

Les médiations dans l'accès aux documents en ligne : pouvoir et autorité des moteurs commerciaux

Brigitte Simonnot

► **To cite this version:**

Brigitte Simonnot. Les médiations dans l'accès aux documents en ligne : pouvoir et autorité des moteurs commerciaux. Actes de la troisième conférence Document numérique et Société " Documents, contenus numérique : politique en question ", Sciences Po Aix-en-Provence, Nov 2010, France. pp.175-191. sic_00667498

HAL Id: sic_00667498

https://archivesic.ccsd.cnrs.fr/sic_00667498

Submitted on 7 Feb 2012

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.

Les médiations dans l'accès aux documents en ligne :
pouvoir et autorité des moteurs commerciaux¹

Brigitte Simonnot

simonnot@univ-metz.fr

Le développement des publications en ligne et l'émergence de nouveaux acteurs puissants dans ce domaine ont renouvelé les approches scientifiques de la recherche d'information et, plus largement, de l'accès aux documents en ligne. L'un des objectifs affichés par les concepteurs de dispositifs d'accès à l'information est d'en faciliter l'accès à un plus grand nombre d'individus, et se traduit souvent par la tentative de mettre de l'ordre dans les documents. Parmi ces dispositifs, les moteurs commerciaux sont encore peu analysés en termes des présupposés dont ils procèdent (Dervin, 1994) et de la politique qu'ils mettent en œuvre pour arriver à leurs objectifs (Introna & Nissenbaum, 2000).

À partir d'une sélection de publications récentes sur les moteurs de recherche, nous proposons d'examiner l'évolution de la politique des moteurs de recherche et les questions de pouvoir qu'elle soulève. Notre analyse s'attache aussi à dégager certaines des médiations qui sont à l'œuvre à l'heure actuelle avec la montée en puissance de ces infomédiaires particuliers que sont les moteurs de recherche commerciaux. Ce réexamen souhaite participer à une réflexion plus large sur les alternatives à mettre en place pour permettre l'accès à l'information en ligne et redistribuer le pouvoir dans ce domaine.

1. Pourquoi parler de politique pour les moteurs de recherche ?

Les professionnels de l'information et des bibliothèques connaissent la notion de politique documentaire. Les bibliothèques ont toujours procédé à une sélection des documents qu'elles mettent à disposition en fonction des besoins du public, en prenant en compte un certain nombre de principes « intellectuels » et déontologiques – autorité, pluralité – en plus de leurs contraintes budgétaires. La question des besoins des usagers est souvent placée en première ligne dans l'élaboration de ces politiques. Mais on mobilise peu le concept de politique lorsqu'on évoque les nouveaux entrants que sont les moteurs commerciaux, dans la mesure où ils semblent vouloir collecter tout ce qui est à leur portée sans autre projet que d'agrandir

¹ Texte paru dans E. Broudoux & G. Chartron (dirs) *Enjeux politiques du document numérique. Actes de la troisième conférence Document numérique et Société « Documents, contenus numérique : politique en question »*, (pp. 175-191) Sciences Po Aix-en-Provence, 15-16 novembre.

leurs index et d'augmenter leur audience. La politique d'un moteur peut se lire non seulement à partir de la manière dont il collecte les documents, dont il les sélectionne mais aussi à partir des traitements qu'il opère et des modalités de leur mise en accès.

Régimes de publication en ligne et collecte par les moteurs

Sur le web, la publication est soumise au régime de l'*opt out*, au sens où, par défaut, tout document publié est susceptible d'être utilisé par les moteurs. Le protocole d'exclusion des robots de moteurs de recherche (Koster, 1994) permet cependant aux concepteurs de sites d'indiquer qu'ils ne souhaitent pas que certaines actions soient réalisées par certains moteurs sur leurs documents : pas de collecte, pas de suivi des liens, pas d'archivage en cache sur leurs serveurs. D'abord conçu pour des raisons purement techniques, ce protocole permet aussi de traiter certaines questions « morales » (Thelwall & Stuart, 2006). Il a récemment été mis en avant dans la négociation entre moteurs et éditeurs de presse suite à des conflits juridiques², montrant que les éditeurs traditionnels n'avaient pas forcément connaissance de ce dispositif technique ou, du moins, de la manière dont ils pouvaient le mobiliser. Conçu pour résoudre essentiellement des problèmes techniques, il est bien sûr insuffisant pour traiter les questions légales voire syndicales relatives aux droits d'auteur et de diffusion des contenus en ligne. Soulignons aussi que la généralisation du régime de l'*open access* est une condition nécessaire au développement de l'activité des moteurs, ce qui peut expliquer la manière dont certains s'engagent pour ce principe, tout en jouant un rôle ambivalent du fait des enjeux économiques sous-jacents (Ippolita, 2008).

