit-il ? Voici les résultats de nos tests et de notre enquête...

Vous l'avez certainement remarqué : sur certaines requêtes, Google propose, en dessous du premier lien-résultat, quatre liens vers des zones internes du site en question. Exemple pour le mot clé "[abondance](#)" :



Ces liens sont appelés "SiteLinks" par Google. Le moteur de recherche a fait à ce sujet, dernièrement, une communication très légère, mais qui a le mérite d'exister de façon officielle, à l'adresse suivante :

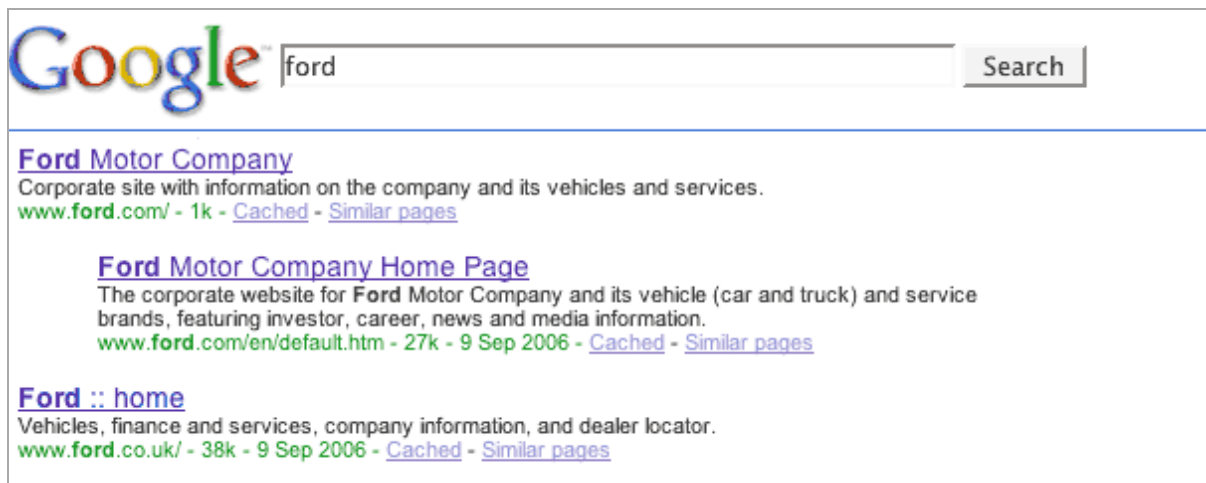
<http://www.google.com/support/webmasters/bin/answer.py?answer=47334&topic=8523>

Il est difficile d'en savoir plus sur ces liens, cependant, certains points sont clairs car provenant de constatations évidentes ou d'infos fournies par Google :

- Ces liens SiteLinks ne s'affichent que pour le premier lien de la page de résultats.
- Le fait qu'ils s'affichent ou pas (présence ou absence de SiteLinks) semble venir de deux critères essentiels : la pertinence (il est vraisemblable que les liens SiteLinks ne s'affichent que si le premier lien est "évident" par rapport à la requête demandée : présence du mot clé dans le nom de domaine et/ou taux de pertinence - valeur connue de Google mais non divulguée - supérieur à une certaine limite, etc.) et la qualité de la structure du site (voir plus loin). D'autres critères, comme le PageRank (les SiteLinks ne s'afficheraient que si la page d'accueil du site a un PageRank réel supérieur à une certaine "note"), sont aussi possibles, même si cela nous semble improbable.
- Les SiteLinks ne semblent s'afficher que pour des pages d'accueil de sites.
- Le système est entièrement automatisé (pas d'intervention humaine), même si Google envisage d'incorporer dans le futur des données fournies par les webmasters (certainement au travers de l'outil "Google SiteMaps").
- Selon Google, un critère important est la "structure du site", donc la façon dont les liens internes y sont proposés, sans qu'il soit possible d'en savoir plus sur la façon dont le moteur de recherche choisit tel lien plutôt que tel autre...
- Les SiteLinks peuvent pointer vers une page interne du site lui-même (www.votresite.com/repertoire/pageinterne.html) soit vers une page d'un sous-domaine (sousdomaine.votresite.com/repertoire/pageinterne.html) de ce même site.

- Les SiteLinks peuvent être au format texte, image (l'option "alt" fournit alors le texte affiché) ou JavaScript dans la page d'accueil.

