
LES MOTEURS DE RECHERCHE

SIRE, G. (2016). Les moteurs de recherche. Paris: La Découverte.

Kimberley André, Clara Baldi, Camille Boudoire, Emna Drira



L'AUTEUR

Guillaume Sire est écrivain et maître de conférences en Sciences de l'Information et de la Communication à l'Université de Toulouse et à l'Institut Français de Presse. Ses sujets de prédilection sont les moteurs de recherche, le journalisme en ligne, les industries culturelles et le code informatique.

Il a publié de nombreux articles scientifiques et plusieurs ouvrages à propos des **mutations sociales liées au numérique** et a notamment consacré une thèse à l'analyse de la **relation entre Google et les entreprises** de presse.

PRESENTATION GENERALE DE L'OUVRAGE

Le moteur de recherche est un **outil computationnel** chargé de prendre connaissance des documents présents sur le web et de les classer pour l'utilisateur. Une fois les informations identifiées, le moteur de recherche **évalue** et **hiérarchise** les résultats par ordre de pertinence grâce à un **algorithme**. Sa fonction est de mettre en avant certains documents, en les rendant accessibles et donc visibles, tandis que d'autres resteront en arrière, accessibles mais invisibles...

La question du rôle joué par les moteurs de recherche est d'autant plus prégnante que le moteur de recherche Google est en **position hégémonique** dans la plupart des pays du monde et représente notre premier réflexe lors d'une recherche d'information. La place prépondérante de cet acteur inquiète.

Cet ouvrage a la particularité de pouvoir toucher **plusieurs publics** : les webmasters, qui y trouveront des renseignements sur les systèmes d'indexation ainsi que sur le

référencement, ainsi que les utilisateurs comme toutes les personnes qui s'intéressent aux enjeux globaux des moteurs de recherche à l'échelle de la société.

Guillaume Sire opte pour un ton objectif : il expose et analyse les faits sans pour autant condamner les moteurs de recherche. Il démontre qu'il est primordial pour les usagers des moteurs de recherche de **comprendre leur fonctionnement et enjeux**. Dans son ouvrage l'auteur s'intéresse aux rouages du référencement, à la **récolte massive de données** et à **l'influence des moteurs de recherche** sur l'information, la connaissance et la société.

REFLEXION PROPOSEE

Dans son livre Guillaume Sire nous aiguille vers une autre dimension des moteurs de recherche en soulevant par exemple la notion de **droit** et de **devoir** vis-à-vis du contenu et le **pouvoir normatif** de cet outil. Il évoque aussi les questionnements que soulèvent les moteurs de recherche : la question du **droit d'auteur** ou bien de la **protection de la vie privée**.

On se méfie davantage de la nature d'une information et de sa véracité que de son cheminement. Pour l'auteur, les **résultats proposés** par les moteurs de recherche et leur référencement ne sont **ni neutres, ni naturels**. Il faut alors se demander pourquoi les moteurs de recherche nous mènent à cette information plutôt qu'à une autre.

Ce livre nous dévoile les **stratégies des moteurs de recherche** et les **conséquences** que peuvent avoir les technologies d'indexation sur l'information. Néanmoins, Guillaume Sire n'a pas la volonté de dénoncer Google ou bien de nous dissuader d'utiliser les moteurs de recherche. Ce livre nous amène à réfléchir à ces outils et au fonctionnement du web comme "**faiseurs d'information**" et aux **conséquences** que le geste anodin d'une recherche sur le net peut avoir sur la **société**.

LES MOTEURS DE RECHERCHE COMME TECHNOLOGIE POLITIQUE

Les moteurs de recherche traitent les requêtes et diffusent des pages de résultats en fonction de leur pertinence par rapport à l'utilisateur. Ces outils de recherche puissants donnent accès à des milliers de sources d'information en ligne, et structurent l'information à laquelle nous avons accès. Ils ont de ce fait la possibilité d'orienter nos opinions, ce qui fait d'eux **plus qu'un outil technique, une technologie politique**.

DES ALGORITHMES NECESSAIREMENT OPAQUES

Dès le lancement d'une requête sur Google un algorithme **classe les résultats en fonction de critères de pertinence**. Mais ce dernier est réputé pour son **opacité**, et **évolue** en permanence de manière à ce que personne ne puisse le connaître.

« Si cet algorithme était rendu public tous les éditeurs de contenu agiraient de manière à optimiser le calcul de pertinence de leurs documents. Cela empêcherait l'algorithme de fonctionner, incapable de choisir entre les documents conçus de la même manière. »

Certains créateurs de contenu agissent en fonction de ce qu'ils pensent savoir de l'algorithme. Ils se plient donc aux **règles**, réelles ou non, utilisées par les moteurs de recherche pour **évaluer** et **hiérarchiser** leurs documents. Même si la firme publie des **recommandations** sur son "centre d'aide aux webmasters" pour **guider** les éditeurs de contenu, elle ne fait aucune promesse de résultat.