Loin de se contenter des documents librement disponibles sous forme numérique, la compagnie Google a entrepris en 2004 une vaste opération de numérisation d'ouvrages et mis en place un moteur de spécialité, Google Books (Google Livres pour la version française). Sans revenir sur les conflits avec les éditeurs et les débats juridiques occasionnés par les modalités de mise en œuvre de cette initiative, force est de constater que la puissance économique de la société et la forme d'expertise technologique que ce projet lui a permis de développer³ en font un interlocuteur incontournable dans les projets de numérisation du

² Par exemple, dans le litige qui a opposé Google et la société Copiepresse à propos de l'exploitation par Google News des articles de presse des titres adhérents à la société belge, une partie de la négociation s'est résolue par l'utilisation de la balise noarchive afin que les articles ne soient pas mémorisés en cache par le moteur. Cette information a été confirmée sur le site de la société <http://copiepresse.be/>

³ Il s'agit de mettre au point des méthodes rapides et non destructives des œuvres, de manière à pouvoir numériser les ouvrages rares ou anciens. <http://books.google.fr/intl/fr/googlebooks/history.html>

patrimoine à large échelle. En 2009, Microsoft, Yahoo et Amazon se sont joints à l'initiative de l'Internet Archives de fonder l'Open Book Alliance⁴, travaillant à contrer les tentations monopolistiques du géant californien. Au-delà des profits prévisibles liés par exemple à l'exploitation de versions numériques de livres épuisés, ces initiatives attestent du fait que les moteurs ne sont plus dans un processus de simple collecte mais dans des actions concrètes visant à augmenter encore les contenus auxquels ils permettent d'accéder. La gratuité partielle de l'accès peut servir, dans ce cas, à mettre en place des objectifs beaucoup plus rentables à long terme.

Pouvoir de sélection des moteurs

La collecte automatique des documents est la pierre angulaire du succès des moteurs. Jusqu'en 2005, les principaux majors commerciaux du domaine ont d'ailleurs rivalisé par divers moyens de communication sur la taille de leurs index respectifs : affichage de la taille de l'index sur la page d'accueil de Google, communiqués de presse des moteurs de Microsoft ou de Yahoo faisant état du nombre de pages indexées. Notons au passage que, non seulement les chercheurs indépendants étaient dans l'impossibilité de vérifier ces informations – ils pouvaient tout juste chercher à les estimer – mais en outre que nombre de travaux sur l'estimation de la taille du web se sont appuyés sur les index des moteurs commerciaux, en prenant pour seule précaution d'en inclure plusieurs dans leurs études. C'est dire l'importance stratégique prise par ces nouveaux entrants dans le paysage du web.

Cependant, la politique d'acquisition des moteurs ne saurait être réduite à la collecte automatique puisqu'ils opèrent une sélection dans les documents qu'ils indexent, et celle-ci n'est pas seulement liée à leurs limites techniques. D'une part, ils procèdent à un dédoublement pour éliminer de leurs index les pages similaires. D'autre part, ils se réservent le droit d'exclure de leurs index certaines pages, voire certains sites pour divers motifs. Le premier type de motifs est lié à des décisions judiciaires auxquelles les moteurs doivent se plier. Les guides proposés aux webmestres⁵ en exposent d'autres, de manière plus ou moins claire ou complète. Certains cas ont attiré l'attention sur ces procédés, comme ceux de BMW ou de RICOH en Allemagne dont les sites ont été exclus puis réintégrés dans l'index de

⁴ <http://www.openbookalliance.org/>

⁵ Pour Google, voir <http://www.google.com/support/webmasters/>. Pour Yahoo Search, voir Yahoo! Search Content Quality Guidelines, <http://help.yahoo.com/l/us/yahoo/search/basics/basics-18.html>.

Google en 2006. L'impact médiatique de ces cas a créé l'émotion dans les milieux spécialisés du référencement. Les professionnels de l'information n'ont pas pour mission d'évaluer la qualité des documents auxquels ils donnent accès, ils se fient pour cela aux spécialistes du domaine. L'évaluation à laquelle ils procèdent pour les acquisitions porte davantage sur l'adéquation des documents au public de leur bibliothèque ou de leur centre de ressources. « Le développement d'une collection est une question de prédiction et non d'évaluation » écrivait Patrick Wilson (1996 : 128). Ces prédictions sont néanmoins périodiquement réévaluées, notamment à l'occasion des opérations de désherbage des fonds qui figurent dans les missions des professionnels et dont les critères devraient être inscrits dans toute politique documentaire. Dans le cas des moteurs, leurs procédés de « désherbage » qui supprime des sites de leurs index nécessiteraient des clarifications. Quelques voix, peu nombreuses, s'élèvent d'ailleurs pour les demander au nom du droit à l'audience (Chandler, 2007).

