

Google : connaissez-vous ces trucs et astuces de recherche ? (1ère partie)

Google, vous connaissez, bien sûr... Mais le connaissez-vous si bien que cela ? Nous allons vous proposer, dans les pages qui suivent, une liste de petits "trucs et astuces" souvent bien utiles pour mieux utiliser le moteur leader.

Ce mois-ci, nous allons nous cantonner à la recherche web. Le mois prochain, nous étudierons les Groupes (forums), la recherche d'images, le répertoire, etc.

Syntaxe de recherche

Elle est assez connue mais revenons quand même rapidement sur la syntaxe avancée de Google, souvent assez pratique pour accélérer ses recherches sur le Web.

Syntaxe	Signification	Exemple
Espace	ET (par défaut)	moteur recherche est équivalent à moteur ET recherche (mais l'opérateur ET n'est pas pris en compte par Google)
OR (obligatoirement en majuscules)	OU	google OR altavista
Guillemets	Expression. Les mots demandés les uns à côté des autres et dans cet ordre.	"moteur de recherche"
Signe -	SAUF. Le signe - doit obligatoirement être précédé d'un espace.	moteur -automobile
Parenthèses	Elles sont acceptées par Google dans les expressions booléennes.	(moteur voiture) OR (moteur automobile)
intitle:	Recherche le mot qui suit dans les titres des pages.	intitle: référencement
allintitle:	Recherche tous les mots qui suivent dans les titres des pages.	allintitle: référencement paris
site:	Recherche sur un nom de domaine ou un domaine précis.	apple site: org apple site: microsoft.com apple site: www.microsoft.com
link:	Recherche toutes les pages ayant mis en place un lien vers l'adresse indiquée.	link: http://www.abondance.com/ link: www.abondance.com link: http://docs.abondance.com/oa.ht link: www.abondance.com/oa.html
related:	Recherche des sites similaires à celui dont l'adresse est fournie.	related: http://www.google.com/ related: www.google.com
inurl:	Recherche le mot demandé dans l'url complète des pages.	inurl:faq
allinurl:	Recherche tous les mots suivants dans l'url complète des pages.	allinurl: chimie faq
cache:	Affiche la page dont l'adresse est fournie, dans sa version présente dans le cache de Google.	cache: www.abondance.com

daterange:	Recherche sur les documents dont la date de dernière modification se situe dans une fourchette de dates (au format Julien). Voir un convertisseur de dates juliennes (assez indispensable pour utiliser cette fonctionnalité) ici : http://www.tesre.bo.cnr.it/~mauro/JD/	marée noire daterange:2452640-2452779
filetype:	Recherche sur le type de fichiers (liste des formats à l'adresse http://www.google.com/help/faq_filetypes.html)	confidentiel filetype:ppt
intext:	Recherche dans le texte de la page (exclut le titre, l'en-tête et les textes des liens).	intext:chirac
inanchor:	Recherche dans le texte des liens. Cette fonction ne semble pas marcher actuellement.	inanchor:kelkoo
info:	Fournit un certain nombre d'informations sur la page dont l'adresse est fournie sous la forme de liens correspondants aux fonctions link:, cache:, etc.	info:www.abondance.com
phonebook:	Recherche d'un numéro de téléphone (US uniquement).	phonebook:smith ca
rphonebook:	Recherche d'un numéro de téléphone - Pages Blanches ("Residential" - US uniquement).	rphonebook:smith ca
bphonebook:	Recherche d'un numéro de téléphone - Pages Jaunes ("Business" - US uniquement).	bphonebook:smith ca
stocks:	Recherche d'informations boursières sur une société cotée.	stocks:msft (code de la société Microsoft)

Notons que la fonction spell: (recherche d'un mot dans le dictionnaire), parfois évoquée sur le Web dans des articles parlant du moteur de recherche, semble ne pas fonctionner... En revanche, sur le site américain (uniquement), lorsque vous tapez un mot clé (exemple : "search engine"), dans la barre bleue, en haut de page, le message suivant apparaît : **Searched the web for search engine**. Si vous cliquez sur le mot recherché, vous avez accès à sa définition, grâce à un partenariat conclu entre Google et le site Dictionary.com.