MISE EN AVANT DE L'INFORMATION

Mais comment **pouvons-nous être certains de la pertinence des résultats si nous ne savons pas comment ils sont triés et choisis ?** Nous ne sommes pas en capacité de savoir si Google met en avant les contenus de leurs entreprises partenaires. C'est pourquoi Jean-Noël Jeanneney, président de la Bibliothèque nationale de France, proposait **un algorithme européen unique et transparent**. Cela répondrait à l'accusation des moteurs de recherche qui privilégieraient les sources d'information les plus populaires et leurs partenaires économiques.

CLASSEMENT DE L'INFORMATION

A ce sujet, la chercheuse Susan Gerhart a démontré que les moteurs de recherche ont tendance à atténuer les débats. Pour cela, elle a effectué une série de requêtes à propos de sujets controversés et a constaté que **les premiers résultats avaient tendance à privilégier les informations consensuelles**. Preuve que les critères de pertinence de l'algorithme impactent notre manière de **recevoir** et **comprendre** l'information. Après cette prise de conscience, nous devons questionner sur notre **dépendance** aux moteurs de recherche et leurs **choix de classification**.

TECHNIQUES DE REFERENCEMENT

Dans le but d'être le mieux référencés possible, les éditeurs de contenu utilisent les techniques de **SEO** (Search Engine Optimization) soit d'**optimisation** pour les moteurs de recherche. Ces techniques visent à améliorer la compréhension de la thématique et du contenu des pages d'un site et à **augmenter le trafic naturel** de celui-ci. L'objectif étant d'**améliorer le positionnement** d'une page web dans les pages de résultats de recherche **via des mots-clés ciblés**. Cela rend compte de l'influence que les concepteurs de l'algorithme sur les éditeurs de contenu. Ils sont malgré eux **tributaires** du fonctionnement des moteurs de recherche et répondent aux "exigences" des moteurs. Les éditeurs ont conscience que **la visibilité de leurs documents dépend de leurs actions**, et agissent donc, dans le but d'être mieux référencés. Dès lors les décisions des

concepteurs d'un moteur comme Google sont susceptibles **d'influencer la nature du web tout entier, voire d'impacter la nature des informations auxquelles les usagers ont accès.**

LE POUVOIR D'INFLUENCE DES MOTEURS DE RECHERCHE

POUVOIR NORMATIF

Guillaume Sire nous explique que le **pouvoir normatif** est celui d'édicter des **normes**, des **règles générales** ou des **décisions** qui créent des **droits** ou des **obligations** pour ceux à qui elles sont destinées. Les moteurs de recherche sont dotés de ce pouvoir normatif. Les éditeurs de sites web adhèrent au principe : " *Exister c'est être indexé par un moteur* " Introna & Nissenbaum, 2000. Ce qui pousse ces éditeurs à se plier aux règles et décisions des concepteurs des moteurs de recherche afin d'assurer leur visibilité et indexation sur ces derniers.

THEORIE DES BULLES DE FILTRE

Guillaume Sire nous présente également une théorie développée par Eli Pariser en 2011, la **théorie des bulles de filtre**. Les algorithmes utilisés sur les réseaux sociaux auraient tendance à **renforcer les opinions et croyances des utilisateurs** sans que ceux-ci ne s'en rendent compte, en leur recommandant des contenus en adéquation avec leurs préférences. Grâce aux nouvelles technologies de recherche et de récolte massive de datas, les moteurs sont en mesure de répondre à la tendance générale de la personnalisation et d'offrir une navigation internet "sur-mesure" à leurs utilisateurs.

ENFERMEMENT DE L'INTERNAUTE

Bien que certains considèrent la **personnalisation** comme une façon d'améliorer la pertinence des informations pour les utilisateurs, cette technologie ne convainc pas tout le monde. Le principal enjeu étant "**l'enfermement de l'internaute**". L'utilisation des bulles de filtres conforterait l'internaute dans une **connaissance préexistante** au lieu de l'ouvrir à de nouveaux horizons, c'est ce que l'on appelle la "**cyber-balkanisation**". Dans le cas d'une forte personnalisation on ne qualifierait plus les moteurs de recherches comme des outils d'information mais de confirmation. L'essayiste Eli Pariser va même jusqu'à parler d'une forme de censure.