Face à l'abondance de données qu'ils indexent, le pouvoir de sélection des moteurs s'exerce au moins autant au moment de l'affichage des résultats que de la collecte des documents. Soo Young Rieh (2002) a montré comment les individus, lors de leurs recherches d'information en ligne, procèdent à une évaluation en deux temps : des jugements prédictifs à partir de la description sommaire qui est faite des documents qui leur sont proposés, qui leur permettent de décider de consulter ou non le document en détail, et des jugements évaluatifs après lecture complète de ces documents. Les études menées à partir des historiques d'interrogation des moteurs ont montré que les internautes ne consultent en général que les tout premiers résultats qui leur sont proposés (voir par exemple Jansen & Spink, 2006). Les choix opérés par les moteurs pour présenter leurs résultats ont donc une influence indéniable sur l'information que les internautes consultent réellement. C'est pourquoi la question des « biais » dans le calcul et le classement des résultats est particulièrement sensible, en l'absence d'information complète sur la manière dont ces classements sont opérés. Certains sites sont ainsi surreprésentés dans les résultats naturels – ainsi qualifiés par opposition aux liens commerciaux qui relèvent de la publicité payante. C'est le cas notamment des articles de l'encyclopédie collaborative Wikipedia ou de sites commerciaux dont le référencement a été « optimisé » (Höchstötter & Lewandowski, 2009).

Les compagnies majors qui produisent des moteurs généralistes ont aussi développé des moteurs de spécialités (actualités, images, vidéos, publications scientifiques, recherches locales, par exemple), constituant ainsi des collections distinctes sur des critères de formats ou de provenance des documents. La « recherche universelle », c'est-à-dire le fait de proposer

dans les résultats des moteurs généralistes ceux issus des moteurs de spécialité, contribue certainement à aider les internautes à trouver une information : cela les dispense de décider à quelle application spécialisée adresser leur requête et permet de diversifier la nature des résultats proposés. Par exemple, une recherche sur le moteur généraliste Google.fr à propos d'une publication scientifique propose dans les résultats un lien qui mène à une requête à Google Scholar, (publications académiques), des liens plus généraux et un lien vers les contenus numérisés de Google Books. La recherche universelle permet d'accéder à différentes « collections » constituées par le moteur à travers une interface unique. Stratégiquement, cela permet aussi de mettre en valeur les moteurs de spécialité, souvent moins connus du grand public. Une vision particulière, que l'on pourrait qualifier d'ergonomique, de la pluralité nécessaire à une information de qualité mais qui trouve ses limites dans la clôture informationnelle qu'elle opère autour des différents champs d'activité d'un moteur. Enfin, les efforts récents qui permettent aux moteurs de produire des chronologies à partir de techniques de résumés automatiques montrent aussi une tendance à renvoyer à leurs autres services et les interconnexions qui sont ainsi rendues possibles. Les moteurs se posent en référents dans le monde du web, dont ils seraient les seuls à pouvoir maîtriser le chaos.

2. Représentation de l'utilisateur dans les moteurs

La réification des pratiques informationnelles

Toute politique documentaire repose sur l'analyse des besoins des usagers. C'est aussi l'objectif principal affiché des systèmes de recherche d'information que de répondre aux besoins informationnels des individus (Baeza-Yates & Ribeiro-Neto, 1999). Dans ces applications, l'utilisateur a longtemps été représenté par la requête ponctuelle qu'il soumet au système, supposée décrire son besoin d'information. À présent, les traces laissées sur les serveurs par les internautes peuvent aussi être collectées, conservées et exploitées de manière plus profonde. Ces traces informent les concepteurs sur les sujets privilégiés dans les recherches mais aussi sur la manière souvent très intuitive dont les utilisateurs expriment leurs besoins. Les premiers enseignements tirés de l'étude des historiques d'interrogation des moteurs portent sur la pauvreté de la syntaxe des requêtes : peu de mots, opérateurs ignorés. Les résultats d'études menées sur les moteurs différents concordent sur ce point (pour une revue, voir Ihadjadene & Chaudiron, 2008). L'ergonomie tente de pallier ces déficiences, effaçant derrière des interfaces simples et soignées les langages d'interrogation et la nécessaire interprétation faites par les moteurs de formules lacunaires. C'est l'amélioration

des dispositifs qui permet aux moteurs de justifier la collecte des traces sur les pratiques dont ils font l'objet et, bien que certains s'en émeuvent, les moteurs ne semblent pas prêts à renoncer à ces procédés. Il faut dire que la concurrence à laquelle ils doivent faire face et qui propose des alternatives aux internautes, comme le moteur Cuil par exemple, ne semble pas vraiment représenter une menace pour leurs intérêts économiques tant ils sont installés, à l'heure actuelle dans le paysage de la recherche en ligne.

