Guillaume SIRE, *Les Moteurs de recherche*

Paris, Éd. La Découverte, coll. Repères, 2016, 128 pages

Florence Michet

Édition électronique

URL : <https://journals.openedition.org/questionsdecommunication/16998>

DOI : 10.4000/questionsdecommunication.16998

ISSN : 2259-8901

Éditeur

Presses universitaires de Lorraine

Édition imprimée

Date de publication : 31 décembre 2018

Pagination : 417-419

ISBN : 978-2-8143-0543-4

ISSN : 1633-5961

Référence électronique

Florence Michet, « Guillaume SIRE, *Les Moteurs de recherche* », *Questions de communication* [En ligne], 34 | 2018, mis en ligne le 31 décembre 2018, consulté le 08 janvier 2022. URL : <http://journals.openedition.org/questionsdecommunication/16998> ; DOI : <https://doi.org/10.4000/questionsdecommunication.16998>

Ce document a été généré automatiquement le 8 janvier 2022.

Questions de communication is licensed under CC BY-NC-ND 4.0

Guillaume SIRE, *Les Moteurs de recherche*

Paris, Éd. La Découverte, coll. Repères, 2016, 128 pages

Florence Michet

RÉFÉRENCE

Guillaume SIRE, *Les Moteurs de recherche*, Paris, Éd. La Découverte, coll. Repères, 2016, 128 pages

- 1 Devant le flot informationnel apporté par l'internet, les moteurs de recherche essaient, dans leur diversité, de guider l'utilisateur à la manière d'un outil de navigation « chargé de prendre connaissance des documents accessibles sur le web » (p. 4). Il mettra en avant les documents les plus utilisés considérés comme les plus pertinents reléguant les autres au second plan, les rendant presque invisibles. À l'époque de la rédaction maître de conférences à l'Institut français de presse (université Paris 2 Panthéon-Assas), Guillaume Sire propose d'aider à comprendre le fonctionnement de ces systèmes, Google au premier plan. L'ouvrage est découpé en sept parties développant un panorama complet de ces outils et de leur historique.
- 2 Le premier chapitre, « Généalogie » (pp. 9-18), retrace à la fois l'évolution des supports d'échange de messages et des moyens de stockage et celle de la mise à disposition pour faciliter l'accès : « C'est parce qu'on souhaite mobiliser l'information dans le temps et l'espace, et parce qu'on organise socialement et techniquement les modalités de sa communication, que la recherche est possible » (p. 8). À la fin du XIX^e siècle, Paul Otlet imaginait déjà la recherche informatisée et la mécanisation laissa peu à peu place à l'informatisation. À chaque nouvelle méthode de traitement ou de transmission de l'information, « il fallut concevoir de nouveaux procédés de stockage et de nouvelles procédures d'accès » (p. 10). Dans les années 1950, les chercheurs proposèrent d'abord d'attribuer des mots clés aux documents, puis des scores de pertinence, méthodes qui s'imposèrent et changèrent les procédés d'indexation et de recherche (p. 11). Ces

procédés ne cessèrent d'évoluer, mais l'application au web reposait sur des problèmes de norme de publication, d'indexation et du nombre de documents mis en ligne. Le moteur de recherche Altavista tira son épingle du jeu et fut « le premier moteur à permettre au grand public d'effectuer des recherches en langage naturel » (p. 14). L'annuaire Yahoo constitua une première révolution en proposant des liens vers des sites par catégories et sous-catégories accompagnés d'une description, mais, fin 1990, le nombre de sites était devenu trop important pour cette méthode. Puis vint Google, et son principe de *Pagerank* (l'algorithme d'analyse des liens participant au système de classement des pages web utilisé par le moteur de recherche Google). Ce premier chapitre éclaire sur ce qui a motivé les concepteurs des moteurs de recherche actuels pour la mise à disposition de l'information la plus pertinente et son accès.

