



HAL
open science

Quelles analyses de l'usage des moteurs de recherche ? Questions méthodologiques

Madjid Ihadjadene, Stéphane Chaudiron

► To cite this version:

Madjid Ihadjadene, Stéphane Chaudiron. Quelles analyses de l'usage des moteurs de recherche ? Questions méthodologiques. Questions de communication, 2008, 14, pp.17-32. 10.4000/questionsde-communication.604 . hal-00468777

HAL Id: hal-00468777

<https://hal.science/hal-00468777>

Submitted on 31 Mar 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**QUELLES ANALYSES DE L'USAGE DES MOTEURS DE
RECHERCHE ?
QUESTIONS METHODOLOGIQUES**

MADJID IHADJADENE

Centre de recherche en information spécialisée
Université Paris 10
ihadjade@u-paris10.fr

STEPHANE CHAUDIRON

Groupe d'étude et de recherche interdisciplinaire en information et communication
Université Lille 3
stephane.chaudiron@univ-lille3.fr

Article paru dans *Questions de communication*, 2008, n°14, p. 17-32

Résumé. — Les évaluations des moteurs de recherche sur l'internet sont généralement effectuées selon trois approches : sous l'angle de leurs performances globales, en s'inscrivant dans le paradigme des tests de Cranfield, ou encore selon la perspective de l'analyse du comportement de l'utilisateur durant la session d'interrogation. Dans cette contribution, on se focalisera sur la démarche de consultation des usagers, l'identification des difficultés qu'ils rencontrent, l'exploration des modes d'appropriation des outils de navigation, de catégorisation, de visualisation et, enfin, sur l'étude des stratégies de formulation et de reformulation des requêtes.

Mots clés. — Moteurs de recherche, approche usager, reformulation, catégorisation, évaluation, analyse des *logs*.

Depuis l'étude pionnière de Craig Silverstein (1998) sur le moteur de recherche Altavista jusqu'aux récents travaux de Sherry Koshman, Amanda Spink et Bernard J. Jansen (2006) sur le métamoteur Vivisimo, plus d'une centaine d'études ont été consacrées, principalement dans le monde nord-américain, à l'analyse des pratiques d'accès des internautes aux sources d'information en ligne. En France, plusieurs auteurs se sont également intéressés à cette question. Citons Thomas Beauvisage, Valérie Beaudoin et Houssein Assadi (2007) qui ont utilisé la technique des panels pour effectuer une analyse des pratiques des usagers selon leur profil sociodémographique (âge, sexe, CSP, socio-style, etc.), ou encore Josiane Jouet (2003) qui propose une analyse critique de ces dispositifs de mesures et d'études d'audience. Ces travaux se caractérisent par leur diversité, tant méthodologique qu'en lien avec les objectifs qui les animent – scientifiques et commerciaux –, et les cadres théoriques dans lesquels elles s'insèrent. À côté des études de nature scientifique, il en existe également de nombreuses autres, réalisées par des sociétés de conseil en médiamétrie (Nielson/Netrating, Xiti, etc.), dont les résultats sont notamment présentés dans l'ouvrage de François-Xavier Hussherr et Sophie Neron (2002).

Quelques grandes tendances émergent. Ainsi constate-t-on que, désormais, l'utilisation d'un moteur de recherche est une pratique courante de la vie quotidienne des individus et que les internautes ont tendance à être fidèles à un voire deux outils. Un peu plus d'un cinquième du trafic de l'internet est généré par les outils de recherche (Hussherr, Neron, 2002). La majorité des usagers sont également confiants dans la probité des annuaires et des moteurs de recherche. En revanche, il apparaît qu'ils sont peu au fait des implications politiques, culturelles et économiques liées à l'utilisation de ces outils ; par exemple, selon l'étude de Deborah Fallows pour le *Pew Internet & American Life Project*¹, en 2005, 57 % des internautes ignoraient que les moteurs conservent des données relatives à leur navigation et que celles-ci peuvent être réutilisées dans un cadre commercial.

Dans le cadre de cette contribution, le premier objectif est de savoir ce qu'il en est de l'usage des moteurs de recherche et, plus précisément, la manière dont les études visant à identifier et expliciter ces usages correspondent à des objectifs plus ou moins avoués (elles ne sont pas neutres, sont souvent limitées, parfois discutables dans leurs fondements méthodologiques et, de plus, difficilement comparables). Dans ce contexte, la réflexion est triple.

¹ L'étude est disponible à http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf

Il s'agit d'abord d'identifier les grands courants auxquels correspondent ces études. Nous en avons identifié quatre et les avons structurés de la manière suivante : le premier correspond à l'analyse des tactiques et stratégies de recherche mises en œuvre par les usagers ; le deuxième à l'analyse des modes de lecture et de qualification des résultats ; le troisième à l'analyse par catégorie d'usagers ; le quatrième à l'analyse par différenciation du contexte. Les deux premiers axes correspondent aux approches qui utilisent l'analyse des *logs*² comme outil principal de collecte des données. Les deux derniers privilégient des méthodologies expérimentales et/ou qualitatives pour le recueil des données.