Bien que les algorithmes de traitement de l'information aient beaucoup progressé, ils souffrent encore d'insuffisances : difficile en effet de totalement automatiser des traitements censés résoudre des problèmes mal structurés voire mal définis, comme le sont la plupart des problèmes de recherche d'information. Pour y remédier, les moteurs procèdent de plus en plus à une hybridation entre traitement automatique des documents et implication des individus. On l'observe déjà, par exemple, dans la prise en compte du nombre de liens entrants vers les documents pour classer les résultats (Brin & Page, 1998) ou, plus récemment, dans la mise en place d'applications ludiques pour permettre aux internautes de participer à l'indexation des images par mots-clés, comme dans l'application mise en place par Google, Image Labeller, inspirée du prototype ESP (Von Ahn, 2005). On pourrait considérer qu'il s'agit là de mettre en œuvre des méthodes participatives, où les compétences des internautes seraient mises à contribution pour améliorer le fonctionnement des moteurs. Mais ces contributions sont rarement conscientes et volontaires, la plupart des internautes ignorant la manière dont leurs actions sont exploitées. Leurs pratiques se trouvent réifiées et mises au service d'algorithmes de fouille de données qui cherchent à en dégager des tendances et à nourrir le marché de la publicité commerciale.

Répondre aux besoins informationnels ou susciter le désir ?

Certaines fonctionnalités récentes implémentées dans les moteurs commerciaux portent à remettre en cause la primauté de leur fonction de répondre à des besoins d'information. Par exemple, les fonctions de suggestion lors de la frappe, si elles sont souvent comprises comme des aides orthographiques, montrent aussi la volonté de devancer les besoins des individus, quitte à modifier radicalement leurs demandes. Les logos de fêtes de Google suggèrent également des requêtes toutes faites à l'occasion de certains événements. L'interface de Yahoo propose des articles d'actualités et les tendances du jour. Les suggestions de reformulations de requêtes tendent à prescrire l'exploration de parties des collections selon des critères opaques. Les moteurs ne se contentent plus de répondre aux requêtes ponctuelles des individus, ils cherchent à susciter le désir d'information : si un besoin d'information peut

être assouvi par une réponse ponctuelle, le désir est par essence insatiable. Il est donc essentiel pour les moteurs commerciaux non seulement de fidéliser les internautes mais aussi de stimuler leurs désirs d'information pour s'assurer une audience toujours plus large. Les internautes sont alors moins considérés comme des usagers que comme un public ou des consommateurs par les concepteurs de ces applications.

Prise en compte de l'avis des usagers

Une politique documentaire est censée prendre en compte, périodiquement, l'avis des usagers pour être révisée et adaptée. Les moteurs semblent prendre en compte, dans une certaine limite, les critiques dont ils font l'objet, dont certaines sont éminemment politiques. Le phénomène des bombes Google (Bar-Ilan, 2007), médiatisé régulièrement par la presse, a donné lieu à un grand nombre de protestations de la part de certaines communautés ciblées. D'abord minimisé par les concepteurs, le phénomène a été pris suffisamment au sérieux pour obliger à produire un communiqué réaffirmant la neutralité des opérateurs⁶. Les moteurs commerciaux affirment désormais lutter contre ces phénomènes qui entachent leur réputation.

Le feuilleton relatif à l'exploitation du moteur Google en Chine est certainement très représentatif du caractère sensible que représentent ces services. Si tous les moteurs commerciaux se sont en effet pliés à la censure imposée par le gouvernement chinois sur les contenus en ligne, c'est principalement la position de Google qui a été le plus discutée et la plus controversée. Le fait que Google se soit posé en champion de la démocratie dans les explications données sur son algorithme Pagerank y est sans doute pour quelque chose. Accepter la censure de ses résultats invalide le principe de neutralité si souvent réaffirmé par la société ou du moins le relativise.

La prise en compte de l'avis des usagers consiste, dans le cas des moteurs, non seulement à améliorer leurs services mais à maintenir une forme de crédibilité et à asseoir une réputation acquise à coup de défis et d'innovations technologiques. Les opérateurs savent, en effet, combien ils sont dépendants de leur audience en ligne pour maintenir leurs profits.

