

DIRK LEWANDOWSKI

Hochschule für Angewandte Wissenschaften Hamburg, Deutschland
Fakultät DMI, Department Information
dirk.lewandowski@haw-hamburg.de

NADINE HÖCHSTÖTTER

Topicflux, Deutschland
nsh@topicflux.de

MESURER LA QUALITÉ DES MOTEURS DE RECHERCHE WEB

Résumé. — La qualité des moteurs de recherche est un élément fondamental de l'étude de l'accès à l'information et de certains aspects de la politique de communication. La question de savoir si un moteur européen unique est nécessaire et, si tel était le cas, ce qui différencierait celui-ci de ses concurrents américains, ne peut trouver de réponse qu'à travers l'analyse des moteurs de recherche existants. Notre cadre théorique d'analyse de la qualité d'un moteur de recherche comporte quatre domaines : la qualité de l'index, l'efficacité de la recherche d'information, la qualité des fonctionnalités de recherche et l'utilisabilité du moteur en question. Cette contribution présente une revue de la littérature scientifique et les résultats empiriques des études que nous avons réalisées. Les conclusions montrent, notamment, que de nouveaux paramètres réellement centrés sur l'utilisateur seraient nécessaires à l'évaluation de la qualité des moteurs de recherche.

Mots clés. — Moteur de recherche, évaluation de la qualité, comportement de recherche, qualité de l'index, efficacité de la recherche d'information.

encore trouvé de réponses réellement satisfaisantes. Si l'on observe la concentration du marché des moteurs de recherche où il reste à peine quatre principaux concurrents - à savoir Google, Yahoo, MSN et Ask -, il est important de porter un regard extérieur sur ces questions. En gardant cela à l'esprit, les chercheurs ont le devoir d'examiner le niveau de performance de ces moteurs de recherche et de déterminer les mesures qui devraient être utilisées pour en évaluer la qualité afin de donner aux utilisateurs des standards objectifs de qualité.

Nous savons que les critères traditionnels d'évaluation utilisés en recherche d'information ne permettent pas de mesurer la qualité d'un moteur de recherche dans son intégralité. Mais la mise au point de mesures spécifiques, comme celles proposés par exemple par Liwen Vaughan (2004) ne suffisent pas non plus. La qualité d'un moteur doit trouver une définition plus large et intégrer des facteurs allant bien au-delà des performances de recherche d'information. Nous sommes convaincus que seule une approche intégrée de l'évaluation de la qualité peut conduire à des résultats susceptibles d'être utilisés pour le développement de meilleurs moteurs de recherche. A l'instar de la recherche d'information, de manière générale, nous observons un changement de paradigme d'une perspective plus technique (centrée sur le document) vers une perspective centrée sur l'utilisateur (Ingwersen, Järvelin, 2005).

Dans un premier temps, nous reverrons les études réalisées jusqu'ici sur le comportement des utilisateurs des moteurs de recherche et en proposerons une synthèse. Dans la section suivante, nous présenterons un cadre général d'évaluation de la qualité des moteurs de recherche. Nous analysons ensuite l'évaluation de la qualité selon ces quatre dimensions : la qualité de l'index, l'efficacité de la recherche d'information, la qualité des fonctionnalités de recherche et l'utilisabilité du moteur de recherche. Dans la dernière partie, nous synthétiserons les résultats et proposerons des stratégies susceptibles d'améliorer les moteurs de recherche.

Etat de l'art des observations du comportement de l'utilisateur

Il existe deux approches principales de l'observation du comportement de recherche de l'utilisateur : les méthodes supervisées et celles qui ne le sont pas. Les études et enquêtes menées en laboratoire sont dites supervisées lorsque le sujet sait qu'il est observé par un chercheur Les

cette combinaison est rarement observée comme approche principale. Nous résumerons brièvement ces deux méthodes d'observation du comportement de l'utilisateur pour en déduire des paramètres d'évaluation de la qualité.

Dans certaines enquêtes, on demande parfois directement aux utilisateurs comment ils décriraient leur comportement de recherche. Elles donnent aussi une vision subjective, du point de vue de l'utilisateur; sur ce qu'ils savent des moteurs de recherche et de leurs services complémentaires. Des utilisateurs ayant des niveaux de connaissance différents montrent des comportements de recherche différents (Schmidt-Maenz, Bomhardt, 2005). Dans la plupart des cas, les études en laboratoire portent seulement sur de petits échantillons et, de ce fait, ne sont pas représentatives. Il arrive également que les sujets, se sentant observés, essaient d'adopter une méthode de recherche plus professionnelle, par exemple en utilisant un plus grand nombre d'opérateurs ou de fonctionnalités de recherche. Une façon plus représentative de collecter des informations sur le comportement réel d'un utilisateur passe par l'analyse des historiques de transactions ou les données saisies en direct. Dans ce cas, le problème est qu'on ne dispose pas d'informations complémentaires sur l'utilisateur lui-même, telles ses caractéristiques démographiques et ses habitudes en général.