Le deuxième objectif est de montrer que cette grande variabilité méthodologique influe directement sur les résultats de l'analyse des données et pose des problèmes d'interprétation. En effet, la méthode choisie pour recueillir les données influe plus particulièrement sur l'analyse des usages. Elle n'est donc pas neutre, mais dépend des objectifs de l'analyse qui, pour leur part, sont plus ou moins explicites. Certains sont de nature commerciale (toucher une cible pour affiner une stratégie marketing), d'autres visent à améliorer les moteurs. En ce sens, nous ne nous situons pas dans le cadre de l'évaluation de la performance des moteurs (contrairement aux campagnes d'évaluation de type TREC³) ni de l'analyse comparative des fonctionnalités de recherche.

Le troisième objectif vise à montrer que les études sur les moteurs de recherche s'inscrivent à la fois dans un mouvement de continuité par rapport aux études consacrées aux cédéroms, aux catalogues de bibliothèques en ligne (OPAC) et aux serveurs commerciaux, et dans un processus de rupture. Indiscutablement, la question de la recherche d'information n'est pas nouvelle mais on peut avancer qu'elle se pose en des termes neufs. Rappelons les études notamment publiées dans la revue *CD-Rom and Online Review* dans les années 80⁴ et qui portaient principalement sur l'usage professionnel des serveurs d'informations tel que Dialog ou STN puis sur les usages des cédéroms d'information spécialisée. Aujourd'hui, on constate une rupture, induite par un certain nombre de mutations, dont l'extraordinaire croissance du nombre d'usagers des outils de recherche, leur appropriation par le « grand public » – qui

² Les *logs* sont des fichiers de type journaux où sont enregistrées les opérations effectuées sur un serveur par des utilisateurs ou des processus. Outre la date et l'heure de connexion, ils donnent l'adresse IP des internautes et le texte de leurs commandes.

³ *Text REtrieval Conference*, <http://trec.nist.gov>

⁴ En réalité, les premières études sur les usages des serveurs commerciaux datent de 1972 lorsque ceux-ci ont été connectés sur les réseaux de communication spécialisés tels que Tymnet et Telenet.

influe désormais, même si c'est indirectement, sur l'évolution technologique des moteurs, y compris à usage professionnel –, la volonté d'accéder directement au contenu informationnel primaire (le texte intégral) et non plus aux seules références bibliographiques, une évolution des habiletés informationnelles requises (les internautes n'ont plus besoin de maîtriser les langages documentaires pour interroger un moteur de recherche).

Dans un premier temps, nous présenterons les quatre axes identifiés, en ne retenant que ceux qui permettent de discuter quelques points saillants. Introduites dans la première partie, ces questions seront synthétisées dans la seconde.

Présentation critique des études

Les études recensées sont regroupées en quatre axes qui correspondent à la fois aux « objets » analysés, aux variables étudiées et à la méthodologie retenue.

Tactiques et stratégies de recherche mises en œuvre par les usagers

Compte tenu de la spécificité méthodologique de l'approche, les études qui relèvent de la première catégorie permettent de caractériser le comportement des internautes à travers la diversité des fonctionnalités des moteurs de recherche. Elles présentent des résultats sur le nombre de sessions et de mots clés saisis, la reformulation des requêtes, l'usage des opérateurs, etc. Dans leur ouvrage sur les usages des moteurs de recherches, Amanda Spink et Bernard J. Jansen (2004) montrent que les requêtes des usagers ne comportent qu'un, voire deux mots. Les utilisateurs n'ont presque jamais recours aux fonctionnalités avancées, et il est fréquent de trouver des fautes d'orthographe dans la requête. Les auteurs constatent également que la grande majorité des requêtes ne contiennent pas d'opérateurs booléens (5 à 10 % seulement en contiennent) et que les usagers ont encore des difficultés à maîtriser l'usage des opérateurs. Ils montrent également que l'opérateur booléen AND est de loin le plus utilisé et que les utilisateurs recourent très rarement à la possibilité de créer des expressions (termes entre guillemets).

Des travaux plus récents sur d'autres moteurs de recherche confirment ces données. Bernard J. Jansen et Amanda Spink (2006) présentent une comparaison de neuf moteurs – quatre européens et cinq américains – sur une période de sept ans. On constate des similitudes et des différences dans chacune des utilisations. Les utilisateurs réagissent de la même manière pour

ce qui est de la durée des sessions, de la taille des requêtes et du nombre de pages de résultats lus. En revanche, les opérateurs de requêtes américains sont un peu plus utilisés par les usagers des moteurs américains (quel que soit leur lieu de connexion). Dans une session, la distribution du nombre de requêtes indique qu'environ 71 % des sessions se composent d'une ou deux questions.

De leur côté, Soyeon Park, Joon Ho Lee et Hee Jin Bae (2005) ont analysé le comportement de recherche d'informations sur la toile des utilisateurs coréens du moteur de recherche Naver. À partir du recueil de *logs* portant sur plus de 40 millions de requêtes collectées sur une semaine, ils montrent que 83 % des sessions se composent d'une ou deux requêtes et que près de 90 % de celles-ci sont composées d'un seul mot. Dans le cas d'AOL (Beitzel, 2006), le nombre de termes par question est de 2,7 alors qu'il est de 3,14 pour le métamoteur Vivisimo (Koshman *et al.*, 2006). La même étude montre que la plupart des utilisateurs conduisent plusieurs recherches successives sur un même sujet mais de manière discontinue, modifiant plusieurs fois leur requête et complétant leurs informations. Concernant la reformulation des requêtes, Amanda Spink et Bernard J. Jansen (2004) montrent que près de la moitié des requêtes sur le moteur Excite sont modifiées. Cette reformulation porte d'abord sur la substitution d'un ou plusieurs termes de la requête, puis sur un ajout et, enfin, sur une suppression. Bien qu'Excite soit un système capable de retourner des pages similaires à un lien donné en résultat, il s'avère que seulement 5 % des utilisateurs utilisent cette possibilité (Jansen, Spink, 2006).