⁶ "The beliefs and preferences of those who work at Google, as well as the opinions of the general public, do not determine or impact our search results" Google, s.d., *An explanation of our search results*, <http://www.google.com/explanation.html> (consulté le 17/07/2010)

3. Médiations informationnelles

Le rôle des médiateurs

Dans les années 1990, les professionnels de l'information en contact direct avec leurs usagers se sont revendiqués couramment comme des médiateurs. L'étymologie du terme, le verbe latin *mediare*, évoque l'idée de se situer à l'exact milieu entre deux pôles, comme tous les points d'une médiatrice se trouvent à égale distance des deux extrémités d'un segment de droite. Dans le domaine social, politique ou législatif, le médiateur est celui qui s'interpose entre deux entités pour permettre le traitement des conflits à travers les échanges entre les deux parties et s'assurer qu'elles se comprennent. Des analyses sociologiques empiriques et situées ont proposé de distinguer plusieurs types de médiation : les « médiations de conciliation » et les « médiations de conformation », où la médiation intervient comme moyen de « traitement para-institutionnel du désordre » (voir, dans le cadre d'une entreprise de transport, Macé, 1997). Il s'agirait de bénéficier des compétences de certains individus et de leurs connaissances du contexte pour leur faire jouer le rôle de médiateurs et trouver des arrangements précaires qui maintiennent un ordre établi. Les médiateurs n'ont pas de pouvoir sanction, ils participent à trouver des solutions à travers les négociations qu'ils rendent possibles entre les protagonistes. Dans le cas des moteurs, nous avons vu qu'ils s'arrogeaient un pouvoir de sanction en excluant certains sites de leurs index par exemple, ou en déclassant des documents à l'occasion de changements dans leurs algorithmes. Cela remet en cause la neutralité qu'ils affichent : en s'arrogeant de tels pouvoirs, ils prennent délibérément position contre certaines pratiques de publication, officiellement celles liées au référencement abusif. Mais en l'absence d'informations complètes sur les critères qu'ils appliquent, on peut s'interroger sur le fait qu'il s'agisse bien là d'une volonté de préserver la qualité de leurs résultats naturels. Ne s'agit-il pas plutôt d'obliger les concepteurs de pages à se plier à leurs recommandations ?

Dans le domaine culturel, le rôle de médiateurs est moins de résoudre des conflits que de permettre la rencontre d'individus avec celle-ci, d'être des « passeurs » de culture (Briant & Palau, 1999 : 49). Il s'agit de donner à un plus grand nombre de personnes l'occasion d'accéder aux œuvres culturelles et de servir d'intermédiaires entre un patrimoine et le public, en œuvrant souvent au sein d'une institution. Antoine Hennion (1993) a mis en évidence la tentation sociologiste de faire des médiateurs les acteurs d'un refoulement social que les sciences humaines seraient là pour dénoncer. En prenant l'exemple de la passion musicale, il montre que la médiation se fait dans la performance, le jeu de l'œuvre. Elle fait intervenir

musiciens et instruments (instruments de musique ou lecteurs d'enregistrements), partitions et scènes.

Les moteurs comme infomédiaires

Pouvons-nous transposer ces réflexions dans le domaine de l'information ? L'information n'est pas non plus un donné en soi. Comme la musique, elle est portée par des objets informationnels, des documents aux flux sonores, qui nécessitent d'être interprétés. Pour accéder à l'information, il ne suffit pas de localiser des documents, encore faut-il les lire (ou les écouter). Les actions des professionnels qui permettent au public d'y accéder sont diverses : mise en forme des textes, descriptions, organisation des collections, résumés, critiques, etc. Attachés à préserver le contexte des documents, ils participent à produire des éléments qui seront pris en compte par les usagers, notamment dans leurs jugements préliminaires. Cependant, contrairement au cas des musiciens, les actions de ces professionnels sont le plus souvent invisibles aux yeux du grand public et des profanes. Si l'œil exercé du documentaliste ou du chercheur peut remettre en cause le traitement documentaire réalisé pour tel document, ce sera rarement le cas pour le « lecteur ordinaire », de plus en plus pressé et ignorant de ce qu'il aurait pu trouver à la place de ce qu'on lui propose. Dans le cas des publications en ligne, le volume d'information potentielle est tel qu'il rend dérisoire toute initiative visant à évaluer l'exhaustivité d'une collection et par là même la qualité de l'application qui permet d'y accéder. Si les moteurs sont de nouveaux infomédiaires (Ihadjadene & Chaudiron, 2002 ; Rebillard & Smyrnaio, 2010), leur médiation s'exerce entre les concepteurs de contenus en ligne et les internautes désireux d'y accéder. Elle tend à les imposer comme acteurs incontournables dans ce domaine. Moins que de permettre les échanges entre les deux parties, ils suscitent la rivalité entre les concepteurs de pages. L'ordre qu'ils cherchent à mettre en place concerne surtout les pratiques d'écriture, rôle qui était jusque-là dévolu aux éditeurs, et les pratiques de référencement via leurs services.