- 3 Le deuxième chapitre, « Description sociotechnique » (pp. 19-35), explique en détail les tâches caractérisant tous les moteurs de recherche et modes opératoires, révélant ainsi ce qui se cache derrière une interface souvent simpliste. Pour fonctionner, un moteur de recherche suit quatre étapes essentielles entre les informations disponibles et les résultats d'une requête : d'abord le travail des *crawlers* ou collecteurs repérant les informations possibles ; ensuite, l'indexation pour permettre une recherche pertinente et rapide ; puis, la formulation de la requête par l'internaute, le moteur devant faire le lien avec la base de données ; enfin, la présentation des résultats et le classement des informations. Guillaume Sire prend le temps d'expliquer la genèse, la domination de Google et la mise au point de son efficacité qui en fait l'un des moteurs les plus complets (p. 30). Il est à noter que ce chapitre est difficile à comprendre pour le novice parce qu'il utilise des termes très spécifiques.
- 4 Dans le chapitre intitulé « Modèles et stratégies économiques » (pp. 37-54), l'auteur reprend les différentes tentatives de monnayer les contenus et services proposés sur le web, concurrencées par les moteurs gratuits présentant des bannières publicitaires ou une concentration des services sur une même page afin d'augmenter le nombre de visites et, par conséquent, le revenu. Yahoo et Google, chacun à sa manière, occupèrent le marché, l'un sous la forme d'un annuaire avec une interface développée, l'autre sous celle d'un moteur de recherche automatisé avec une interface simple. Les deux ne cessèrent de se diversifier, mais dans des voies toujours différentes, l'un avec un positionnement moteur et l'autre avec un positionnement média (p. 39). Puis Bing (Microsoft) fit son apparition, doté lui aussi d'une interface très épurée, « une stratégie d'imitation » (p. 40). Même si Yahoo accepta un partenariat avec Bing, le proposant comme moteur de référence, cette alliance ne fit pas vaciller Google. La diversification de ce dernier avec d'autres services (Gmail, Google translation, Google Chrome, Android, Youtube...) accentua sa domination et lui permit de collecter une quantité croissante de données des individus s'identifiant et augmentant ainsi la pertinence des résultats du moteur (p. 41). Et, bien que ses fondateurs Sergey Brun et Larry Page ne souhaitent pas, au début, nouer des partenariats avec des annonceurs publicitaires, leur entreprise est devenue « la plus importante régie publicitaire en ligne » (p. 43). L'auteur explique aussi comment la publicité à peu à peu envahi l'internet, Google allant jusqu'à facturer à l'annonceur le simple clic de l'internaute visitant les pages. Sa position dominante est liée au « coût d'opportunité » (p. 50), c'est-à-dire à l'habitude de se servir de Google et à la difficulté engendrée par l'adaptation à un moteur de recherche différent. Le problème du recueil des données des utilisateurs depuis 1998, appelées *query logs*, utilisées pour mesurer la pertinence des contenus et la personnalisation des résultats sont évoqués. Selon Guillaume Sire, le succès de Google

serait dû à l'étendue de sa base de données plus qu'à son algorithme, même si ses concurrents l'accusent de manipulation des résultats de recherche.

