



Le formulaire de recherche de Google

Google est né d'un programme d'analyse d'hyperliens, le célèbre Page Rank (<http://www.google.com/intl/fr/profile.html>). Les opérateurs de recherche ont été greffés à ce programme et cette greffe a échoué. En effet, il y a discordance entre l'usage des opérateurs logiques de la case de recherche de la page d'accueil et le formulaire.

Les opérateurs de recherche comprennent les opérateurs logiques, de proximité et les opérateurs linguistiques. Ils sont représentés par des libellés dans un formulaire de recherche, par des symboles et par des mots dans une recherche libre. Toutes ses représentations donnent le même résultat, c'est, du moins, ce à quoi nous nous attendons.

Google fonctionne différemment selon que l'on utilise la case de recherche de la page d'accueil et le formulaire. Le sujet choisi pour le démontrer est *l'implication des États-Unis et de la Grande-Bretagne, mais non du Canada, dans la guerre en Irak*. Ce sujet a été amplement traité depuis 2001.

Voici le sujet traduit en requête dans le formulaire comprenant les deux expressions **Great Britain** et **United Kingdom** :

Find results	with all of the words	war Irak
	with the exact phrase	United States of America
	with at least one of the words	"Great Britain" "United Kingdom"
	without the words	Canada

La case de l'expression exacte comprend **United States of America**. Dans celle d'au

moins un mot, les deux noms représentant la Grande Bretagne - ils peuvent être ensembles ou seuls dans un texte - sont entre guillemets anglais conformément à l'aide de Google. Nulle part n'y a-t-il un avis contraire pour le formulaire.

Le programme traduit cette requête en ignorant les guillemets anglais et en introduisant l'opérateur **OR** entre les parties composant les expressions de départ. La requête traduite n'est plus conforme à l'originale.

A screenshot of the Google search interface. The search bar contains the text "war Irak "Great OR Britain" OR "United OR King". Below the search bar, there are radio buttons for "the web" (selected) and "pages from Canada". Navigation links for "Web", "Images", "Groups", "News", "Local", and "more" are visible at the top.

Cette requête donne environ 10 000 notices.

En retournant dans le formulaire, nous constatons que le logiciel de recherche a retiré les guillemets anglais.

Find results	with all of the words	war Irak
	with the exact phrase	United States of America
	with at least one of the words	Great Britain United Kingdom
	without the words	Canada

Il est écrit dans l'aide :

«Certains caractères sont reconnus par Google comme connecteurs d'expression : tiret (-), barre oblique standard (/), point (.), signe égal (=). »

Les guillemets sont remplacés par le tiret ou le trait d'union entre les parties de l'expression afin de vérifier s'il est plus efficace.

Find results	with all of the words	war Irak
	with the exact phrase	United States of America
	with at least one of the words	Great-Britain United-Kingdom
	without the words	Canada

Le résultat de la recherche est maintenant d'environ 800 notices.

Voici la traduction de la requête par Google :



Search: the web pages from Canada

L'opérateur d'union **OR** est maintenant bien placé. **Great Britain** et **United Kingdom** sont recherchés chacun en un seul mot. Google a placé **United States of America** entre guillemets. Nous observons aussi que la syntaxe de la requête est différente de celle illustrée par le formulaire. Le programme traite en premier les unitermes en intersection, puis il forme un ensemble avec les termes unis, ensuite viennent les expressions, enfin les termes exclus.

Si nous revenons dans le formulaire, Google aura réécrit la requête ainsi :

Find results	with all of the words	war Irak United-States-of-Ameri
	with the exact phrase	Great Britain
	with at least one of the words	United-Kingdom
	without the words	Canada

Le programme a fait passer **United-States-of-America** de la case d'expression exacte à la case d'intersection en considérant qu'il est devenu un mot. **Great Britain** a monté de la case d'union à celle d'expression exacte et **United-Kingdom** est demeuré dans la case d'union. En validant cette requête, le programme la re-formule.



Search: the web pages from Canada

Cette fois-ci, Google a retiré tous les opérateurs d'union pour clore la recherche avec environ 100 notices.

Nous constatons que le programme de Google est instable lorsque nous voulons modifier une requête. Il faut réécrire la requête au complet si l'on désire la modifier, par exemple en restreignant la recherche dans le temps ou à domaine.

Le comportement de Yahoo ! est stable comparativement à Google.

Voici la même requête dans le formulaire de Yahoo !

Show results with	all of these words	war Irak
	the exact phrase	United States of America
	any of these words	"Great Britain" "United Kingdom"
	none of these words	Canada

Yahoo ! traduit exactement ce qui est inscrit dans le formulaire. Remarquez que les expressions sont entre guillemets et l'opérateur est entre les deux expressions et non entre les mots des expressions.



Le nombre de notices est d'environ 500 000.

En revenant dans le formulaire, le programme de Yahoo ! n'a pas modifié la requête. Nous pouvons poursuivre la recherche en restreignant la requête dans le temps ou à un domaine par exemple **gov** pour trouver des documents du gouvernement des États-Unis.

Conclusion:

Le formulaire de Google doit être utilisé selon des paramètres précis.

La stratégie de recherche doit être composée avant de commencer la recherche si le format de fichiers, le domaine ou un laps de temps sont requis. Sinon, vous aurez à reformuler la requête.

Les mots ne doivent pas être composés à moins que ses parties ne soient unies par un trait d'union. S'il n'y a pas de traits d'union, le tiret est obligatoire au lieu des guillemets an-

glais.

La case d'expression est inutile. Le programme de Google dénature les expressions guillemetées. Mieux vaut placer l'expression dans la première case, les parties des mots jointes par des traits d'union.

Le programme de Google remplace les guillemets par des traits d'union dans la case d'exclusion. Autant utiliser les traits d'union dès la première inscription.

Utilisez les préférences pour décider du nombre de notices par page afin de ne pas avoir la surprise de reprendre la formulation de la requête.

Les requêtes de plus de trois mots com-

prenant des expressions et des champs doivent être formulées avec les préfixes de la recherche avancée.

Finalement, il faut se résoudre à utiliser une stratégie simple avec le formulaire. Heureusement pour Google, les internautes font majoritairement des requêtes avec moins de trois mots.

Marc Duval
Service de recherche documentaire DSI
1280-4, boul. de Montarville.
Boucherville. Québec. Canada
J4B 8B4
marcduval@dsi-info.ca