

Moteurs de recherche

Usages et enjeux

Brigitte Simonnot



Édition électronique

URL : <http://journals.openedition.org/questionsdecommunication/592>

DOI : 10.4000/questionsdecommunication.592

ISSN : 2259-8901

Éditeur

Presses universitaires de Lorraine

Édition imprimée

Date de publication : 1 décembre 2008

Pagination : 7-15

ISBN : 978-2-86480-981-4

ISSN : 1633-5961

Référence électronique

Brigitte Simonnot, « Moteurs de recherche », *Questions de communication* [En ligne], 14 | 2008, mis en ligne le 11 janvier 2012, consulté le 30 avril 2019. URL : <http://journals.openedition.org/questionsdecommunication/592> ; DOI : 10.4000/questionsdecommunication.592

BRIGITTE SIMONNOT

Centre de recherche sur les médiations

Université Paul Verlaine-Metz

brigitte.simonnot@univ-metz.fr

MOTEURS DE RECHERCHE. USAGES ET ENJEUX

Alors qu'ils ont à peine plus de dix ans, les moteurs de recherche Web sont devenus familiers et parfois indispensables. Leur usage s'est banalisé dans des situations très diverses de la vie quotidienne, que ce soit en contexte professionnel ou privé (Savolainen, 1995). Qu'il s'agisse de planifier un déplacement, de s'informer sur l'actualité, sur des questions de santé, lors d'étapes importantes de la vie personnelle ou plus simplement pour les loisirs, les individus se tournent de plus en plus souvent vers les ressources en ligne. Les moteurs de recherche commerciaux capitalisent cette audience : de nombreuses enquêtes montrent que l'activité de recherche d'information (RI) arrive en bonne place dans les usages de l'internet, désormais pratiquement à égalité avec les outils de communication comme la messagerie. Selon Comscore Inc.¹, pour le seul mois de juin 2008, 11,5 milliards de requêtes ont été soumises par les Américains vers les cinq principaux sites qui se partagent l'essentiel de la recherche d'information en ligne : Google (61,5 %), Yahoo (20,9 %), Microsoft (9,2 %), Ask (4,3 %) et AOL (4,1 %). En France, en mai 2008², ce sont 2,9 milliards de requêtes qui ont été soumises par 26 millions d'internautes, soit en moyenne 3,6 recherches

¹ Étude menée sur un panel de plus de 2 millions d'internautes : « ComScore Releases June 2008 U.S. Search Engine Rankings », *Comscore Inc.*, <http://www.comscore.com/press/release.asp?press=2337> (consulté le 20/07/08).

² Voir « comScore Releases May 2008 French Search Rankings », *Comscore Inc.*, 18/07/08 <http://www.comscore.com/press/release.asp?press=2327> (consulté le 20/07/08).

par jour, avec une part de 82 % pour Google. Les chiffres donnés par Médiamétrie/Netratings pour le mois de mai 2008 sont proches : selon cette étude³, Google a fait l'objet de 26,6 millions de visiteurs uniques en France. Ces résultats reflètent la banalisation de l'usage des moteurs de recherche dans les pratiques des internautes et soulignent les enjeux sous-jacents en termes de marketing publicitaire. Ils montrent aussi que, si le nombre de moteurs en ligne est considérable⁴, une très petite partie d'entre eux se partage les faveurs du grand public.