LA PERSONNALISATION N'ENFERME PAS L'INTERNAUTE

Cependant le sociologue français Dominique Cardon réfute l'idée selon laquelle la personnalisation réduirait l'utilisateur à **ses propres centres d'intérêts**. L'**opacité** et le **flou** qui entourent les modalités de fonctionnement des principaux moteurs de recherche ne contribuent en rien à clore le débat au sujet de la personnalisation.

MIEUX COMPRENDRE LA STRATEGIE DE GOOGLE POUR MIEUX L'UTILISER

LA STRATEGIE DE DIVERSIFICATION DE GOOGLE

Aujourd'hui, personne n'ignore que **Google détient le titre de leader**. En effet, le moteur détient plus de **90% des parts de marché** dans de nombreux pays tels que la France, l'Espagne, l'Allemagne et les pays Anglophones. Cette position hégémonique donne à la firme ce que l'on appelle un pouvoir normatif. Pour asseoir cette position archi-dominante la firme a développé une **stratégie de diversification** en proposant des services extrêmement variés (Gmail, Picasa, Google Translate, Google Document, ou encore le navigateur Google Chrome). Ces activités sont toutes susceptibles de servir à l'activité du moteur de recherche ou à l'activité publicitaire. Ils permettent au géant du web de **récolter des données de façon massive** sur les utilisateurs et de personnaliser au mieux les résultats proposés, mais également les publicités contextuelles. D'autre part, cette diversification a pour but de garder un trafic substantiel puisque quand l'internaute cherche une information il est renvoyé en priorité vers les outils Google.

COLLECTE DE DONNEES

Depuis 1998 Google stocke les données de ses utilisateurs, ainsi la firme connaît toutes les requêtes effectuées, par qui, quand, où et sur quels liens l'utilisateur a cliqué. Cette base de données **agit comme une barrière à l'entrée du marché** dans la mesure où si un nouvel acteur voulait entrer sur le marché, même avec un très bon algorithme, il ne pourrait en aucun cas proposer des résultats aussi bien personnalisé que ceux de la firme.

GOOGLE ET LA PUBLICITE

La technologie assurant les revenus de ces **activités publicitaires** et qui participe à la fortune de Google est l'idée des **liens sponsorisés** grâce à l'**achat de mots clés**. Habituellement les annonceurs achètent un espace publicitaire dans lequel leur annonce apparaît, juxtaposée à d'autres publicités. Cette fois l'internaute a **formulé une requête** qui lui fournit les résultats "**naturels**" ainsi que les sites d'**annonceurs ayant acheté les mots clés** utilisés. C'est ce que l'on appelle aussi les "**AdWords**", qui sont toujours en haut de page ce qui permet de bien les différencier des résultats naturels afin d'éviter toute confusion. En 2007 Google a fait l'acquisition de la régie publicitaire DoubleClick et s'est ouvert aux agences de publicités qui peuvent désormais diffuser leurs annonces sur les réseaux partenaires de la firme.

ENJEUX DES MOTEURS DE RECHERCHE

ENJEUX DE SOCIÉTÉ : A QUI PROFITENT CES TECHNIQUES DE PERSONNALISATION ?

Les moteurs de recherche utilisent des outils qui leur permettent de récolter des données sur leurs utilisateurs. Parallèlement, une autre question subsiste : **à qui profitent ces techniques de personnalisation ?**

Si certains auteurs comme Goldman pensent que la personnalisation peut profiter aux internautes, d'autres affirment qu'elle profiterait aux annonceurs. En effet, le moteur offre aux annonceurs de cibler les utilisateurs en fonction de leur historique de navigation. Selon Granka, ce biais risquerait d'amplifier la formation des bulles de filtre et contribuerait à enfermer l'internaute dans ses croyances.

On peut donc dire que la personnalisation n'est pas néfaste, mais qu'il est **important de continuer à interroger les systèmes et algorithmes**, que ce soit dans le domaine politique, culturel ou économique.

"Nous devrions réfléchir à des solutions qui permettraient de minimiser les effets négatifs [des technologies de personnalisation] et de maximiser les effets positifs plutôt que d'essayer de nous en débarrasser totalement".

ENJEUX DE DROIT

Guillaume Sire s'attache également aux controverses soulevées par les moteurs de recherche dans le domaine du droit, et notamment du droit d'auteur et de la responsabilité éditoriale.

DROITS D'AUTEUR ETATS-UNIEN

Etant donné le caractère international des moteurs de recherche, lors de conflits portés en justice, le juge peut se référer soit au droit états-unien, soit au droit du pays de l'éditeur ou encore au droit du pays de l'utilisateur qui a lancé la requête. Bien évidemment plusieurs conditions entrent en jeu pour trancher sur le droit à appliquer.