D'autres études se sont intéressées aux aspects multilingues des moteurs de recherche. Ainsi Hae-Young Rieh et Soo Young Rieh (2005) montrent-ils que les chercheurs n'utilisent pas les capacités multilingues des moteurs : ils en font un usage monolingue, essentiellement en anglais. À cet égard, on peut se demander si cette pratique linguistique est due aux contraintes techniques des moteurs de recherche ou à la représentation que s'en font les usagers. Comme le montre Judit Bar-Ilan (2005), les trois moteurs – Alltheweb, Altavista et Google – ignorent les caractéristiques des langues autres que l'anglais, en particulier les signes diacritiques. De son côté, Haidar Moukdad (2004) a testé six moteurs (Alltheweb, Altavista, Google, Albahhar, Ayna, Morfix) et montré que l'utilisation des formes exactes de mots arabes dans les moteurs généralistes aboutit à une perte d'information, du fait de l'absence d'un analyseur morphologique dans ces systèmes. L'ensemble des recherches conduites montrent que les

pratiques d'accès sont en deçà des fonctionnalités proposées. Cela correspond aux usages des systèmes d'informations traditionnels (Markey, 2007).

Au-delà des résultats concernant le nombre de sessions, de mots clés utilisés ou l'usage des opérateurs booléens, l'analyse des *logs* permet de constater des pratiques « détournées » de certaines fonctionnalités des moteurs de recherche. En effet, ces derniers sont parfois utilisés comme des outils de vérification orthographique ou comme des outils de localisation d'un site dont on connaît l'URL, mais que l'on ne souhaite pas saisir dans le champ approprié (Ravid *et al.*, 2007). Ce mode d'appropriation correspond à une forme de « braconnage » de fonctionnalités pour reprendre un concept développé par Michel de Certeau (1990).

Lecture et qualification des résultats

Un deuxième groupe d'études est centré sur l'appropriation des résultats fournis par les moteurs et leur validation par les usagers. Beaucoup portent sur le nombre de pages lues, telle celle de Steve Beitzel (2006) qui montre que, dans le cas d'AOL, 79 % des usagers ne regardent que le premier écran, 15 % le second, et 6 % trois écrans ou plus. En moyenne, les usagers visualisent 1,7 page de résultats par session pour Excite, 2,2 pour Alltheweb (Jansen, Spink, 2006), 1,1 page pour le moteur coréen Never (Park, Bae, 2005).

Une analyse oculométrique (méthode du *eyetracking*) menée sur MSN et Google montre que si les internautes lisent linéairement les résultats d'un moteur de recherche, le temps de lecture décroît très rapidement entre le premier et le dernier résultat (Pan *et al.*, 2007). La lecture du premier dure 2,2 secondes, contre 0,2 seconde à 0,3 seconde pour le dernier. Ce sont donc les premiers résultats qui captent le plus l'attention de l'internaute. La même étude révèle que les utilisateurs de l'internet accordent une confiance exagérée au système de classement des résultats utilisé par le moteur de recherche Google. Ce constat a des conséquences importantes du point de vue des enjeux commerciaux, ce qui explique que certaines techniques de référencement de sites utilisent le *spamming*⁵ pour améliorer artificiellement le positionnement de leurs sites dans les pages de résultats.

⁵ Le *spamming* ou *spamdexing* désigne l'utilisation abusive des techniques de référencement destinées aux robots d'indexation de moteur de recherche, notamment en utilisant des mots clés dans des pages Web pour améliorer leur classement dans les moteurs de recherche.

L'une des solutions apportées pour aider les usagers à « lire » et naviguer dans l'espace des résultats est la catégorisation et la classification. C'est le cas de Vivisimo, étudié par Sherry Koshman, Amanda Spink et Bernard J. Jansen (2006). Ce métamoteur regroupe automatiquement les réponses, aussi bien en listes qu'en en catégories thématiques potentiellement intéressantes. L'analyse de l'interaction des usagers indique que la moitié seulement des interactions consiste en l'affichage d'une catégorie. En revanche, les usagers n'interagissent pas avec l'arborescence pour afficher les sous-catégories.

Une autre solution pour aider les internautes à consulter les résultats est proposée par les outils de visualisation de l'information. Mais il existe très peu d'études d'usage concernant ces outils. Walky Rivadeneira et Benjamin B. Bederson (2003) ont comparé trois systèmes : le moteur Crocker, sa version textuelle, et le métamoteur Vivisimo. Les usagers ont une préférence pour les versions textuelles. Christophe Tricot *et al.* (2006) ont réalisé une expérience mettant l'utilisateur dans une situation de recherche d'information à partir d'un système de visualisation de données hiérarchiques. Les arbres de cônes en 3D ont posé des problèmes pour les usagers novices. Alexandra Ciaccia *et al.* (2006) ont évalué deux moteurs cartographiques, Kartoo et Mapstan, et montrent que la visualisation semble cognitivement coûteuse : en général, elle est associée à des temps de traitement plus longs et à plus d'activités de manipulation, sans pour autant conduire à des meilleures performances, par comparaison avec une présentation en liste.