Alors que les recherches scientifiques en RI étaient auparavant cantonnées au domaine des bibliothèques, de l'information scientifique et technique ou stratégique, le développement de l'information en ligne leur a donné un nouvel essor. Le corpus très particulier auquel les moteurs de recherche sont dédiés – hétérogène, instable et en croissance très rapide – tout comme la nature des usages dont ils font l'objet ont obligé leurs concepteurs à étendre leur champ d'action à toutes sortes d'informations, à tout ce qui peut présenter un intérêt potentiel pour quelqu'un. Techniquement, un moteur de recherche en ligne repose sur quatre composants principaux : un robot de collecte (*spider* ou *crawler*) qui parcourt la Toile pour repérer les documents et les mémoriser dans une base de données, un module d'indexation qui les caractérise, notamment à partir des mots clés qu'ils contiennent ou des citations dont ils font l'objet, une interface utilisateur qui permet à l'internaute de saisir sa requête et de consulter les résultats et, enfin, un module d'appariement qui, pour une requête donnée, prépare un classement des pages susceptibles de répondre aux demandes⁵. Chacun de ces modules comporte lui-même plusieurs composants qui raffinent le traitement réalisé à chaque étape. Le grand public a une vision réduite de la complexité de ces dispositifs dont les concepteurs s'attachent à simplifier toujours plus l'utilisation.

Les recherches en sciences humaines et sociales sur les usages des moteurs restent rares en France alors qu'elles font l'objet d'une littérature anglo-saxonne abondante et de publications nombreuses dans les revues scientifiques spécialisées (notamment *Information Processing & Management*, *Journal of the American Society for Information Science and Technology*, *Journal of Documentation* ou *Information Research*). Pour ce dossier de la revue *Questions de communication*, deux grands types de contributions ont été retenues : d'une part, des travaux qui s'appuient

³ Données collectées sur un échantillon de 8500 internautes âgés de 2 ans et plus. Voir « L'audience de l'internet en France en mai 2008 », Médiamétrie/Netratings, 30/06/08, http://www.mediametrie.fr/resultats.php?rubrique=net&resultat_id=557 (consulté le 20/07/08).

⁴ Voir par exemple (Tjondronegoro, Spink, 2008)

⁵ Pour une synthèse, voir l'ouvrage de V. Mesguich et A. Thomas (2006).

sur un état de l'art dans le domaine des analyses d'usage des moteurs et en proposent une lecture enrichie ; d'autre part, des résultats récents de travaux empiriques d'analyse des moteurs ou d'observation des usages dont ils font l'objet.

On peut noter que la terminologie francophone de ce domaine de recherche est relativement pauvre à l'heure actuelle. L'expression « recherche d'information » est employée indifféremment pour désigner le domaine de l'*information search* qui, en anglais, se réfère à l'activité de recherche, à la quête d'information des usagers, et celui de l'*information retrieval*, dénomination introduite par le chercheur américain Calvin Mooers (Chiaromella, Mulhem, 2007 : 14) qui signifie littéralement « récupération d'information » et désigne l'ensemble des méthodes et algorithmes permettant la collecte, l'indexation, l'appariement et le classement des documents par rapport à des requêtes. Quant à l'expression communément utilisée de « moteur de recherche », traduction de l'anglais *search engine*, elle reste peu probante pour désigner ces services, notamment lorsqu'ils ont une vocation commerciale⁶ : le terme *engine* fait référence tout à la fois à la mécanique et à l'ingénierie sous-jacente, cette dernière notion n'étant pas apparente dans sa traduction française. Dans notre langue, un moteur est un appareil qui transforme en énergie mécanique toute autre forme d'énergie, une force qui dirige le mouvement⁷. Au figuré, c'est aussi ce qui incite ou pousse à agir. Les moteurs, en facilitant la recherche d'information, incitent-ils les internautes à s'informer davantage ? En 1959, Calvin Mooers (in Austin, 2001 : 607-608) dénonçait la croyance selon laquelle mettre à la disposition des individus des systèmes de RI suffirait à ce qu'ils soient utilisés : il émettait l'hypothèse que de nombreux individus ne veulent pas d'information et évitent d'utiliser un système qui, précisément, leur en donnerait. Disposer d'une information peut être pénible et incommode puisqu'il faut la lire, essayer de la comprendre et y réfléchir, activités pour lesquelles les systèmes de RI ont longtemps été de peu d'utilité. Les moteurs de recherche, s'ils ne permettent pas forcément de mieux s'informer, contribuent à mettre à la disposition du grand public des informations qui, sans eux, seraient probablement restées confidentielles.