Attention : Il semblerait que Google ait parfois quelques soucis selon l'ordre dans lequel on lui soumet les diverses syntaxes de recherche. Ainsi, si la requête **inurl:faq intitle:chimie** fonctionne bien, les résultats semblent bien moins évidents lorsque l'on inverse les deux termes : **intitle:chimie inurl:faq**. Bizarre... Selon les tests que nous avons effectué lors de l'écriture de cet article, cela marchait parfois, et parfois non. "Order matters", comme le dit Google plus loin dans ces paragraphes... ;-)

Notons également que les guillemets ne sont pas la seule possibilité pour demander un expression sur Google : certains caractères sont reconnus par le moteur comme "connecteurs" : tiret (-), "slash" ou barre oblique standard (/), point (.), signe égal (=).

Ainsi, toutes ces requêtes sont équivalentes :

"moteur de recherche"
moteur-de-recherche
moteur/de/recherche
moteur.de.recherche
moteur=de=recherche

Dernier point : sur le site américain, lorsque vous tapez une adresse, Google vous propose un plan d'accès (partenariat avec Yahoo! Maps / Mapquest). Exemple :

<http://www.google.com/search?q=165+University+Ave+Palo+Alto+CA&btnG=Google+Search>

Tableaux des affinités syntaxiques

Les différentes syntaxes de Google souffrent, pour la plupart d'entre elles, de certaines "difficultés sociales à vivre en bonne communauté". En d'autres termes, vous ne pouvez utiliser certaines fonctions avec d'autres, et vice-versa. Nous avons tenté d'indiquer, dans le tableau suivant, qui fonctionne (ou ne fonctionne pas) avec qui :

La syntaxe :	
intitle:	<ul style="list-style-type: none"> - Fonctionne seule (intitle:abondance) - Lorsqu'un mot clé lui est adjoint (référencement intitle:abondance ou intitle:abondance référencement), la recherche dans le titre semble ne plus s'effectuer. Les 2 mots sont recherchés dans le corps de la page. - Fonctionne avec les fonctions inurl: (intitle:google inurl:faq), daterange: (intitle:google daterange:2452640-2452779), filetype: (intitle:google filetype:pdf), intext: (intitle:google intext:brin). - Ne fonctionne pas avec les fonctions site:, link:, related:, cache:, inanchor:, info:.
allintitle:	<ul style="list-style-type: none"> - Fonctionne seule (allintitle:abondance google) - Lorsqu'un mot clé ou une autre fonction syntaxique lui est adjoint (référencement allintitle:google pagerank), la recherche semble ne plus fonctionner. - Ne fonctionne pas avec les fonctions site:, inurl:, link:, daterange:, filetype:, related:, cache:, inanchor:, info:, intext:.
site:	<ul style="list-style-type: none"> - Ne fonctionne pas seule (site:www.abondance.com), un mot clé doit lui est adjoint (référencement site:www.abondance.com). Le mot clé peut être placé avant (référencement site:www.abondance.com) ou après (site:www.abondance.com référencement) la fonction site:. - Fonctionne avec la fonction intext: (intext:google site:www.abondance.com). - Ne fonctionne pas avec les fonctions intitle:, inurl:, daterange:, filetype:, link:, related:, cache:, inanchor:, info:.
link:	<ul style="list-style-type: none"> - Fonctionne seule (link:www.abondance.com) - Ne fonctionne pas lorsqu'un mot clé lui est adjoint (google link:www.abondance.com) - Ne fonctionne pas avec les fonctions intitle:, inurl:, daterange:, filetype:, related:, cache:, inanchor:, info:, intext:.
related:	<ul style="list-style-type: none"> - Fonctionne seule (related:www.abondance.com) - Ne fonctionne pas lorsqu'un mot clé lui est adjoint (moteur related:www.abondance.com) - Ne fonctionne pas avec les fonctions intitle:.