Un autre enjeu lié à la lecture à l'écran concerne la validation, par les usagers, des différents liens proposés par le moteur. À cet effet, de nombreuses études montrent que les usagers font peu de distinction entre les liens sponsorisés et les autres. Sur Google, près de 62 % des internautes ne font aucune distinction entre les informations publicitaires et les autres (*Pew Internet & American Life Project*, 2005). La présentation des résultats et des liens publicitaires est conditionnée par les stratégies d'indexation. Certains auteurs remettent en cause la neutralité de l'indexation et du classement des résultats ; à travers la notion de « *Googlearchy* », d'autres montrent que l'audience ne se concentre que sur les sites vers lesquels pointent le plus de liens. Pour Gloria Origgi (2003) : « La structure des liens produite culturellement par les choix individuels des créateurs de pages web est interprétée par les algorithmes de recherche comme une hiérarchie de valeur à travers les pages. La hiérarchie des résultats influencera à son tour les choix des utilisateurs en cercle vertueux de collaboration entre humains et automates ».

À l'inverse, Santo Fortunato *et al.* (2006) démontrent que, malgré cette tendance à l'uniformisation liée aux algorithmes de type Pagerank de Google, qui privilégient les sites les plus cités, la diversité des mots clés utilisés par les internautes produit « un effet d'égalisation ». Selon les auteurs, cet effet permet à des sites peu connus d'être néanmoins présents dans les pages de résultats. En ce sens, les grands sites d'information et les moteurs de recherche paraîtraient plus propices à la rencontre d'opinions opposées que d'autres sites comme les blogs.

Analyse par catégorie d'utilisateur

Un troisième groupe d'études, utilisant essentiellement des approches expérimentales, prend en compte les dimensions individuelle et cognitive des usagers. Ces dernières s'appuient par exemple sur une différenciation des usagers selon leur compétence conceptuelle (connaissance du domaine ou non) et leur savoir faire méthodologique (expertise en recherche d'informations), la typologie et la variabilité des tâches de recherche ou l'identification des styles cognitifs. Ainsi plusieurs études empiriques ont-elles pu mettre en évidence que les stratégies de recherche et de navigation sont soumises à des variabilités interindividuelles. De nombreux auteurs ont constaté que l'expertise dans un domaine était un facteur important pour la performance de recherche d'information (Marchionini, 1995).

Dans le cas des moteurs de recherches, des études (Tricot *et al.*, 2000 ; Lazonder *et al.*, 2000 ; Palmquist, Kim, 2000) ont tenté de répondre à cette question dans le cadre d'une expérience comparant une recherche d'information entre deux groupes dont l'un est composé de non-utilisateurs du web et l'autre d'utilisateurs réguliers, les deux étant considérés comme maîtrisant de la même façon le domaine sur lequel portait la recherche d'informations. Aucune différence significative entre les deux types de groupes n'a été mise en évidence, les utilisateurs réguliers du web n'ayant pas fait de recherches d'information plus performantes. Les chercheurs ont plus précisément remarqué que les utilisateurs familiers font un usage aussi intempestif qu'inefficace du site web : ouverture fréquente de pages non pertinentes, nombre très grand de liens activés. Par conséquent, une précision très faible caractérise leurs résultats et leur démarche semble irrationnelle et inefficace d'un point de vue informationnel. Ard Lazonder *et al.* (2000) montrent que les experts de l'internet parviennent à rechercher de l'information plus rapidement que les novices. Un des résultats intéressants de telles études

(Palmquist, Kim, 2000) est que l'expérience de recherche en ligne peut énormément réduire l'effet de dépendance vis-à-vis du domaine.

Ce résultat est corroboré par les travaux de Madjid Ihadjadene *et al.* (2003) qui suggèrent l'existence d'un mécanisme de compensation entre les deux types de compétences (dans le domaine et dans la recherche d'informations). D'une manière générale, les experts font davantage appel à leurs méta-connaissances, en ce sens qu'ils possèdent les connaissances et les heuristiques nécessaires à l'optimisation de leurs stratégies de recherche. Ainsi s'orientent-ils plus souvent vers des sites spécialisés pour lesquels ils font l'hypothèse qu'ils contiennent les réponses à leurs questions (Tabatabai, 2005). D'autres études montrent néanmoins que, malgré leur expérience, les experts ont rarement recours aux fonctionnalités avancées telles que les expressions booléennes, la troncature de mots ou les opérateurs de proximité lexicale dans leur interrogation des moteurs de recherche (Choo *et al.*, 2000).