⁶ À l'instar des Canadiens qui nomment « fouineurs » ou « fureteurs » les navigateurs Web, nous pourrions peut-être utiliser l'expression « prospecteurs Web » pour désigner ces outils afin de rendre mieux compte de leur réalité.

⁷ Voir les définitions données par le Trésor de la Langue française informatisé (TLFI), <http://www.cnrtl.fr/definition/moteur>

Analyser les usages des moteurs

Les analyses d'usages des moteurs de recherche Web ont deux principaux objectifs : mieux comprendre l'activité informationnelle humaine et améliorer le fonctionnement des outils et la conception des interfaces. La contribution de Madjid Ihadjadene et Stéphane Chaudiron⁸ s'intéresse aux méthodologies mises en œuvre dans ces études à partir d'un état de l'art étendu sur la question. Leur contribution est doublement intéressante : d'une part, au vu du grand nombre de travaux consacrés à ces questions et leur disparité ; d'autre part, du point de vue des conclusions qu'ils en tirent. Ils distinguent quatre grandes catégories de recherches. Les deux premières s'inscrivent dans le cadre des études de comportements des utilisateurs de moteurs : elles cherchent à mettre en évidence, d'un côté, les tactiques et les stratégies qu'ils mettent en place, de l'autre, les manières dont ils lisent et évaluent les résultats. Ces études sont confrontées à deux difficultés : pour que les résultats soient représentatifs, elles doivent porter sur un grand nombre de données. Or, celles-ci sont difficilement accessibles aux chercheurs scientifiques indépendants. Les historiques de transaction des moteurs (*transaction logs*) livrent des informations qui peuvent sembler succinctes : l'adresse IP d'où émanent les requêtes, la date et l'heure, les mots clés et les opérateurs employés par les internautes pour leurs demandes, parfois les options choisies. Ils permettent néanmoins aux chercheurs scientifiques qui ont la chance de pouvoir y accéder, et aux concepteurs de moteurs, d'analyser la manière souvent très intuitive dont les internautes formulent leurs requêtes et consultent les résultats. Une troisième catégorie de travaux, menés principalement en psychologie cognitive, cherche à différencier les usagers, en élaborant des typologies d'utilisateurs ou des typologies d'usages. Dans cette catégorie, la différenciation la plus commune est sans doute celle opérée entre experts et novices. Madjid Ihadjadene et Stéphane Chaudiron montrent que nombre de ces études souffrent d'insuffisances, notamment du fait d'une définition floue ou inconsistante des notions d'expert ou de novice. Dernière catégorie traitée : les travaux qui s'intéressent plus globalement au contexte de la recherche d'information et qui s'inscrivent dans le courant général de l'étude du comportement informationnel (*information behaviour*) des individus. De cet état de la littérature scientifique, les auteurs concluent que les méthodologies sont encore imparfaites et parcellaires. Ils préconisent une approche holistique permettant de prendre en compte la complexité des variables qui interviennent dans les activités de RI des internautes.

⁸ Cette contribution a été reçue en mars 2007 et la version définitive acceptée en novembre 2007.

Les jeunes sont des utilisateurs coutumiers des moteurs à tel point que l'on parle parfois de « génération Google » pour désigner ceux qui sont nés après 1993 (CIBER, 2008). Dans ce cadre, Nicole Boubée analyse les stratégies des jeunes chercheurs d'information en ligne. Son étude, qui repose sur des observations directes d'élèves de collèges et de lycées en situation de recherche d'information et sur des entretiens d'explicitation, remet en cause certains résultats antérieurs, notamment quant à l'affirmation de la préférence chez les jeunes pour la navigation par rapport à la recherche « directe » par requête. Revenant sur la distinction trop convenue novice/expert, elle s'attache à montrer les inventions et la créativité de ces jeunes dans leurs pratiques des moteurs, lorsqu'ils tentent d'élaborer des stratégies de recherche pour atteindre leurs buts ou de comprendre le fonctionnement des applications. Elle dénonce une « perspective théorique dominante » susceptible d'enfermer dans la stigmatisation des lacunes des jeunes en matière d'activité informationnelle et appelle à développer des recherches qui prennent réellement en compte la complexité de cette activité que les moteurs commerciaux s'attachent toujours à simplifier davantage, du moins en apparence.

