



Focus 2 - Moteurs de recherche et bulles de filtres

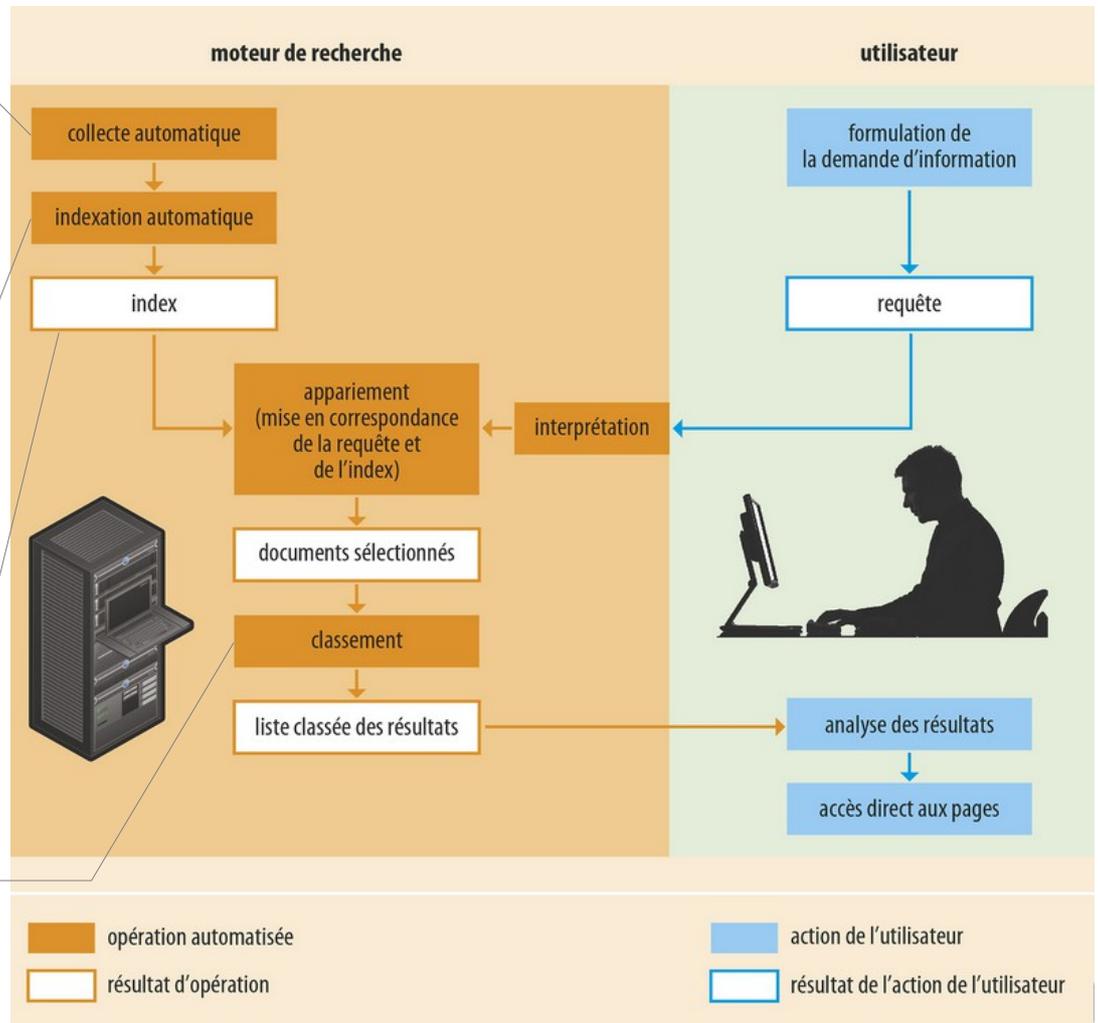
1- Principe de base d'un moteur de recherche

La **collecte** est réalisée par un robot (spider/crawler/bot) qui parcourt inlassablement les pages du web et collecte les adresses url rencontrées

L'**indexation** consiste, pour chaque document du corpus, à récupérer le code html et à l'analyser pour construire une représentation dudit document. Le robot extrait 3 types d'infos :
 - le contenu textuel (mots clés)
 - les métadonnées
 - les liens url vers l'extérieur

L'**index** est la base de données où sont enregistrés les mots clés extraits des docs

Chaque moteur a sa propre recette qui combine un grand nombre de **critères** exprimés dans un **algorithme** tenu secret. Les plus performants sont basés sur un **modèle vectoriel** qui permet de mettre en relation le poids des différents mots de la requête avec le poids de ceux contenus dans les documents. Google utilise la technique du PageRank qui pondère encore le classement par un indice de popularité des pages.



Source : Encyclopédie Universalis

2- Comment les algorithmes nous enferment

2.1- Quand l'économie s'en mêle

Les moteurs de recherche

Google, Bing ou Yahoo sont des entreprises commerciales qui vivent de la publicité tout en fournissant un service de recherche gratuit aux internautes.