Malgré son apparente facilité méthodologique, le modèle novice/expert sur lequel se fondent nombre d'évaluations n'est pas sans poser problème et présente plusieurs limites que nous ne ferons qu'évoquer. Les études donnent non seulement des résultats contradictoires mais, surtout, la notion d'expertise y est souvent mal définie et y est variable. Comme le montrent les typologies à la fois « flottantes » et hybrides des catégories (plus ou moins novice, intermédiaire, plus ou moins expert...), la frontière entre novice et expert n'est pas facile à établir.

Un autre aspect souvent étudié est l'analyse des tâches de recherche. Dépassant la dichotomie question ouverte/question fermée utilisée dans les études portant sur les systèmes de recherche classiques (catalogues, cédéroms, hypertextes...), plusieurs auteurs suggèrent de distinguer trois types de tâches dans l'analyse des moteurs de recherche sur l'internet. Ainsi Andrei Broder (2002) distingue-t-il trois objectifs dans le processus de recherche : trouver une page ou un site particulier comme un « ancrage » à la navigation, repérer une information nécessitant la consultation de plusieurs pages ainsi qu'une autre afin d'accomplir une action telle qu'un achat en ligne. Selon le même auteur, la répartition des requêtes entre ces trois tâches est respectivement de 50 %, 20 % et 30 %, alors qu'elle serait de 60 %, 14 % et 25 % pour Daniel E. Rose et Danny Levinson (2004).

Analyse par différenciation du contexte

La dernière catégorie de travaux s'intéresse à la différenciation des contextes de recherche. Mais la notion de contexte y est souvent polysémique. Elle peut correspondre à une distinction entre les lieux d'activité de la recherche d'informations, les objectifs et motivations de celle-ci ou les catégories sociodémographiques considérées. Ainsi certaines études s'appuient-elles sur la dichotomie contexte privé vs contexte professionnel pour l'analyse des moteurs de recherche. Ainsi Soo Young Rieh (2004) montre-t-il qu'il existe des différences entre l'usage du web au domicile et au travail, quand bien même les tactiques et stratégies de recherche des utilisateurs semblent-elles identiques. D'autres études consacrées à l'analyse des moteurs de recherche utilisés dans des intranets d'entreprises ont confirmé celles portant sur les moteurs de recherche généralistes. Pendant trois ans, Dick Stenmark et Taline Jadaan (2006) ont analysé l'usage de moteur intranet de *SwedCorp* en Suède ; ils ont constaté que le nombre moyen de termes utilisés est inférieur à 1,5 et qu'il est resté stable entre 2000 et 2004. Ils mettent également en évidence que près de 67 % des requêtes ne contiennent qu'un seul terme et que le nombre moyen de pages consultées est inférieur à 2.

Dépassant cette opposition privé/professionnel, de récents travaux en communication (Godard, 2007) s'intéressant aux interactions entre le monde du travail et la sphère privée, montrent une hybridation des pratiques et l'existence d'un *continuum* entre les usages privé et professionnel des technologies de l'information, et donc des moteurs. On peut supposer que, parmi les facteurs qui sous-tendent cette hybridation, se trouve la question du rapport au temps et à l'espace de travail. Cette question est notamment mise en évidence dans le cadre du télétravail et, plus spécifiquement, dans la gestion des supports de stockage des résultats de la recherche d'informations (les outils de stockage mobiles comme la clé USB, les espaces de stockage en ligne, la gestion des favoris, l'usage mixte privé/professionnel de la messagerie électronique ou la création de bibliothèques numériques personnelles).

De même, de nombreuses études s'intéressent aux objectifs et motivations de la recherche d'information. Si, avant la généralisation de l'usage des moteurs de recherche sur l'internet, les travaux portant sur les systèmes d'accès à l'information traditionnels concernaient exclusivement l'information à usage professionnel (utile à l'individu au travail) au détriment de celle à usage privé, il n'en est plus de même aujourd'hui. L'analyse de la recherche d'information à usage privé est devenue un courant autonome (*every day life information seeking*). De plus, l'analyse des comportements des usagers des moteurs en ligne suggère une

hybridation des pratiques ; dans une même session, un usager peut rechercher des informations de nature différente, professionnelle et privée.

Enfin, il existe un troisième groupe d'études qui explicitent le contexte en prenant en compte les facteurs sociodémographiques tels que l'âge, le genre, le capital symbolique, la richesse du réseau social, l'autonomie des utilisateurs... Depuis quelques années, de plus en plus de recherches s'intéressent à la question du genre et de son influence dans la recherche d'informations. Ainsi, pour Manon Arcand et Jacques Nantel (2005), il semblerait que les hommes consultent plus de pages, alors que les femmes prennent plus de temps dans la lecture des contenus des sites. Pour Marguerite Roy et Michelene T. H. Chi (2003), les hommes auraient tendance à passer plus de temps à saisir les requêtes et consulter les pages de résultats alors les femmes privilégieraient la consultation directe des sites indiqués dans les pages de résultats. De leur côté, Andrew Large *et al.* (2002) observent des différences significatives dans le comportement des filles et des garçons. Ils montrent que, par minute, les garçons utilisent moins de mots dans les requêtes, passent moins de temps sur chaque page, et cliquent sur plus de liens que les filles. À l'inverse, d'autres études, comme celles de Melius Weideman et Corrie Strümpfer (2004) ou d'Amit Bhatnagar et Sanjoy Ghose (2004), montrent que le genre n'est pas un facteur de différenciation, sauf pour des tâches de recherche transactionnelle, par exemple, quand il s'agit d'une recherche d'information pour l'achat en ligne (Ha, Stoel, 2004).