À la recherche du moteur parfait

Les deux contributions suivantes s'intéressent aux fonctionnalités que devraient comporter le moteur de recherche parfait. Judit Bar-Ilan se situe dans le contexte très spécifique des usages des chercheurs qui analysent le Web, c'est-à-dire dans le domaine de la webométrie ou de la webologie. C'est l'occasion pour elle de revisiter la liste des fonctions idéales qu'elle a proposée dans un article précédent (Bar-Ilan, 2005), d'analyser si ces fonctionnalités sont offertes ou non par les principaux moteurs commerciaux et si oui, sous quelle forme. Son étude met en évidence que ces services ne sont pas des outils adaptés à la webométrie ni à la recherche scientifique sur le Web. Les chercheurs qui exploitent ces applications pour constituer leurs corpus doivent en être avertis : les données collectées de manière empirique par les moteurs commerciaux ne peuvent en aucun cas être considérées comme représentatives du Web tout entier. Les résultats des moteurs sont aussi caractérisés par une forte instabilité dans le temps et des disparités dans leur couverture des ressources des différents pays. Les principaux moteurs commerciaux ont été conçus pour analyser des documents en anglais et la façon dont ils ont été adaptés à la recherche dans d'autres langues reste peu claire. L'auteur impute certaines de ces limites au fait que ces moteurs sont destinés au grand public – pour qui ces fonctions seraient moins importantes, les chercheurs en webométrie ayant une exigence quantitative quant aux résultats (Thelwall, 2008) – mais aussi à la vocation commerciale de ces services. En effet, les moteurs commerciaux entretiennent le secret sur le

détail de leurs algorithmes pour des raisons de concurrence⁹ et de lutte contre les spameurs. Le manque de transparence et les inconsistances de ces applications obligent à la plus grande prudence lorsque l'on souhaite les exploiter à des fins scientifiques.

Dirk Lewandowski et Nadine Höchstötter prolongent et élargissent ces remarques en proposant un cadre de référence pour « mesurer la qualité des moteurs de recherche Web » et les comparer entre eux. S'appuyant à la fois sur un état de l'art et sur leurs propres investigations, ils montrent les difficultés auxquelles les chercheurs sont confrontés dans ce domaine. Ils soulignent eux aussi les problèmes méthodologiques posés par certaines études, qui empêchent de généraliser ou de comparer les résultats obtenus. Les auteurs plaident pour que s'intensifient les recherches en la matière en Europe, de manière à ce que les différents pays disposent d'outils adaptés à leurs propres ressources informationnelles. Si, dans le domaine du patrimoine culturel, le plaidoyer de Jean-Noël Jeanneney (2005) a été entendu avec l'initiation du projet de bibliothèque numérique Europeana, peu d'acteurs européens ont réussi à s'imposer dans le domaine de la RI généraliste où les enjeux économiques sont pourtant importants. Le cadre d'évaluation proposé prend en compte à la fois les performances techniques des moteurs et leur adaptation aux besoins des internautes, suivant en cela l'appel de Kalervo Jäverlin et Peter Ingwersen (2004 ; Ingwersen, Jäverlin, 2005). Comme en attestent ces deux études, les moteurs peuvent encore faire l'objet de nombreuses améliorations non seulement techniques, mais aussi quant à l'information qu'ils donnent sur leur mode de fonctionnement et à l'utilisabilité de leurs interfaces.