- Attention aux redirections : exemple sur la requête "ford" sur Google.com :



La première page identifiée (www.ford.com), considérée comme très pertinente par Google certainement par le fait que de nombreux liens pointent vers elle, est un document d'1 ko contenant une redirection JavaScript - donc non détectée par Google - vers la page du dessous (www.ford.com/en/default.htm). Comme la première page ne contient pas de liens, il a été impossible à Google de déterminer quels SiteLinks il pourrait proposer. Du coup, rien n'est affiché...

- Il ne semble pas que Google tienne compte d'informations incluses dans le code HTML de la page d'accueil qui bénéficie des SiteLinks. En effet, pourquoi, dans ce cas, choisirait-il les liens "Audit", "Outils de recherche" ou "Blog" plutôt que d'autres, traités de la même façon, dans la page d'accueil du site Abondance.com ? Ce ne sont pas non plus les liens proposés en premier dans le code source, etc.

- Il ne s'agit pas non plus des résultats qui ressortent en premier sur l'utilisation de la requête "site:" pour n'obtenir que les pages issues du site en question. Exemple avec la requête [abondance site:abondance.com](http://www.google.com/search?q=site:abondance.com) ou [site:abondance.com](http://www.google.com/search?q=site:abondance.com). Les pages proposées par Google dans ces deux cas n'ont pas grand chose à voir avec celles affichées comme SiteLinks...

- Une hypothèse intéressante et crédible pourrait être la suivante : Google utilise les données renvoyées par les Google Toolbars de ses utilisateurs pour déterminer quels liens sont le plus souvent cliqués par les internautes lorsqu'ils se trouvent sur la page d'accueil en question. Selon le site Social Patterns (<http://www.socialpatterns.com/search-engine-marketing/traffic-determines-google-ui-snippet-links/>), il existerait de fortes similitudes entre les liens affichés dans les SiteLinks et les données de trafic de l'outil Alexa, comparables à ce dont Google pourrait se servir... A la lumière de cette information, regardez à nouveau les liens proposés en SiteLinks et regardez les pages d'accueil des sites proposés : n'avez-vous pas l'impression que les SiteLinks sont les liens sur lesquels "vous seriez le plus tentés de cliquer" ??

De plus cette hypothèse résoudrait de nombreuses interrogations : pourquoi les SiteLinks seraient-ils trouvés par Google lorsqu'ils sont issus de liens JavaScript alors que ses spiders ont encore du mal à suivre ce type de lien ? Cela résoudrait également le cas du site Ford (voir ci-dessus). Une page de redirection ne peut avoir ce type de statistiques sur les clics à partir d'elle. Etc.

Notre hypothèse est donc la suivante :

- **Le choix d'afficher les SiteLinks ou non** est basé pour Google sur quatre critères principaux :

1. Adéquation entre la requête demandée et le site, et notamment son nom de domaine (exemple pour la requête "[référencement 2000](http://www.google.com/search?q=r%C3%A9f%C3%A9rencement+2000)"). Si le nom de domaine du premier résultat affiché contient

les mots de la requête, il y a de fortes chances que ce site soit affublé de SiteLinks. Ce n'est cependant pas vrai à 100% car la requête [time magazine](http://www.time.com) propose le site www.time.com...

2. Le premier résultat proposé doit être une page d'accueil (www.siteweb.com).
3. Le taux de pertinence du site par rapport à la requête (donnée non communiquée par Google mais dont nous pouvons attester de l'existence) doit être supérieur à une certaine limite, malheureusement inconnue.
4. Des liens doivent être détectables au sein de cette page d'accueil.

- **Le choix des SiteLinks eux-même** s'effectuerait alors grâce à 2 critères principaux :

1. Liens pointant vers une page interne du site ou d'un de ses sous-domaines (pas de lien externe).
2. Nous pensons fortement, au vu de nombreux exemples analysés, que ces liens sont fournis par des statistiques issues de la barre d'outils de Google. Il semble s'agir des liens le plus souvent cliqués sur la page en question par les internautes ayant installé la barre d'outils de Google sur leur navigateur.

Rien ne nous permet d'être affirmatif à 100% sur ces informations. Mais disons qu'il s'agit là d'une "intime conviction" :-)

Bien sûr, toute information, observation ou suggestion supplémentaire à ce sujet de votre part sera la bienvenue...