Un autre facteur étudié est l'influence de l'âge. Si plusieurs évaluations ont été réalisées sur les enfants, souvent en milieu scolaire ou étudiant, très peu ont porté sur les autres âges de la vie et, en particulier, sur les personnes âgées. De ce point de vue, Dania Bilal (2004) a montré que, auprès des élèves, la navigation donne de meilleurs résultats que la recherche par mot clé. L'expérience de navigation du web aide ces utilisateurs à identifier et localiser plus de sites pertinents et plus rapidement. En revanche, elle ne les aide pas à trouver les informations précises à l'intérieur même des sites.

Concernant les personnes âgées, plusieurs études ont mis en évidence qu'une majorité de sites web et moteurs de recherche sont difficiles à utiliser par des personnes âgées. Aideen J. Stronge, Wendy A. Rogers et Arthur D. Fisk (2006) montrent que, intrinsèquement, cette inefficacité s'explique plus par le choix de stratégies de recherche souvent inefficaces que par l'âge. Mais, Madjid Ihadjadene (2005) montre qu'avec un niveau égal d'expérience d'usage

de l'internet, les seniors atteignent autant d'objectifs de recherche que les étudiants, et les plus expérimentés y parviennent dans des délais comparables à ceux de leurs cadets. Ces résultats semblent confirmer la mise en place de mécanismes de compensation et rejoignent les travaux réalisés sur le maniement de logiciels.

On le voit, ces exemples attestent d'une très grande hétérogénéité des approches. Elles sont conduites par des communautés scientifiques différentes (sociologie des usages, économie et gestion, sciences cognitives, sciences de l'information...), avec des objectifs eux-mêmes divers (scientifiques ou commerciaux).

Conclusion

En dépit des avancées significatives concernant les pratiques des moteurs de recherche dont témoignent ces études, nous constatons des problèmes de différentes natures : terminologique, méthodologique, d'interprétation. Sur le plan terminologique, la définition des principaux concepts opératoires est souvent absente, peu explicite ou divergente. C'est en particulier le cas des notions clés comme le « besoin informationnel », la « pertinence », la « compétence », ou l'« expertise » de l'utilisateur, le « contexte », la « navigation », la « lecture » des résultats... De même, il n'existe pas de définition stable et partagée de ce que sont un « internaute », une « session » de recherche ou une « tâche de recherche ».

Sur le plan méthodologique, les résultats sont difficilement comparables, notamment du fait de la variabilité des contextes et publics analysés. Comme le montrent Kirstie Hawkey et Melanie Kellar (2004) – dans une méta-analyse portant sur 31 études qui correspondent aux deux dernières catégories –, de nombreuses variables sont peu ou mal renseignées. Ainsi seulement une étude sur deux précise-t-elle le genre de l'échantillon de la population considérée ou l'expérience et la compétence des usagers. La date de l'expérimentation n'est indiquée que dans 25 % des études, et la signification statistique des données n'est précisée que dans 1/3 des cas. Plus globalement, on peut s'interroger sur la manière d'interpréter et de comparer les résultats des études et, ensuite, sur la possibilité de généraliser les observations. Se pose également la question de la pertinence des modèles théoriques visant à modéliser le comportement informationnel des usagers et des tâches qu'ils accomplissent. En effet, il est souvent difficile de faire un lien opérationnel entre le cadre et les finalités méthodologiques des études et les cadres théoriques portés par les modèles de l'usage qui sont proposés, tels

que le modèle cognitif de Peter Ingwersen, le modèle stratifié de Tefko Saracevic, le modèle généralisé de T. Wilson ou la théorie *Sense Making* de Brenda Dervin (Chaudiron, Ihadjadene, 2002).

La prise en compte des différents facteurs de l'usage nécessite une analyse multidimensionnelle, c'est-à-dire une approche que l'on peut qualifier d'holistique (Chaudiron, Ihadjadene, 2004). Celle-ci impose de prendre en compte d'autres dimensions : affective (confiance, anxiété, satisfaction..., générées au cours du processus de recherche), technique, ergonomique, économique, et, en particulier, socioculturelle (capital social des individus, partage de croyances sur la qualité des sources et des moteurs, respect des habitus propres aux diverses communautés...).

Le nombre important et la diversité des études sur les moteurs de recherche renseignent tout d'abord sur le foisonnement de l'activité de recherche d'informations sur l'internet qui est devenu un phénomène socio-économique, politique et culturel de grande ampleur. Alors que, pendant des décennies, les travaux portaient sur les usages professionnels des systèmes d'accès à l'information, désormais, l'appropriation de ces outils par le « grand public » transforme ce dernier en objet de recherche.

Cette généralisation de l'usage des moteurs de recherche en ligne induit un transfert d'habitudes et de pratiques vers l'usage des catalogues de bibliothèques en ligne et/ou des bibliothèques numériques. Conditionnés par leurs pratiques sur l'internet, les usagers s'attendent à accéder à l'information en texte intégral, non plus seulement à des notices bibliographiques (Yu, Young, 2004). Par ailleurs, on ne peut limiter l'accès à l'information aux seuls moteurs de recherche, tant sont divers les sources et moyens d'information ainsi que leurs complémentarités (navigation, sérendipité, outils de diffusion, fils RSS, alertes par courriel, réseaux sociaux, etc.).

