

Requêtes moteur et "stop words"

[Retour au sommaire de la lettre](#)

Nous vous proposons chaque mois une série de "trucs et astuces" qui devraient vous simplifier la vie dans le cadre de vos recherches sur le Web. Ce mois-ci : on a souvent tendance à dire que les moteurs de recherche ignorent les "mots vides" ou "stop words" comme "le", "de" ou "et". Pourtant, une requête sur "moteur de recherche" donne un résultat différent de "moteur recherche" sur la plupart des moteurs. Pourquoi ?

On a souvent tendance à dire que la plupart des moteurs de recherche ignorent les "stop words" comme, en anglais : "the", "a", "of", etc. ou en français : "le", "la", "un", "de", "et", etc.

Et c'est vrai puisque Google, par exemple, nous le dit : sur la requête "moteur de recherche" (<http://www.google.fr/search?q=moteur+de+recherche>), il nous explique que :
"**"de" étant un mot très courant, il a été ignoré lors de cette recherche**".

Dans l'aide en ligne du moteur, il est ainsi expliqué que :
"*Google ignore les chaînes de caractères dont le poids sémantique est trop faible (également désignés "mots vides" ou "bruit") : le, la, les, du, avec, vous, etc., mais aussi des mots spécialisés tels que "http" et ".com" et les lettres/chiffres d'un seul caractère, qui jouent rarement un rôle intéressant dans les recherches et risquent de ralentir notablement le processus.*"

On pourrait donc logiquement s'attendre à ce qu'une requête sur les expressions "moteur de recherche" et "moteur recherche" donnent les mêmes résultats. Eh bien non. Faites le test :

Requête "moteur de recherche" :

<http://www.google.fr/search?q=moteur+de+recherche>

Requête "moteur recherche" :

<http://www.google.fr/search?q=moteur+recherche>

S'il y a un certain recouvrement entre les deux pages de résultats, elles ne sont pas identiques, loin de là...

Alors, pourquoi cette différence ?

Eh bien, cela semble venir du fait que Google tient compte de la proximité des mots entre eux dans son algorithme de pertinence. Par exemple, sur la requête "moteur de recherche", Google ne tient pas compte du "de" mais "se souvient" tout de même qu'il existe un mot entre les deux termes. Alors que sur la requête "moteur recherche", les pages qui contiennent ces deux mots l'un à côté de l'autre seront mieux positionnées, toutes choses égales par ailleurs, que celles qui contiennent l'expression "moteur de recherche"...

Pour être plus clair, raisonnons sur un exemple : sur l'expression "franklin roosevelt" (<http://www.google.fr/search?q=franklin+roosevelt>), la majorité des pages identifiées comme répondant à la requête contiennent le nom ainsi orthographié : "Franklin Roosevelt". Insérons maintenant un "stop word" "bidon" entre les deux termes et lançons la requête "franklin le roosevelt" (<http://www.google.fr/search?q=franklin+le+roosevelt>). Résultat : la plupart des pages contiennent le nom différemment orthographié, sous la forme "Franklin *quelque-chose* Roosevelt"... Google s'est donc souvenu que la requête était sur trois termes, même si le deuxième n'a pas été pris en compte. Et ça change tout au niveau des résultats...

Vous voulez une autre démonstration ? Tapez la requête "franklin * roosevelt" (http://www.google.fr/search?q=franklin+*+roosevelt) et vous obtiendrez quasiment la même réponse que pour "franklin le roosevelt". Rappelons que l'astérisque, sur Google, permet de remplacer un mot, quel qu'il soit, dans une requête. Là encore, le moteur s'est souvenu que la requête s'effectuait sur trois termes, le premier et le dernier seulement étant pris en compte...

Comment faire ?

Comment faire, alors, pour que Google prenne en compte le "stop word" s'il vous semble important pour votre recherche ? Il existe deux façons de le faire : soit avec les guillemets, soit avec le signe "+".

- Les guillemets vont vous permettre d'effectuer la requête ""moteur de recherche"" (<http://www.google.com/search?q=%22moteur+de+recherche%22>), les trois mots dans cet ordre et les uns à côté des autres. Dans ce cas, Google prend bien en compte le mot vide dans son algorithme.

- Le signe + vous permet, à l'aide de la requête "moteur +de recherche" (<http://www.google.com/search?q=moteur+%2Bde+recherche>) d'inclure de façon obligatoire le mot vide dans la recherche. En revanche, par rapport à l'exemple ci-dessus (guillemets), les mots ne seront pas obligatoirement dans cet ordre et les uns à côté des autres dans le résultat. L'option "guillemets" est donc plus restrictive...

Et les autres moteurs ?

Sur Yahoo! et MSN, une recherche avec "stop word" sera différente d'une recherche sans... Cependant, aucun des deux moteurs n'indique s'il prend ou non en compte les "stop words". Aucun message du type ""le" étant un mot trop commun, etc." n'est affiché. A notre avis, aucun des deux n'a mis en place ce type de filtre... Cependant, là aussi, l'utilisation de guillemets vous permettra sans problème d'inclure de façon sûre vos "mots vides" favoris dans vos recherches...