Derrière l'interface des moteurs

L'interface du moteur Google est souvent citée comme un exemple d'ergonomie. Brigitte Simonnot relève que cette interface fait appel au jeu sous plusieurs formes, une manière peut-être d'attirer la sympathie des utilisateurs, de créer une connivence. Elle analyse aussi certains jeux créés par et autour du célèbre moteur. Ils sont souvent représentatifs non seulement des enjeux économiques – référencement en ligne, l'autre face des moteurs commerciaux –, mais aussi des enjeux politiques et éthiques du secteur. Si l'objectif principal des systèmes de RI traditionnels était de permettre à un usager de trouver une information pertinente qui réponde

⁹ Par exemple, si les moteurs donnaient leurs statistiques exactes sur les liens pointant vers un site donné, tout bon mathématicien pourrait chercher par approximation à en déduire le poids accordé à ces liens dans le calcul du rang de classement des pages, ce qui permettrait de découvrir une partie des arcanes des algorithmes.

à son besoin, celui des moteurs commerciaux est en effet plus ambigu : les stratégies marketing des concepteurs s'adaptent en permanence aux usages des internautes qui, à leur tour, s'approprient les applications et les détournent parfois des usages attendus. En facilitant l'activité de recherche d'information, voire en la rendant ludique, les prestataires de moteurs entretiennent une confusion sur leurs objectifs réels et travaillent sur la séduction, probablement pour améliorer la confiance des utilisateurs en leurs applications et les fidéliser¹⁰.

Certes, les moteurs s'attachent à faciliter les recherches d'information en ligne et aident à se repérer dans les innombrables contenus disponibles. Cependant, Michael Zimmer s'interroge sur le « pacte faustien » que peut représenter la recherche de la perfection dans la conception de moteurs, ceux qui visent à donner accès à la connaissance universelle – grâce à une couverture la plus étendue possible des contenus Web – et à se doter d'une capacité de remémoration propre à discerner les intentions des utilisateurs (Jansen *et al.*, 2008). L'auteur souligne combien les risques pour la vie privée sont accrus par la mise à disposition du grand public d'applications robustes de traitement des données audiovisuelles, par le développement du Web 2.0 – le Web comme plateforme universelle de traitement des données (O'Reilly, 2005) et des réseaux sociaux – et l'interopérabilité de ces technologies avec les moteurs de recherche. Il suggère un ensemble de mesures qui devraient être prises pour protéger l'intimité des individus et respecter leur vie privée. En effet, en cherchant à personnaliser toujours mieux les résultats, les concepteurs de moteurs sont conduits à recueillir systématiquement toutes les données susceptibles d'améliorer les performances des applications, données sur lesquelles ils peuvent mettre en œuvre des algorithmes de fouille sophistiqués. Celles-ci décrivent non seulement ce que font les internautes, mais elles peuvent aussi fournir une foule de détails sur leurs centres d'intérêt et sur leurs idées, ce qui constitue une menace réelle pour le respect de la vie privée. Les moteurs de recherche participent à une forme d'« économie de la visibilité » (Heilmann, 2007 : 310), sous le prétexte de prévenir non pas d'éventuels problèmes de sécurité publique mais les besoins, souhaits et désirs des chercheurs d'information en ligne. En France, la Commission nationale de l'informatique et des libertés (CNIL) fête cette année ses 30 ans, conjointement avec l'institution homologue allemande. Sa mission est essentielle : informer les citoyens sur leurs droits et obligations quant à la protection des libertés et de la vie privée, veiller au respect de la législation par les concepteurs d'applications et au droit d'accès de chacun aux données informatisées qui le concernent. Encore faut-il que

¹⁰ Certaines études montrent une corrélation entre l'esthétique des sites et leur crédibilité auprès des internautes (Robins, Holmes, 2008).

les citoyens soient avertis des données qui sont recueillies souvent à leur insu. Les concepteurs d'applications devraient sans aucun doute accroître leur effort pour une meilleure information sur ce qu'ils recueillent ou produisent, et faciliter l'exercice du droit d'accès et de rectification aux données personnelles, tout en prévoyant un « droit à l'oubli ».