→ les moteurs ne traitent pas tous les sites sur un pied d'égalité, mais les classent en fonction de ce qu'ils leur rapportent

Les entreprises du web

Les moteurs de recherche sont un facteur de visibilité pour les entreprises qui cherchent à capter l'attention des internautes

→ il leur faut apparaître en bonne position dans les classements opérés par les moteurs

STRATÉGIES DE WEB-RÉFÉRENCIEMENT DES ENTREPRISES POUR AMÉLIORER LEUR « CLASSEMENT »

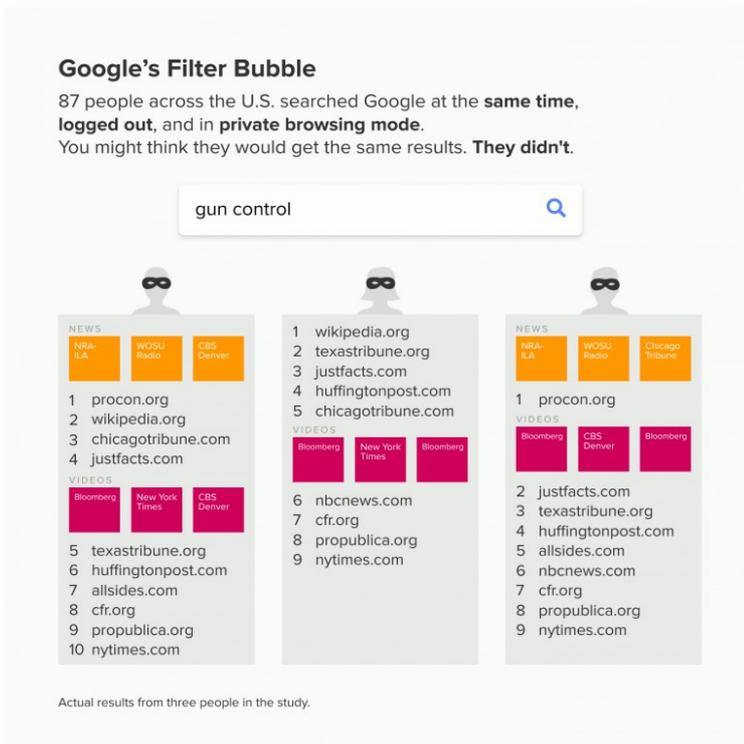
Le référencement dit « payant »

implique de recourir à la publicité sur les moteurs en achetant des mots clés ou des expressions, selon un système d'enchères. Lorsque l'internaute interroge un moteur sur les mots choisis, les annonces publicitaires des offres les plus fortes apparaissent en marge des résultats éditoriaux.

Le référencement dit « naturel »

conduit à adapter les pages d'un site pour que celui-ci apparaisse dans les résultats éditoriaux. Cela peut consister à mettre en avant certains mots clés et à faire en sorte que de nombreux liens pointent vers les pages.

2.2- Quand nos pratiques de navigation s'en mêlent



Source : DuckDuckGo

Mesurer la bulle : Comment Google influence-t-il les clics ?
Étude menée en décembre 2018 auprès de 87 internautes ayant saisi simultanément les mêmes termes de recherche

« En pratique, [la manipulation des résultats de recherche] signifie que les liens sont déplacés vers le haut, le bas ou ajoutés à vos résultats de recherche Google, non pas en fonction de leur pertinence, mais selon le profil de l'utilisateur. Ces résultats sont basés sur les informations personnelles que Google a récoltées sur vous, comme l'historique de vos recherches, votre navigation sur le Web et votre historique d'achat, et vous placent dans une bulle de liens sur lesquels Google suppose que vous êtes le plus susceptible de cliquer » indique l'étude.

Le concept de **Bulles de filtres** a été théorisé par l'activiste américain Eli Pariser en 2011.

« À partir des différentes données collectées sur l'internaute, des algorithmes vont silencieusement sélectionner les contenus qui seront visibles ou non par lui. Le terme de « bulle de filtres » renvoie à l'isolement produit par ce mécanisme : chaque internaute accède à une version différente du web, il reste dans une « bulle » unique et optimisée pour lui. »



Déduisez des énoncés suivants les avantages et les inconvénients de la personnalisation des contenus opérée par les moteurs de recherche

En 2011, un grand nombre d'internautes a soumis pendant plusieurs semaines des requêtes sur Google associant le nom d'une société « la lyonnaise de garantie » et le terme « escroc ». La fonctionnalité *Suggest* de Google en a pris acte, et proposait à toute personne saisissant le nom de l'entreprise d'y accoler ce terme indubitablement injurieux. Dans cette affaire judiciaire qui a fait grand bruit, le moteur de recherche a d'abord été condamné par une cour d'appel, mais la décision a été par la suite censurée par la Cour de cassation au motif que le rapprochement critiqué était le fruit d'un processus purement automatique et ne pouvait être imputé à la volonté de nuire de l'exploitant du moteur de recherche. Emmanuel Netter, *Numérique et droit privé*, 2017

« Si *** se contente de saisir dans le moteur de recherche de son entreprise une suite de caractères en apparence absconse, comme « C-362/14 », l'algorithme comprendra qu'il s'agit d'un numéro d'arrêt rendu par la Cour de justice de l'Union européenne, car cet utilisateur déterminé a l'habitude de consulter le site de la juridiction. La même requête formulée par une personne ayant d'autres centres d'intérêt conduira à des résultats différents. » Francis Donnat, directeur des politiques publiques de Google France, lors d'une conférence donnée le 7 décembre 2015

« Si une personne politiquement marquée à droite et une autre plutôt à gauche lancent une recherche Google sur les lettres « BP », la première reçoit, en tête de page, des informations sur les possibilités d'investir dans la British Petroleum, la seconde sur la dernière marée noire qu'a causée la compagnie pétrolière britannique. » Eli Pariser, *The Filter Bubble*, 2012

« Latanya Sweeney, une chercheuse en informatique afro-américaine, a remarqué que, lorsqu'elle tapait son nom dans le moteur de recherche de Google, elle voyait apparaître la publicité « *Latanya Sweeney arrested ?* ». Cette publicité propose un service de consultation en ligne [...] qui permet [...] de savoir si les personnes ont un casier judiciaire. Or, le nom de ses collègues blancs n'était pas associé au même type de publicité [...]. » Dominique Cardon, *A quoi rêvent les algorithmes*, 2015

les avantages

les inconvénients