L'étude de Marcel Machill, Christoph Neuberger; Wolfgang Schweiger et Weiner Wirth (2003,2004) est constituée de deux parties : une enquête téléphonique auprès de 1 000 participants et une analyse en laboratoire sur 150 sujets. Cette étude montre que 14 % des utilisateurs de moteurs de recherche utilisent des fonctionnalités de recherche avancées. Seulement 33 % des personnes interrogées savent qu'il est possible de personnaliser les interfaces d'un moteur de recherche. Le titre et la description des pages dans les résultats trouvés par le moteur de recherche sont très importants pour évaluer les listes de résultats. Pour 44 % d'entre eux, les utilisateurs sont mécontents lorsque les résultats n'ont rien à voir avec la requête soumise. Par ailleurs, 36 % d'entre eux rejettent ce que l'on appelle les liens morts. Marcel Machill et *al.* (2003) concluent leur étude en soulignant que les utilisateurs attendent une présentation rapide des résultats, des pages de résultats claires et la satisfaction de leurs besoins d'information. Christoph Hoelscher et Gerhard Strube (2000) ont observé des comportements de recherche différents selon que l'utilisateur est expert ou novice. Gord Hotchkiss (2000) distingue différents groupes de comportements de recherche quant à l'examen des pages de résultats. Par ailleurs, les utilisateurs privilégient les résultats naturels par rapport aux liens sponsorisés.

Spink, Jansen, 2004 ; Spink, Jansen, Ozmutlu, 2000 ; Spink, Ozmutlu, Ozmutlu, Jansen, 2002 ; Spink, Wolfram, Jansen, Saracevic, 2001). D'autres ont utilisé les historiques d'Altavista (Silverstein, Henzinger, Marais, Moricz, 1999 ; Beitzel, Jensen, Chowdhury, Grossman, Frieder, 2004) et Alltheweb (Jansen, Spink, 2003 ; Jansen, Spink, 2006). Des chercheurs (Cacheda, Viña, 2001) ont également pu travailler sur l'historique du moteur de recherche espagnol BWE (*Buscador en Internet para la Web en Español*). Christoph Hoelscher et Gerhard Strube (2000) ont analysé l'historique des requêtes de Fireball, un moteur de recherche allemand. Jason Zien, Jörg Meyer, John Tomlin et Joy Liu (2000) ont observé en direct sur une période de 66 jours les requêtes saisies sur Webcrawler. On peut aussi trouver un état de l'art général et un descriptif de toutes ces données (Höchstötter; Koch, 2008). Cette étude indique aussi les paramètres les plus importants pour décrire le comportement de recherche de l'utilisateur, comme la longueur moyenne des requêtes, le pourcentage de requêtes complexes - notamment le pourcentage de recherches sur des expressions - et la proportion de sessions de recherche où seule la première page de résultat est consultée (tableau 1). Il est évident que les requêtes sont très courtes. En second lieu, une partie importante des requêtes ne contient qu'un seul terme. A quelques exceptions près, l'usage des opérateurs booléens est très rare. L'utilisation de la recherche d'expressions est l'une des façons les plus courantes d'affiner les requêtes. Les utilisateurs se contentent généralement d'examiner la première page de résultats. Ces faits démontrent que les utilisateurs des moteurs de recherche formulent leurs demandes de manière intuitive et qu'ils n'évaluent pas chaque résultat de la liste (voir également Schmidt-Maenz, Koch, 2006). Par ailleurs, les utilisateurs ne modifient pas les paramètres de recherche par défaut des moteurs, car ils n'utilisent pas de fonctionnalités de recherche particulières.

Un cadre de référence pour la qualité des moteurs de recherche

Min Xie *et al.* (1998) ont proposé un cadre de référence pour mesurer la qualité des moteurs de recherche puis celui-ci a été développé par Haixun Wang *et al.* (1999). Les auteurs se fondent sur le modèle de SERVQUAL (Service et qualité) appliqué aux moteurs de recherche (Parasuraman, Zeithaml, Berry, 1988). Ils comparent les attentes des utilisateurs aux performances perçues des moteurs de recherche, mais ne pondèrent pas les facteurs observés. Dans le cadre qu'ils proposent, il est clair qu'il

sont pertinents vis-à-vis de la requête ») ne diffère pas significativement d'un moteur à un autre.