Autorité et légitimité *versus* réputation

L'ordre documentaire traditionnel repose en grande partie sur les notions de légitimité et d'autorité. Patrick Wilson (1983 ; 1996) a développé la notion d'autorité cognitive comme l'un des mécanismes qui nous permet de déterminer l'importance d'un texte et décider de le prendre en compte. Il discerne l'autorité de l'auteur et l'autorité de l'éditeur de celle des textes eux-mêmes, mais souligne leur importance et leur participation à construire l'autorité cognitive d'un texte. Lorsque nous cherchons à nous informer, la prise en compte de l'autorité

des textes est déterminante pour nous permettre de distinguer, au sein de la masse publiée sur le sujet, ce qui vaut la peine d'être lu.

Evelyne Broudoux (2007) a proposé de substituer à la notion d'autorité cognitive celle d'autorité informationnelle. L'autorité informationnelle peut être, selon elle, définie en référence aux formes d'attribution d'autorité qui peuvent concerner la figure de l'auteur (autorité énonciative), celle de l'institution (autorité institutionnelle), le contenu du document (son genre, sa qualité, la nature des sources qu'il convoque) et la nature du support (support physique et périodicité de publication). On pourrait y ajouter l'autorité des intermédiaires qui promeuvent les publications et les distinguent de la masse par le traitement qu'ils en font. De même que les éditeurs ont, par leur travail de sélection et de mise en valeur des oeuvres, été gratifiés d'une certaine autorité dans le monde de l'imprimé, de même aujourd'hui on peut observer un déplacement des attributs d'autorité perçue vers les intermédiaires qui donnent accès à l'information en ligne. Les enquêtes menées auprès des utilisateurs des moteurs de recherche montrent en effet, à travers les témoignages recueillis, que le choix de recourir à tel ou tel moteur repose sur le fait qu'il a su s'imposer dans le paysage de l'information en ligne. Ces perceptions des moteurs reposent sur plusieurs critères : leur réputation, leur disponibilité permanente (notamment par leur intégration à l'interface des navigateurs) mais aussi les expériences antérieures des internautes, le fait qu'ils ont été ou non satisfaits de leurs précédentes démarches vis-à-vis de ces services (McKenzie, 2003 ; Simonnot, 2009 ; Hilligoss & Rieh, 2008). Les moteurs tendent ainsi à devenir le point de départ obligé de toute recherche en ligne, lorsqu'ils n'en sont pas l'outil exclusif (Brophy & Bawden, 2005). On pourrait dire que, dans le monde des publications en ligne, la notion d'autorité tend à être remplacée de fait par celle de réputation. L'autorité des publications, comme la réputation, est socialement construite au sein de communautés agissantes mais les deux notions ne procèdent pas des mêmes critères. Les critères de surface priment dans l'attribution de réputation. Quant à la notion de légitimité, elle est supplantée par les performances techniques qui posent les concepteurs de moteurs en experts dans leur domaine du traitement de l'information en ligne.

Les moteurs créent une familiarité qui amène à ne pas s'interroger sur leur fonctionnement. Ils créent des habitudes. Les moteurs habituent les internautes à accéder très rapidement aux informations, pratiquement à la seconde, et à toujours recevoir une réponse fût-elle de mauvaise qualité. Ils les amènent également à multiplier leurs pratiques informationnelles en ligne, en se posant comme des intermédiaires incontournables.

Neutralité *versus* pluralité

Les moteurs de recherche sont souvent présentés comme de nouveaux infomédiaires qui sélectionnent les documents en ligne, en réalisent un traitement (indexation, résumé) afin de permettre à leurs utilisateurs d'y accéder. La médiation qu'ils opèrent semble de nature essentiellement technologique dans la mesure où toutes leurs opérations sont automatisées. À chaque occasion où ils doivent faire face à la critique, ils affichent leur neutralité. Ils font oublier par là leur pouvoir de sélection et le traitement qu'ils opèrent sur les données qu'ils récupèrent et auxquelles ils donnent accès. Par la présentation des résultats qu'ils composent, les moteurs modifient la première perception que les internautes peuvent avoir des documents proposés. Leur classement est souvent considéré comme un jugement de pertinence intrinsèque des documents. Lorsque l'on connaît l'importance de ces jugements préliminaires (Rieh, 2002), où l'internaute décide ou non d'aller voir le document en détail, on appréhende l'impact que cette présentation peut avoir sur les pratiques informationnelles. La neutralité de point de vue, principe fondateur d'une encyclopédie collaborative comme Wikipedia, est en effet une utopie dans le domaine de la recherche d'information, notamment pour les moteurs qui présentent leurs résultats sous forme de listes que l'espace limité de l'écran hiérarchise de fait (Höchstötter & Lewandowski, 2009). Le pouvoir des moteurs est immense : celui de rendre visibles certaines pages au détriment d'autres et de favoriser certaines pratiques. Quant à la volonté de proposer une pluralité de contenus, si l'on tient compte de l'espace limité d'affichage des résultats et de ce qui est réellement vu par les internautes, elle porte davantage sur la pluralité des formats que sur celle des points de vue. C'est qu'il manque encore aux moteurs la capacité à donner du sens aux informations.