Aussi les études d'usage des moteurs de recherche se heurtent-elles à la complexité de l'ensemble des facteurs à prendre en compte. Un moteur de recherche n'est plus simplement le « lieu » où s'apparient différentes structures cognitives dans le cadre d'interactions, mais il est considéré comme un système plus global dans lequel entrent en jeu de multiples variables : l'espace cognitif des acteurs, les caractéristiques contextuelles psychologiques, sociales et organisationnelles, ainsi que le changement des besoins d'information. Il est important

d'appréhender l'utilisateur en situation de recherche d'information de manière beaucoup plus globale que dans les modèles cognitifs et, *a fortiori*, dans l'approche système qui sous-tend encore souvent les études d'usage actuelles des moteurs de recherche.

Références

- Arcand M., Nantel., 2005 « Gender Differences in Processing Information: Implications for Online Search Patterns and Task Performance », *Innovation and Knowledge Economy : Issues, Applications, Case studies. Part 2*, pp. 1334-1341, in : Cunningham P., Cunningham M., eds., IOS Press, <http://www.chairerbc.com/chairerbc/fr/doc.asp?id=189> (consulté le 01/07/07).
- Bar-Ilan J., Gutman T., 2005, « How do search engines respond to some non-English queries? », *Journal of Information Science*, 31(1), pp. 13-28.
- Beauvisage T., Beaudouin T., Assadi H, 2007, « Internet 1.0: early users, early uses », *Annales des Télécommunications*, 62(3-4), pp. 283-304.
- Beitzel S. M., 2006, *On Understanding and Classifying Web Queries*, PhD Thesis in Computer Science, Illinois Institute of Technology, May, http://www.webir.org/resources/phd/Beitzel_2006.pdf (consulté le 15/07/07).
- Bhatnagar A., Ghose S., 2004, « Online information search termination patterns across product categories and consumer demographics », *Journal of Retailing*, 80, pp. 221-228.
- Bilal D., 2004, « Research on children's information seeking on the Web », *Youth Information-Seeking: Theories, Models, and Approaches*, pp. 271-291, in : Kay Chelton M., Cool C., eds., Lanham, MD : Scarecrow Press.
- Broder A., 2002, « A Taxonomy of Web Search », *SIGIR Forum*, 36(2), pp. 3-10.
- Chaudiron S., Ihadjadene M., 2002, « Quelle place pour l'utilisateur dans l'évaluation des SRI ? », pp. 211-213, in : *Recherches récentes en sciences de l'information : Convergences et dynamiques, actes du colloque MICS-LERASS*, 21-22 mars, Toulouse, Paris, ADBS Éd.
- Chaudiron S., Ihadjadene M., 2004, « Évaluer les systèmes de recherche d'information : nouveaux modèles de l'utilisateur », *Hermès*, 39, pp. 170-178.
- Choo C., Detlor B., Turnbull D., 2000, « Information Seeking and the Web : An Integrated Model of Browsing and Searching », *First Monday*, 5(2), http://www.firstmonday.dk/issues/issue5_2/choo/ (consulté le 10/11/06).
- Ciaccia A., Ihadjadene M., Martins D., 2006, « Évaluation des moteurs de recherche cartographiques », *Document numérique*, 9(2), pp. 111-125.

- De Certeau M., 1990, *L'invention du quotidien, I/ Arts de faire*, Gallimard.
- Fallows D., 2005, *Search Engine Users*, Rapport du *Pew Internet & American Life Project*, http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf
- Fortunato S., Flammini A., Menczer F., Vespignani A., 2006, « Topical interests and the mitigation of search engine bias », *Proceedings of the National Academy of Sciences of the United States of America* , 103(34), pp. 12684-12689.
- Godard F., 2007, « Vie publique et vie privée : de nouveaux régimes temporels », *Réseaux*, 140(1), pp. 29-65.
- Ha Y., Stoel L., 2004, « Internet apparel shopping behaviours: the influence of general innovativeness », *International Journal of Retail and Distribution Management*, 32(8), pp. 377-385.
- Hawkey K., Kellar M., 2004, *Recommendations for reporting context in studies of web browsing behaviour*, Tech Report N° CS-2004-16, Faculty of Computer Science, Dalhousie University, Halifax, Canada, <http://www.cs.dal.ca/research/techreports/2004/CS-2004-16.shtml> (consulté le 10/07/07).
- Hussherr F., Neron S., 2002, *Comportement de l'internaute*, Paris, Dunod.
- Ihadjadene M., Chaudiron S., Martins D., 2003, « The effect of individual differences on searching the Web », *Proceeding ASIST 2003 Annual Meeting*, pp. 240-246.
- Ihadjadene M., Le Rouzo M.-L., Martins D., 2005, « La recherche d'information chez les seniors : analyse exploratoire », *Colloque Enjeux et usages des TIC : aspects sociaux et culturels*, université Michel de Montaigne, Bordeaux, 22-23-24 sept.
- Jansen B.J., Spink A., 2006, « How are we searching the World Wide Web ? A comparison between nine search engine transaction logs », *Information Processing & Management*, 42(1), pp. 248-263.
- Jouet J., 2003, « La pêche aux internautes : les études d'audience d'Internet », *Hermès*, 37, pp. 203-211.
- Koshman S., Spink A., Jansen B. J., 2006, « Web Searching on the Vivisimo Search Engine », *Journal of the American Society for Information Science and Technology*, 57(14), pp. 1875-1887.
- Large A., Beheshti J., Rahman T., 2002, « Gender differences in collaborative Web searching behavior: an elementary school study », *Information Processing and Management*, 38(3), pp. 427-443.