	<p>inurl:, daterange:, filetype:, link:, related:, cache:, inanchor:, info:, intext:.</p>
inurl:	<ul style="list-style-type: none"> - Fonctionne seule (inurl:faq) - Fonctionne lorsqu'un mot clé lui est adjoint (moteur inurl:faq). Le mot clé peut être placé avant (moteur inurl:faq) ou après (inurl:faq moteur) la fonction inurl: . Attention, les résultats seront alors différents. - Fonctionne avec les fonctions daterange: (inurl:faq daterange:2452640-2452779), intext: (inurl:faq intext:google). - Ne fonctionne pas avec les fonctions intitle:, inurl:, daterange:, filetype:, link:, related:, cache:, inanchor:, info:.
allinurl:	<ul style="list-style-type: none"> - Fonctionne seule (allinurl:faq google) - Ne fonctionne pas lorsqu'un mot clé lui est adjoint (allinurl:faq google) - Ne fonctionne pas avec les fonctions intitle:, inurl:, daterange:, filetype:, link:, related:, cache:, inanchor:, info:, intext:.
cache:	<ul style="list-style-type: none"> - Fonctionne seule (cache:www.abondance.com) - Ne fonctionne pas lorsqu'un mot clé lui est adjoint (référencement cache:www.abondance.com) - Ne fonctionne pas avec les fonctions intitle:, inurl:, daterange:, filetype:, link:, related:, cache:, inanchor:, info:, intext:.
daterange:	<ul style="list-style-type: none"> - Ne fonctionne pas seule (daterange:2452640-2452779), un mot clé doit lui est adjoint (référencement daterange:2452640-2452779). Le mot clé peut être placé avant (référencement daterange:2452640-2452779) ou après (daterange:2452640-2452779 référencement) la fonction daterange: . - Fonctionne avec la fonction inurl (inurl:faq daterange:2452640-2452779), intext: (intext:google site:www.abondance.com). - Ne fonctionne pas avec les fonctions intitle:, filetype:, link:, related:, cache:, inanchor:, info:.
filetype:	<ul style="list-style-type: none"> - Ne fonctionne pas seule (filetype:pdf), un mot clé doit lui est adjoint (référencement filetype:pdf). Le mot clé peut être placé avant (référencement filetype:pdf) ou après (filetype:pdf référencement) la fonction filetype: . - Fonctionne avec les fonctions intitle: (intite:google filetype:pdf), inurl (inurl:faq filetype:pdf), intext: (intext:google filetype:pdf). - Ne fonctionne pas avec les fonctions link:, related:, cache:, inanchor:, info:.
intext:	<ul style="list-style-type: none"> - Fonctionne seule (intext:google) - Fonctionne lorsqu'un mot clé lui est adjoint

	<p>(faq intext:google). Le mot clé peut être placé avant (faq intext:google) ou après (intext:google faq) la fonction intext:. Attention, les résultats seront alors différents.</p> <p>- Fonctionne avec les fonctions site: (intext:google site:www.abondance.com), intitle (intext:google intitle:faq), daterange: (intext:faq daterange:2452640-2452779), inurl (intext:google inurl:faq), filetype (intext:google filetype:pdf).</p> <p>- Ne fonctionne pas avec les fonctions link:, related:, cache:, inanchor:, info:.</p>
info:	<p>- Fonctionne seule (info:www.abondance.com)</p> <p>- Ne fonctionne pas lorsqu'un mot clé lui est adjoint (référencement cache:www.abondance.com)</p> <p>- Ne fonctionne pas avec les fonctions intitle:, inurl:, daterange:, filetype:, link:, related:, cache:, inanchor:, info:, intext:.</p>

Tableau récapitulatif

Affinités syntaxiques	seule	avec mot clé	intitle:	allintitle:	site:	link:	related:	inurl:	allinurl:	cache:	daterange:	filetype:	intext:	info:
intitle:	X							X			X	X	X	
allintitle:	X													
site:		X											X	
link:	X													
related:	X													
inurl:	X	X	X								X	X	X	
allinurl:	X													
cache:	X													
daterange:		X	X					X					X	
filetype:		X	X					X					X	
intext:	X	X	X		X			X			X	X		
info:	X													

Nota : le tableau ci-dessus a été dressé à partir de nos propres tests. Parfois, il nous a semblé que certains couples de fonctionnalités n'avaient qu'une affinité partielle. La syntaxe demandée ne fonctionnait pas parfaitement. Nous avons quand même admis, dans la plupart des cas, que l'affinité, alors, était positive. Si vous trouvez des "contre-indications" à ce tableau, merci de nous les faire parvenir, nous le mettrons à jour périodiquement si cela s'avère nécessaire.