Conclusion : moteurs et pouvoir

Dans un article présentant l'algorithme Pagerank (Brin & Page, 1998), les fondateurs de Google Inc. utilisaient l'analogie avec la démocratie pour expliquer son fonctionnement. Douze ans plus tard, il est facile de percevoir l'illusion de cette analogie. La prise en compte des actions des utilisateurs par les moteurs et de la participation des internautes se fait généralement à leur insu. Ils sont placés dans l'obligation de participer, ce qui contribue à les déresponsabiliser dans leurs pratiques informationnelles.

Il y a presque 10 ans, Dominique Boullier (2001) écrivait : « les moteurs de recherche d'internet n'ont, pour l'instant, aucun pouvoir de structuration des contenus d'origine : de ce fait, leurs indexations demeurent sommaires et leurs traitements des requêtes à base purement

statistique ». Dix ans plus tard, une tel énoncé peut être remis en cause. Les moteurs, par leurs guides à l'intention des webmestres et les pratiques de référencement qu'ils induisent, incitent en effet à modifier la présentation des pages, voire leur contenu. Par les opérations de numérisation auxquels ils participent, ils constituent des fonds selon des choix qui leur sont propres.

L'histoire récente montre que le pouvoir acquis par les moteurs n'est pas seulement celui de rendre visibles certaines pages au détriment d'autres. Ils ont permis aux recherches d'information en ligne de prendre une ampleur qui était encore insoupçonnable il n'y a pas si longtemps. Ils participent aussi à formater ces pratiques et à les promouvoir. Ce ne sont pas tant les pratiques individuelles qui intéressent les concepteurs que les tendances qui se dessinent et les « arts de faire », voire les détournements qui sont autant de suggestions pour mettre en place de nouvelles fonctionnalités. Les services documentaires traditionnels ont été longtemps conçus pour des usagers experts, maîtrisant des langages documentaires complexes, ce qui a contribué à écarter les moins habiles. Récemment, des efforts importants ont été mis en œuvre pour en faciliter l'utilisation. Les moteurs ont quant à eux très vite travaillé, au contraire, à faciliter leur utilisation, conscients qu'il s'agissait d'une condition nécessaire pour généraliser les pratiques dont ils font l'objet. Beaucoup déplorent l'emprise de ces services en ligne au détriment des services traditionnels des bibliothèques, voire de la presse. Faute d'autres repères, les moteurs sont considérés par les internautes comme des « mécanismes collectifs de désignation de valeur » (Hennion, 1993), un point d'entrée unique qui permet de s'orienter sans avoir à choisir entre plusieurs applications. Cette situation leur procure un avantage décisif à l'heure des indicateurs dérisoires dérivés d'une société du spectacle marchandisé, qui a inventé l'audience comme mesure de valeur.

L'introduction de nouvelles technologies dans un secteur amène en général à réévaluer et à renégocier l'ordre établi (Strauss, 1992 : 250). Dans le cas des moteurs de recherche commerciaux, on peut observer l'ampleur des contestations et des négociations qui touchent à la fois aux questions de droit et aux conditions d'accès et d'usage de l'information en ligne. Sans nier les progrès technologiques qu'ils ont permis, et sans se contenter de la dénonciation récurrente de leurs tentations monopolistiques, il est temps de prendre conscience des changements qu'ils ont opérés dans les habitudes et de les analyser pour proposer des alternatives, en supposant que l'avènement de nouveaux entrants dans ce secteur soit encore possible. Et c'est bien une volonté politique qui est nécessaire pour rassembler les savoir-faire

et les moyens indispensables pour imaginer de nouveaux services capables de dépasser les lois du marché.