- Lazonder A., Biemans H., Wopereis G., 2000, « Differences between novice and experienced users in searching in the World Wide Web », *Journal of the American Society for Information Science*, 51(6), pp. 576-581.
- Marchionini G., 1995, *Information seeking in electronic environments*, New York, University Press.
- Markey K., 2007, « Twenty-five years of end-user searching, Part 1 : Research findings », *Journal of the American Society for Information Science and Technology*, 58(8), pp. 1071-1081.
- Moukdad H., 2004, « Lost in Cyberspace: How do search engines handle Arabic queries? », *Access to Information : Technologies, Skills, and Socio-Political Context, Proceedings of the 32nd Annual Conference of the Canadian Association for Information Science*, Winnipeg, June, http://www.cais-acsi.ca/proceedings/2004/moukdad_2004.pdf (consulté le 03/07/07).
- Origgi G., 2003, « Recherches sur Internet », *La Rivista dei Libri*, trad. de l'italien par A.-M. Varigault, <http://gloriaoriggi.free.fr/Internetrecherche.htm> (consulté le 27 mars 2006).
- Palmquist R. A., Kim K., 2000, « The Effect of Cognitive Style and Online Search Experience on Web Search Performance », *Journal of the American Society of Information Science*, 51(6), pp. 558-567.
- Pan B., Hembrooke H., Joachims T., Gay G., Granka L., 2007, « In Google we trust : Users' decisions on rank, position and relevancy », *Journal of Computer-Mediated Communication, special issues on The Social, Political, Economic and Cultural Dimensions of Search Engines*, 12(3), <http://www.ota.cofc.edu/pan/new-JCMC-paper.pdf> (consulté le 27/07/07).
- Park S., Lee J. H., Bae H. J., 2005, « End user searching: A Web log analysis of NAVER, a Korean Web search engine », *Library and Information Science Research*, 27(2), pp. 223-231.
- Ravid G., Bar-Ilan J., Baruchson-Arbib S., Rafaeli S., 2007, « Popularity and findability through log analysis of search terms and queries : The case of a multilingual public service Web site », *Journal of Information Science*, 33(5), pp. 567-583.
- Rieh H. Y., Rieh S. Y., 2005, « Web searching across languages: Preference and behavior of bilingual academic users in Korea », *Library & Information Science Research*, 27(2), pp. 249-263.

- Rieh S. Y., 2004, « On the web at home: Information seeking and web searching in the home environment », *Journal of the American Society for Information Science and Technology*, 55 (8), pp.743-753.
- Rivadeneira W., Bederson B. B., 2003, « A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable User Interfaces », University of Maryland HCIL, Technical Report HCIL-2003-36.
- Rose D. E., Levinson D., 2004, « Understanding User Goals in Web Search », pp. 13-19, in : *Proceedings of the World Wide Web Conference (WWW 2004)*, New York, NY, USA.
- Roy M., Chi M.-T. H., 2003, « Gender differences in patterns of searching the Web », *Journal of Educational Computing Research*, 29(3), pp. 335-348.
- Silverstein C, 1998, « Analysis of a very large Web search engine query log », *Technical note #1998-014*, Digital SRC, Oct.
- Spink A., Jansen B. J., 2004, *Web Search: Public Searching of the Web*, Springer-Verlag, Berlin.
- Stenmark D., Jadaan T., 2006, « Intranet Users'Information-Seeking Behaviour : A Longitudinal Study of Search Engine Logs », *Proceedings of ASIS&T 2006*, Austin, Texas, Nov.
- Stronge A. A., Rogers W. A., Fisk A. D., 2006, « Web based Information Search and Retrieval : Effects of Strategy Use and Age on Search Success », *Human Factors and Ergonomics Society*, 48(3), pp. 434- 446.
- Tabatabai D., Shore B., 2005, « How experts and novices search the Web », *Library & information science research*, 27, 2, pp. 222-248.
- Tricot A., Drot-Delange B., Foucault B., El Boussarghini R., 2000, « Quels savoir-faire les utilisateurs réguliers du Web acquièrent-ils ? », *Journal d'Intelligence Artificielle*, 14 (1-2), pp. 93-112.
- Tricot C., Roche C., Foveau C. E., Reguigui S., 2006, « Cartographie sémantique de fonds numériques scientifiques et techniques », *Document numérique*, 9(2), pp.12-35
- Weideman M., Strumpfster C., 2004, « The effect of search engine keyword choice and demographic features on internet searching success », *Information Technology and Libraries*, 23(2), pp. 58-65.
- Yu H., Young M, 2004, « The Impact of Web Search Engines on Subject Searching in OPAC », *Information Technology and Libraries*, 23(4), pp. 168-180.