Troncature

Rappelons que Google ne permet pas la troncature ("Tout ce qui commence par..."). En revanche, l'astérisque permet de remplacer un mot entier.

Exemple : "george * bush" (avec les guillemets) trouvera les occurrences George Bush, George W Bush, Georges W. Bush, etc.

Cela peut, par exemple, vous permettre de trouver toutes les pages d'un site web incluses dans l'index de Google par la syntaxe :

site:www.abondance.com inurl:"*.html"

Si toutes les pages du site se terminent par la suffixe ".html", bien sûr...

Notons qu'un astérisque (*) remplace un mot, deux astérisques (**) deux mots, trois astérisques (***) trois mots, etc.

Dans un sens ou dans l'autre ?

On le sait, l'accentuation et la casse des lettres n'ont pas d'importance sur Google :

bébé = bébé = bebe

IBM = Ibm = ibm

En revanche, Google tient compte de l'ordre des mots demandés. Si vous saisissez l'expression **paris dakar**, il privilégiera les pages qui contiennent le mot **paris** avant le mot **dakar** dans son algorithme de classement. Une page contenant la phrase "Je suis allé de Dakar à Paris par le vol de 12h57" sera donc mal classée pour l'expression **paris dakar**, mais elle devrait être, en revanche, bien positionnée sur la requête **dakar paris**.

Tenez donc compte de ce point lors de la saisie de vos requêtes et saisissez toujours la version la plus "logique", la plus "vraisemblable" pour vos requêtes à plusieurs mots clés. Si vous cherchez des renseignements sur le rallye Paris-Dakar, pourquoi donc taperiez-vous cette requête à l'envers ?

Le fait de prendre en compte l'ordre des mots est assez logique et assez fin et intelligent dans le cadre d'une recherche web. Bravo Google ! En revanche, un point plus étonnant concerne le nombre de résultats fournis dans les deux cas. Sur Google France (<http://www.google.fr/>), on obtient les résultats suivants :

Requête **paris dakar** : 206 000 résultats

Requête **dakar paris** : 204 000 résultats

Si on demande les mêmes mots, pourquoi n'obtient-on pas le même nombre de résultats ? Il semble y avoir plusieurs réponses possibles à cette question :

- Le résultat fourni n'est qu'une approximation, Google ne se "donnant pas la peine", assez logiquement puisqu'il ne donne accès qu'aux mille premiers liens, de calculer à l'unité près le nombre exact de pages contenant les mots demandés. Donc, il y a peu de chances que 2 approximations donnent le même résultat, même si les données de départ sont identiques.
- Comme sur beaucoup de moteurs de recherche, il se pourrait que seule une partie de l'index est interrogée sur Google pour une requête donnée, pour gagner du temps. Il suffit que ce ne soit pas la même partie qui soit interrogée dans les 2 cas pour que les résultats diffèrent.
- Google utilise peut-être un algorithme de recherche à plusieurs niveaux. Par exemple : "je prends les pages qui contiennent les mots demandés dans les 1000 premiers caractères, puis, à l'intérieur de celles-ci, je classe mieux les pages qui contiennent les mots demandés dans le titre, puis je prends en compte le PageRank, etc." Du coup, l'ordre des mots peut devenir important et certaines pages être prises en compte dans un cas ou dans l'autre, et rejetées éventuellement...

Pour en avoir le coeur net, nous avons posé la question à Google, avec force détails et exemples pour bien leur faire comprendre notre interrogation. Voici leur réponse : "*The number of results for 'paris dakar' and 'dakar paris' differs because the order of the words in your query affects the search results. Order matters.*"

Ah, ok, ok, ok... "Le nombre de résultats est différent sur Google parce que l'ordre des mots est important pour Google". Ah ben, vous voyez, j'y aurais pas pensé tout seul. Merci Google :-))

Bref, pas d'explications officielles de la part de Google (vu que j'avais mis 3 semaines à recevoir la réponse précédente, je n'ai pas osé insister :-)), mais on peut penser que l'une des trois hypothèses proposées ci-dessus (voire une conjonction de plusieurs d'entre elles) apporte une réponse plausible... Dans tous les cas, le fait que la réponse puisse être considérée comme un dysfonctionnement du moteur de recherche explique certainement la formule lapidaire que nous avons reçue en réponse à notre question...