Références

- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. New York: ACM Press, Addison-Wesley .
- Bar-Ilan, J. (2007). Google Bombing from a Time Perspective. *Journal of Computer-Mediated Communication* , 12 (3), p. article 8.
- Birnholtz, J. P. (2006). What Does It Mean to Be an Author? The Intersection of Credit, Contribution, and Collaboration in Science. *Journal of the American Society for Information Science and Technology* , 57 (13), 1758–1770.
- Boullier, D. (2001). Les conventions pour une appropriation durable des TIC: Utiliser un ordinateur et conduire une voiture. *Sociologie du travail* , 43, 369–387.
- Briant, V. d., & Palau, Y. (1999). *La médiation. Définition, pratiques et perspectives*. Paris: Nathan.
- Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks* , 30, 107-117.
- Brophy, J., & Bawden, D. (2005). Is Google enough? Comparison of an internet search engine with academic library resources. *Aslib Proceeding: New Information Perspectives* , 57 (6), pp. 498-512.
- Broudoux, E. (2007). Construction de l'autorité informationnelle sur le web. Dans R. Skare, N. W. Lund, & A. Vårheim, *A Document (Re)turn: Contributions from a Research Field in Transition*. Frankfurt: Peter Lang.
- Chandler, J. A. (2007). A Right to Reach an Audience: An Approach to Intermediary Bias On The Internet. *Hofstra Law Review* , 35 (3), 1095-1137
http://law.hofstra.edu/pdf/Academics/Journals/LawReview/lrv_issues_v35n03_i06.pdf.
- Choo, C. W., Detlor, B., & Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday* , 5 (2), URL:
http://firstmonday.org/issues/issue5_2/choo/index.html.
- Dervin, B. (1994). Information ↔ Democracy: An Examination of Underlying Assumptions. *Journal of the American Society for Information Science* , 45 (6), 369-385.
- Hennion, A. (1993). L'histoire de l'art : leçons sur la médiation. *Réseaux* , 11 (60), 9-38.
- Hilligoss, B., & Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing and Management* , 44, 1467-1484.

- Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences* , 179 (12), 1796-1812.
- Ihadjadene, M., & Chaudiron, S. (2002). La recherche et la diffusion d'information sur Internet : vers de nouvelles médiations ? *Actes du XIIIe Congrès de la SFSIC Les recherches en information et communication et leurs perspectives - Histoire, objet, pouvoir, méthode : Vers une éthiquedecommunicationnelle* (pp. 163-173). Marseille: SFSIC.
- Ihadjadene, M., & Chaudiron, S. (2008). L'étude des dispositifs d'accès à l'information électronique : approches croisées. Dans F. Papy, *Problématiques émergentes dans les sciences de l'information* (pp. 183-207). Paris: Hermès Lavoisier.
- Ihadjadene, M., & Chaudiron, S. (2008). Quelles analyses de l'usage des moteurs de recherche ? Questions méthodologiques. *Questions de Communication* (14), 17-32.
- Introna, L., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The Information Society* , 16 (3), 169-185.
- Ippolita. (2008). *La face cachée de Google*. (M. Rovere, Trad.) Paris: Payot & Rivages.
- Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web ? A comparison of nine search engine transaction logs. *Information Processing and Management* , 42, pp. 248-263.
- Koster, M. (1994). *A standard for robot exclusion*. The Web Robots Page, <http://www.robotstxt.org> .
- Macé, E. (1997). Les contours de la médiation : institution, conciliation, conformation. A propos d'un dispositif de "médiation" de la RATP. *Revue Française des Affaires Sociales* (2), 225-244.
- McKenzie, P. J. (2003). Justifying cognitive authority decisions: Discursive strategies of information seekers. *The Library Quarterly* , 73 (3), 261-288.
- Rebillard, F., & Smyrnaiois, N. (2010). Les infomédiaires, au coeur de la filière de l'information en ligne : les cas de Google, Wikio et Paperblog. *Réseaux* (160-161), 164-194.
- Rieh, S. Y. (2002). Judgement of Information Quality and Cognitive Authority in the Web. *Journal of the American Society for Information Science and Technology* , 53 (2), pp. 145-161.
- Simonnot, B. (2009). De l'usage des moteurs de recherche chez les étudiants. Dans B. Simonnot, & G. Gallezot, *L'entonnoir. Google sous la loupe des sciences de l'information et de la communication* (pp. 31-57). Caen: C & F éditions.
- Strauss, A. (1992). *La trame de la négociation : sociologie qualitative et interactionnisme*. Paris : L'harmattan.

Thelwall, M., & Stuart, D. (2006). Web Crawling Ethics Revisited: Cost, Privacy, and Denial of Service. *Journal of the American Society for Information Science and Technology* , 57 (13), 1771–1779.

Von Ahn, L. (2005). *Human Computation*. PhD thesis in computer science: Université Carnegie Mellon, CMU-CS-05-193.

Wilson, P. (1996). Information Retrieval and Cognitive Authority. Dans R. L. Ruggles, *Knowledge management tools* (pp. 121-144). Newton, MA, USA: Butterworth-Heinemann .

Wilson, P. (1983). *Second-hand knowledge:an inquiry into cognitive authority*. Westport CT: Greenwood Press.