- 5 Ensuite, le chapitre « Droits et devoirs vis-à-vis du contenu » (pp. 55-71) reprend les différentes dispositions prises face à la loi d'abord pour les États-Unis, puis pour l'Union Européenne dont la France. La législation est complexe et difficile à cerner. L'auteur explique que le droit états-unien – *Copyright Act* – reste très faible. Pour utiliser certaines données, un tiers doit se prévaloir du *fair use*, sans quoi il s'expose à des poursuites ; ce qui permet à l'auteur de conclure que le droit américain accorde aux moteurs une « relative sécurité juridique » (p. 57). Le droit américain apparaît ici assez nébuleux : « Aucune solution ne permet, sans risquer de conduire à une forme de censure, de satisfaire un individu ayant émis le désir de faire retirer un contenu de l'index d'un moteur de recherche sous prétexte que ce contenu est calomnieux » (p. 61). Guillaume Sire s'intéresse ensuite au droit européen, dont l'adoption d'une directive en 2001 (*EU Copyright Directive*) concernant le droit d'auteur et proposant un « triple test », qui reste assez floue. Un encadré traite du droit français, où la directive n'a été transposée qu'en 2006, qui fait que la justice doit trancher au cas par cas (p. 62).
- 6 Puis l'ouvrage expose la question du « pouvoir normatif » (pp. 73-84). Le chapitre qui y est consacré s'intéresse aux éditeurs de sites et aux stratégies qu'ils mettent en œuvre pour augmenter les chances d'être visités par l'apparition dans les résultats de recherche et un positionnement haut dans la page (p. 73). Guillaume Sire reprend ici l'influence indéniable des moteurs de recherche et de leurs concepteurs sur le contenu de certains sites exerçant encore plus leur domination. L'algorithme de Google reste un mystère et heureusement, nous dit-il, sinon il serait incapable de faire des choix de pertinence. Bien que le moteur ne divulgue pas les détails de son algorithme, il propose des recommandations ou normes de publication ainsi que certains outils pour optimiser les chances – parmi lesquels « tendances de recherche » (p. 75) pour indiquer les sujets les plus recherchés conduisant les éditeurs à faire des choix rédactionnels particuliers ou à utiliser des mots clés spécifiques. La vitesse de réponse d'un site doit aussi être élevée pour être mieux référencé par Google : « performance du contenant pour pertinence du contenu » (p. 77). Dans ce chapitre, il est aussi question du code source (p. 78), des formats (*ibid.*), de l'arborescence d'un site web (p. 79), des ancres ou documents où se trouvent des liens (p. 80) et des interactions sociales ou commentaires des internautes comme autant d'indicateurs de pertinence d'un site. Est expliqué ce que sont les chapeaux blancs, les éditeurs de contenus qui se conforment aux prescriptions du moteur, et les chapeaux noirs, ceux qui les contournent : « L'algorithme est le résultat d'une influence mutuelle entre les actions des concepteurs du moteur de recherche et les actions des éditeurs de contenus » (p. 84).
- 7 Le chapitre « Un traitement de l'information controversé » (pp. 85-95) interroge les controverses dont le moteur de Google fait l'objet : opacité des algorithmes, des classements des résultats, des critères de pondération compliquant les tâches des éditeurs. Ce dernier aspect est néanmoins essentiel et ferme la porte aux éditeurs qui tentent de contourner le moteur pour proposer des documents moins pertinents. À ce sujet, Guillaume Sire pose la question de savoir si le moteur ne privilégie pas ses propres entreprise. Pour certains, souvent critiquée, la personnalisation possible des résultats enfermerait les connaissances des internautes en les maintenant dans une « bulle-filtre » informationnelle (p. 87). Pour d'autres, au contraire, elle permettrait de ne pas s'éparpiller par une moindre exposition aux informations. Un autre aspect

évoqué est celui des privilèges accordés ou non à certaines sources d'information populaires ou à de grands portails qui relèguerait les autres au second plan. Il est aussi question des données, dont le traitement reste obscur, surtout lorsqu'elles permettent à Google de traiter les résultats. Une enquête de la Commission nationale de l'informatique et des libertés (Cnil) a conclu que Google ne renseignait pas assez sur l'utilisation de ces données, raison pour laquelle cette autorité l'a condamné à une amende, mais sans chercher plus à régler ce problème (p. 92). En outre, jusqu'en 2011, via Google Analytics, Google a pu fournir aux éditeurs des données précieuses et très précises sur les requêtes effectuées (p. 93). Depuis, la firme ne le peut plus. L'auteur évoque aussi, dans un encart (p. 94), le projet Google Books et les difficultés législatives inhérentes.

- 8 Dans le chapitre « Alternatives et perspectives » (pp. 97-106), l'auteur rappelle que, même si le monopole de Google est avéré, certains moteurs tentent de le contrer. Un panorama des autres moteurs est dressé dans un tableau (p. 98) reprenant une typologie spécifique : moteurs généralistes, métamoteurs, moteurs géographiquement situés (langue ou espace particulier), sans publicité, solidaires, spécialisés accompagnés (p. 102). Quel sera l'avenir de la recherche d'information ? Il se situera au niveau des interactions sociales de plus en plus nombreuses (p. 103) et sera fortement lié à l'espace-temps, heure et lieu de la requête (p. 104) sans nécessairement changer d'interface. Web sémantique et plateforme de contenus sont au programme avec des avancées majeures sur la reconnaissance photo ou vidéo (p. 106).
- 9 Le moteur de recherche est donc multiple : physique (serveurs, câbles, routeurs), logiciel (*crawlers* et algorithmes), social (interactions), économique (organisation), informationnel (mots, images, sons...) et communicationnel. Mais il demeure entre « machine et machination » (p. 108) parce qu'il est difficile de réellement savoir ce qui se cache derrière son interface. Les prochains moteurs engloberont encore plus d'informations, donc de réponses, et il est primordial de chercher dès maintenant à les comprendre. Mais, bien que ce livre soit éclairant sur ce que sont les moteurs de recherche, il reste assez technique et obscur sur certains points, justement parce que les moteurs ne se dévoilent jamais entièrement.

AUTEURS

FLORENCE MICHET

Mica, université Bordeaux Montaigne, F-33607
florence.michet[at]etu.u-bordeaux-montaigne.fr