Forcer la présence d'un mot vide ou rechercher une occurrence accentuée exacte

Google ne prend pas en compte les "stop words" ou "mots vides" comme "de", "la", "le", etc. Exemple : sur la requête "moteur de recherche", le message "*de*" étant un mot très courant, il a été ignoré lors de cette recherche apparaît en haut de page.

Le signe + permet de forcer la recherche à prendre en compte ce mot vide : **moteur +de recherche**. La signe + doit être indiqué devant chaque mot vide que l'on désire voir pris en compte par le moteur : **+la femme +du boulanger +de Marseille**.

Idem pour les mots sur une seule lettre qui sont normalement ignorés par Google : **Star Wars Episode +I**

De la même façon, une recherche sur toutes les pages qui contiennent l'adresse www.abondance.com se fera par la requête : **" +www.abondance.+com "** (*www* et *com* sont, sinon, ignorés par Google).

D'autre part, le signe + sert également à rechercher une occurrence unique d'un mot, notamment accentué. La requête **+électricité** ne cherchera que cette occurrence (avec les deux accents aigus) alors que, normalement, Google ne tient pas compte de l'accentuation des mots.

Faire s'ouvrir les résultats dans une fenêtre indépendante

Lorsque vous utilisez Google et que vous cliquez sur des liens que propose le moteur dans sa page de résultats, les documents désirés s'affichent par défaut dans la même fenêtre du navigateur. Mais les préférences du moteur (<http://www.google.fr/preferences?hl=fr>) vous permettent de les afficher dans une fenêtre indépendante de façon automatique, au travers de l'option "*Fenêtre des résultats / Montrer les résultats de recherche dans une nouvelle fenêtre de navigateur.*"

Une fois cette option prise en compte, chaque fois que vous cliquerez sur un des liens de la page de résultats du moteur, cette dernière restera rapidement disponible quoi qu'il arrive, "*à portée de souris*". Essayez, il est difficile de ne plus s'en servir une fois qu'on y a goûté... et qu'on a pris la discipline de refermer les fenêtres une à une et au fur et à mesure... ;-))

Afficher plus de liens-réponse en une seule fois

Autre fonction inéressante des Préférences du moteur (<http://www.google.fr/preferences?hl=fr>) : le choix "*Nombre de résultats / L'affichage par défaut (10 résultats) est le plus rapide. Afficher XXX résultats par page.*"

Si vous avez une connexion haut débit, n'hésitez pas à demander 30, 50, voire 100 liens par page de résultats, vous gagnerez du temps à ne pas passer d'une page à l'autre, surtout vous faites partie des internautes qui ne se contentent pas des 10 premiers liens fournis par le moteur. L'affichage est légèrement plus lent dans ce cas, mais reste cependant tout à fait acceptable !

Effectuer des recherches bilingues ou trilingues

La recherche avancée de Google (http://www.google.fr/advanced_search?hl=fr) propose un menu déroulant permettant de limiter sa recherche à telle ou telle langue. Très pratique. Mais comment faire lorsqu'on désire faire des recherches en français ET en anglais ? ou en espagnol Et en italien ? Etc.

La encore, les Préférences du moteur (<http://www.google.fr/preferences?hl=fr>) vont nous aider au travers de la fonction "*Rechercher uniquement les pages écrites dans la ou les langues suivantes*".

Cochez les langues que vous désirez et Google vous "*fabrique*" une interface de recherche spécifique, dans laquelle un nouveau bouton radio, nommé "*Personnalisé*" va être créé aux côtés de ceux qui existaient déjà auparavant : "*Web*", "*Pages Francophones*" et "*Pages : France*".

Au mois prochain pour des astuces de recherche au sujet des autres outils de Google : Groupes (forums), Images, News, etc.