Conclusion

À l'heure où les concepteurs de moteurs travaillent à mettre toute l'information disponible à portée de clic, il n'est pas question de tomber dans un soupçon généralisé vis-à-vis de ces applications, fort utiles au demeurant. Mais connaître les limites de ces outils et dénoncer le mythe d'une activité informationnelle qui serait simple, facile et sans contrepartie est un premier pas vers une régulation des pratiques dans ce secteur. Doter la langue française d'un vocabulaire plus précis dans ce domaine scientifique est certainement un autre chantier à approfondir. Toutefois, la traduction de trois articles soumis en anglais – ceux de Judit Bar-Ilan, Dirk Lewandowski et Nadine Höchstätter, et Michael Zimmer – s'est efforcée de rester au plus près du vocabulaire établi.

De plus, les travaux en sciences humaines et sociales participent aussi à analyser plus finement les conséquences sociales et les enjeux éthiques et culturels de ces applications. Certes, des processus de rationalisation et de normalisation de la publication en ligne sont en cours, en particulier dans le cadre du Web sémantique. Il reste néanmoins beaucoup à faire dans ce secteur où, comme souvent, les innovations technologiques précèdent la réflexion sur leurs conséquences. Puissent se développer les publications en français de travaux dans ce domaine, un élément stratégique pour amener davantage de chercheurs francophones à développer leurs recherches et les jeunes chercheurs à s'intéresser à ces questions.

Références

- Austin B., 2001, « Mooers' Law : In and Out of Context », *Journal of the American Society for Information Science and Technology*, 52 (8), pp. 607-609.
- Bar-Ilan J., 2005, « Expectations versus reality - Search engine features needed for Web research at mid 2005 », *Cybermetrics*, 9(1), <http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html>, consulté le 25/08/08.
- Chiarabella Y., Mulhem P., 2007, « La recherche d'information. De la documentation automatique à la recherche d'information en contexte », *Document numérique*, 10(1), pp. 11-38.
- CIBER, 2008, *Information behaviour of the researcher of the future*, Londres, UCL, <http://www.bl.uk/news/pdf/googlegen.pdf>, consulté le 26/08/08.

- Heilmann É., « Surveiller (à distance) et prévenir: Vers une nouvelle économie de la visibilité », *Questions de communication*, 11, pp. 303-322.
- Ingwersen P., Järvelin K., 2005, *The Turn : Integration of Information Seeking and Retrieval in Context*, Dordrecht, Springer.
- Jansen B.J., Booth D.L., Spink A., 2008, « Determining the informational, navigational, transactional intent of Web queries », *Information Processing & Management*, 44(3), pp. 1251-1266.
- Järvelin K., Ingwersen P., 2004, « Information seeking research needs extension towards tasks and technology », *Information Research*, 10(1), paper 212, <http://informationr.net/ir/10-1/paper212.html> (consulté le 22/08/08).
- Jeanneney J.-N., 2005, *Quand Google défie l'Europe. Plaidoyer pour un sursaut*, Paris, Éd. Mille et une nuits.
- Mesguich V., Thomas A., 2006, *Net recherche : le guide pratique pour mieux trouver l'information utile*, Paris, Éd. ADBS.
- O'Reilly T., 2005, « What is Web 2.0 - Design Patterns and Business Models for the next generation of software », *O'Reilly Network*, 30/09/05 <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (consulté le 27/09/06).
- Robins D., Holmes J., 2008, « Aesthetics and credibility in web site design », *Information Processing & Management*, 44 (1), pp. 386-399.
- Savolainen R., 1995, « Everyday life information seeking: Approaching information seeking in the context of "way of life" », *Library & Information Science Research*, 17(3), pp. 259-294.
- Thelwall M., 2008, « Quantitative Comparisons of Search Engines », *Journal of the American Society for Information Science and Technology*, 59 (11), pp. 1702-1710.
- Tjondronegoro D., Spink A., 2008, « Web search engine multimedia functionality », *Information Processing & Management*, 44(1), pp. 340-357.