

## La rubrique du chercheur geek

### Parlez-vous le Google ? (Partie 1)

On parle beaucoup de la qualité des moteurs de recherche mais finalement assez peu des moyens d'améliorer les résultats d'une requête. Or, leur qualité dépend à la fois de la qualité de l'indexation (parcourir le contenu et l'indexer efficacement) sur laquelle vous n'avez pas de prise mais également de la qualité du traitement de la requête (interpréter la requête et renvoyer les résultats correspondants) qui, là, dépend de vous !

Cet article a pour objectif d'optimiser les requêtes envoyées à un moteur de recherche, en posant la meilleure question possible. Pour cela, nul besoin de s'y connaître en sémantique vectorielle, *text mining*, similarité cosinus et autres joies du *Big Data Processing*, puisque nous ne mobiliserons que des éléments basiques de l'algèbre de Boole (une approche algébrique de la logique) ainsi que quelques mots-clés utiles. Ces mots-clés sont ceux du moteur de recherche le plus utilisé et le plus connu, et celui que nous utiliserons : Google avec ses cinq milliards de requêtes et ses plus de 20 milliards de sites Web parcourus... par jour.

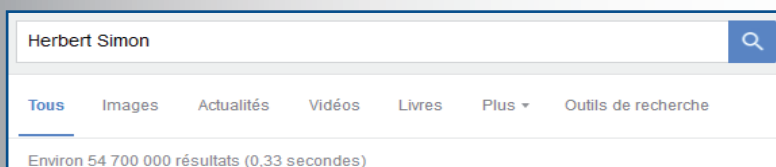
*Dis-moi qui, quoi ou pas, mais surtout dis-le-moi !*

En pratique, poser la meilleure question revient à laisser à Google le moins de liberté possible pour l'interprétation de ce que l'on recherche. Pour cela, il faut être spécifique concernant ce que l'on cherche (qui, quoi, combien), là où il faut rechercher (où, comment), et la période concernée (quand).

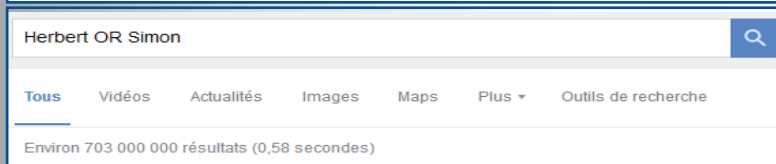
Lorsque vous saisissez une requête dans un moteur de recherche, celui-ci interprète chaque terme de la recherche et connecte chacun de ces termes à l'aide d'opérateurs logiques (nommés également opérateurs booléens). La liaison logique dans un moteur de recherche peut être de trois types :

- la conjonction (l'opérateur AND), qui valide les résultats uniquement si les termes connectés par AND sont tous présents ;
- la disjonction inclusive (l'opérateur OR) qui valide les résultats si au moins un des termes connectés par OR est présent ;
- la négation (l'opérateur NOT), qui valide les résultats si le terme connecté par NOT n'est pas présent.

Par défaut, Google ajoute l'opérateur logique AND entre chaque terme de la requête, il est donc inutile de spécifier cet opérateur dans une requête. En revanche, si vous souhaitez rechercher les pages renvoyant un ou plusieurs des termes mais pas forcément l'ensemble des termes, il faut utiliser l'opérateur OR entre les termes.

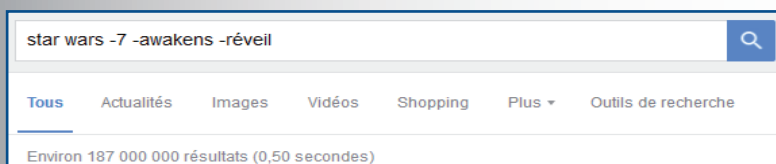


La requête [Herbert Simon]<sup>1</sup> renvoie les pages contenant Herbert ET Simon, soit approximativement 54 millions de pages.



La requête [Herbert OR Simon] renvoie quant à elle les pages contenant Herbert, Simon, ou Herbert Simon (ou Simon Herbert), soit approximativement 706 millions de pages.

Le dernier opérateur logique utilisable est l'opérateur de négation. Google ne reconnaît pas l'opérateur NOT, qu'il faut remplacer par le signe de la soustraction. En accolant le signe « - » à un terme, vous excluez celui-ci des résultats.