Dès lors, le procès jugé valable et le droit de référence établi, **le juge se trouve face aux divergences des textes de loi relatifs aux droits d'auteur de différents pays**. Si on cite les Etats Unis comme exemple, siège des mastodontes des moteurs de recherche, il existe dans le droit de ce pays le '**Fair Use**' : *«Une exception au copyright permettant à un tiers, si certaines conditions sont remplies, d'utiliser un contenu ou une partie d'un contenu qui ne lui appartient pas.»* Dans ce cas, le juge analysera certaines conditions qui doivent être remplies pour bénéficier de cette exception.

L'auteur note qu'il existe un texte de loi aux Etats-Unis qui stipule que les propriétaires des moteurs de recherche doivent rémunérer les éditeurs dont le contenu est repris dans

ses résultats, c'est le '**Hot News misappropriation**'. Néanmoins les entreprises des moteurs de recherche bénéficient d'une jurisprudence favorable quant à la sécurité juridique.

Ce n'est donc pas par hasard qu'on trouve les plus grandes entreprises des moteurs de recherche (Google, Yahoo, Msn) basées aux Etats Unis. **Ils bénéficient ainsi librement du 'Fair Use'**, ce qui leur a permis une croissance explosive.

DROITS D'AUTEUR EN EUROPE ET EN FRANCE

L'auteur décrit les controverses dans la législation américaine mais ne manque pas d'informer sur l'Union Européenne. L'UE introduit la directive '**Triple Test**'. Celle-ci, dont l'interprétation est largement débattue par les juristes, stipule que des exceptions pourront être faites aux droits d'auteur. Toutefois, « la mise en mémoire cache et l'absence de toute compensation versée par le propriétaire d'un moteur aux éditeurs posent particulièrement problème dans le cas du triple test ».

D'autre part, en France, les articles de reproduction de la propriété intellectuelle dénoncent l'activité des moteurs de recherche. Le propriétaire d'un moteur de recherche pourrait ainsi être accusé de reproduction illicite. C'est une tendance notable **en Europe** puisque, contrairement aux Etats-Unis, **les juges ont tendance à pencher du côté des éditeurs dont le contenu serait dérobé par les moteurs.**

RESPONSABILITE EDITORIALE

La responsabilité éditoriale est également sujet de controverses dans le cadre des moteurs de recherche. La responsabilité éditoriale se manifeste lorsque le moteur de recherche remonte dans les résultats des pages dont le contenu ne respecte pas les droits d'auteur ou affiche des informations calomnieuses ou dégradantes. Intervient alors le '**Safe Harbor**' qui consiste à traiter et retirer les résultats délicats des moteurs.

ALTERNATIVES ET PERSPECTIVES

Guillaume Sire propose également dans son ouvrage une ouverture sur le futur de la recherche d'information.

Dans un premier temps à travers la **prise en compte des interactions sociales** comme dans l'outil de recherche Facebook GraphSearch grâce à des données récoltées dans les réseaux sociaux, messageries et peut-être plus tard dans les échanges téléphoniques. Et dans un second temps à travers la **prise en compte de l'espace-temps**, c'est à dire que les résultats ne seront pas les même si on se trouve à la maison ou au bureau, idem le matin le midi ou le soir.

Par ailleurs l'auteur axe sa réflexion sur la **distribution de contenus** à savoir proposer aux utilisateurs directement du contenu dans ses résultats de recherche plutôt que de

l'envoyer vers un site tiers. Cette stratégie permettrait aux moteurs de recherche de garder dans leurs giron les internautes.

Guillaume Sire se penche également sur la **recherche sémantique**, basée sur la compréhension de l'objectif de recherche afin de générer des résultats plus pertinents. C'est ce que les ingénieurs de Google ont nommé le KnowledgeGraph et qui propose dans le cadre d'une recherche des informations annexes et complémentaires afin de cibler au mieux le besoin de ses utilisateurs. Avec le web sémantique **il ne sera plus question pour les moteurs de recherches d'être seulement des relais d'information mais de devenir des sources à part entière.**

Bien que le moteur de recherche Google occupe une grande place dans l'ouvrage de Guillaume Sire, l'auteur n'oublie pas de mettre en lumière d'autres **moteurs de recherches dits "alternatifs"**. Ces moteurs ne sont pas les mastodontes en matière de recherche d'information et ne le seront probablement jamais. Cependant, ils fournissent à l'utilisateur un accès à l'information **avec pour chacun une spécificité**. Certains moteurs sont basés sur l'algorithme de Google comme le moteur solidaire Lilo ou le moteur situé géographiquement Egerin pendant que d'autres prennent le contrepied du géant comme Qwant le moteur respectueux de la vie privée.