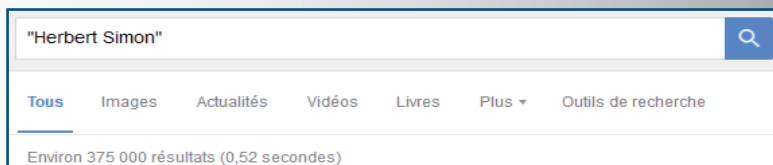
La requête [star wars -7 -awakens -réveil] renvoie les pages avec star ET wars mais ne contenant NI awakens, NI réveil, NI le chiffre 7 (et évite ainsi de connaître des éléments du dernier opus de la saga Star Wars).

<sup>1</sup> Les crochets représentent ce qui est saisi exactement dans le moteur de recherche ; ces crochets ne doivent pas être saisis dans la requête.

### Guillemets, métacaractère et imbrication

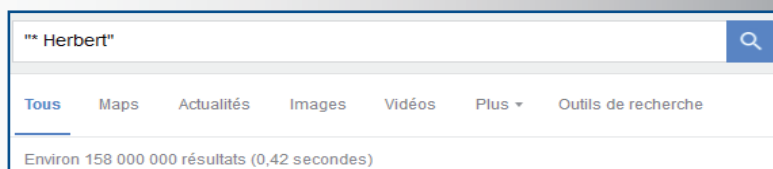
En plus des opérateurs logiques, il existe des éléments pour affiner les recherches. Le plus connu de ces éléments est l'utilisation de guillemets qui permettent de rechercher une expression exacte.

La recherche ["Herbert Simon"] renvoie les pages contenant spécifiquement (et dans cet ordre) « Herbert Simon », soit 375 000 résultats, représentant moins de 1 % du total des pages de la requête [Herbert Simon] sans les guillemets.

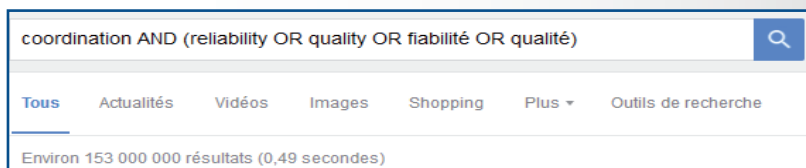


L'autre élément est l'utilisation du métacaractère de l'astérisque (appelé également *wildcard* ou caractère joker) pour compléter une expression dont vous savez qu'il y a un terme mais dont vous n'êtes plus certain.

Si vous cherchez l'auteur Frank Herbert mais que vous ne vous souvenez plus de son prénom, vous pouvez saisir ["\* Herbert"] indiquant ainsi que vous cherchez une expression avec un ou plusieurs termes avant Herbert, spécifiquement et dans cet ordre. Ceci vous permet de filtrer les résultats ne contenant que Herbert, ou Herbert précédé de quelque chose.



Enfin, il est possible d'utiliser la distributivité des opérations (le *nesting* en terme indigène) dans les recherches pour imbriquer différents opérateurs. En effet, les termes d'une recherche peuvent être mis entre parenthèses afin qu'ils soient traités comme un ensemble. Prenons l'exemple suivant : vous souhaitez effectuer une recherche renvoyant les résultats relatifs à la coordination, la fiabilité et la qualité, à la fois en anglais et en français. Une chance pour vous, le terme coordination peut être utilisé pour l'anglais et le français, mais pour les deux autres termes, il faudra effectuer une requête avec les termes anglais et avec les termes français. Vous vous retrouvez donc avec quatre requêtes : [coordination AND reliability], [coordination AND quality], [coordination AND fiabilité] et [coordination AND qualité] (l'opérateur AND étant accessoire dans ces quatre requêtes). En utilisant la propriété de distributivité, vous pouvez traiter cela en une seule requête, dans laquelle le terme « coordination » est « distribué » aux autres éléments de la requête.



Au terme de cette première partie nous avons fait un tour d'horizon des fonctions élémentaires pour effectuer une recherche dans un moteur de recherche, à savoir définir précisément le ou les termes cherchés. Dans le prochain numéro, nous aborderons

les fonctions avancées des moteurs de recherches qui permettent d'intégrer dans une recherche là où il faut rechercher (où, comment), la période concernée (quand), ainsi que certaines fonctions particulières.

\*\*\* pour aller plus loin \*\*\*

Explication plus détaillée de la phase d'indexation des moteurs de recherche :  
How does Google Search work? – <https://www.youtube.com/watch?v=KyCYyoGusqs>

**Geoffrey Leuridan**  
IAE Lille, LEM UMR CNRS 9221

**Cécile Chamaret**  
Université Paris Sorbonne Abu Dhabi