À l'opposé de telles approches centrées sur l'utilisateur; il en est également une - dite système « classique » - qui essaye de mesurer la performance des systèmes de recherche d'informations d'un point de vue plus « objectif ». Tefko Saracevic (1995) distingue deux grandes catégories d'approches pour l'évaluation des systèmes de recherche d'informations, chacune comportant trois niveaux : les niveaux d'évaluation centrés système - niveau d'ingénierie (e.g. performances du matériel ou du logiciel), niveau d'entrée (couverture du corpus) et niveau du traitement (e.g. performance des algorithmes) ; les niveaux d'évaluation centrés utilisateur - niveau de sortie (interaction avec le système, rétroaction), niveau de l'usage et de l'utilisateur (où sont soulevées des questions d'application à des problèmes et tâches donnés) et le niveau social (qui tient compte de l'impact sur l'environnement).Tefko Saracevic (1995 : 141) conclut que les résultats d'un niveau d'évaluation ne disent rien sur la performance du même système quant aux autres niveaux d'évaluation et que « cette séparation des niveaux d'évaluation pourrait être considérée comme une insuffisance de base de toutes les évaluations de RI ». Selon nous, ce constat s'applique également à l'évaluation des moteurs de recherche. Seule une combinaison des deux approches, celle centrée sur le système et celle centrée sur l'utilisateur; peut donner une image plus précise de la qualité globale d'un moteur de recherche.

C'est pourquoi nous étendons le cadre d'analyse de qualité, appliqué une première fois par Dirk Lewandowski (2006c), aux quatre rubriques suivantes :

1. Qualité de l'index ; elle souligne l'importance des bases de données des moteurs de recherche pour obtenir des résultats pertinents et complets ; les mesures appliquées dans cette rubrique comprennent la couverture du Web, les biais liés aux pays et la mise à jour
2. Qualité des résultats ; c'est là que sont appliqués les tests dérivés de RI classique. Comme le montre de l'analyse des mesures de RI évoquée *supra*, il convient de se demander quelles mesures devraient être appliquées et si de nouvelles mesures sont nécessaires pour satisfaire le caractère spécifique des moteurs de recherche et de leurs utilisateurs. Par exemple, on devrait appliquer ici une mesure supplémentaire, à savoir la singularité des résultats par rapport aux autres moteurs de recherche.
3. Qualité des fonctionnalités de recherche ; le moteur propose un ensemble conséquent de fonctionnalités (comme la recherche avancée) qui donnent des résultats fiables.
4. Utilisabilité du moteur de recherche ; elle renseigne rétroactivement sur le comportement de l'utilisateur et est évaluée par le biais d'enquêtes auprès des utilisateurs ou d'analyses des historiques de transactions. Elle fournit des paramètres que l'on peut comparer sur la conception de l'interface - la visualisation des liens

ue eeue ruorique.

Qualité de l'index

Les moteurs de recherche sont uniques dans la manière dont ils construisent leurs bases de données. Tandis que les systèmes traditionnels de recherche d'information s'appuient habituellement sur des bases de données construites manuellement par des indexeurs humains à partir de sources sélectionnées, les moteurs de recherche doivent utiliser la structure en réseau du Web pour trouver leurs documents en parcourant le Web de lien en lien. C'est un énorme défi que de construire et de maintenir un index généré par des robots de collecte. Knut Magne Risvik et Rolf Michelsen (2002) en donnent une bonne synthèse. On peut considérer de plusieurs façons la qualité de l'index d'un moteur de recherche donné. En premier lieu, les index doivent être étendus (c'est-à-dire qu'ils doivent couvrir une grande partie du Web).

Taille des index

Un moteur de recherche idéal serait celui qui conserve une copie complète du Web dans sa base de données. Or pour divers motifs, cela est impossible. Un index important est nécessaire pour deux raisons. La première est le cas où l'utilisateur souhaite obtenir une liste complète de résultats, par exemple pour se familiariser avec un sujet. Le second cas est celui des requêtes peu claires qui ne donnent que peu de résultats. Ici, le moteur qui dispose d'un index plus grand a de meilleures chances de trouver davantage de résultats. On compare parfois les index des moteurs de recherche les uns par rapport aux autres, en s'appuyant sur leurs tailles respectives absolues. Mais certaines études qui traitent de la taille du Web s'intéressent également à la proportion qui est couverte par les moteurs de recherche.

Il existe trois façons d'obtenir des estimations chiffrées pour analyser la taille du Web et la couverture des moteurs de recherche : les chiffres déclarés par les prestataires de moteurs (pour une synthèse, voir Sullivan, 2005), les analyses de recouvrement et l'échantillonnage aléatoire. Krishna Bharat et Andrei Broder (1998) utilisent un robot de collecte pour parcourir une partie du Web et construire un vocabulaire à partir duquel ils sélectionnent des requêtes et les envoient aux quatre principaux moteurs de recherche. À partir de l'ensemble aléatoire des pages trouvées, les auteurs calculent le pourcentage de couverture de ces moteurs de recherche. L'étude la plus

que le Web indexable en 2005 contient au moins 1,5 milliards de pages. La couverture par les moteurs de recherche de l'ensemble de données (formé de toutes les pages retrouvées par au moins un des moteurs) se situe entre 57 et 76 % pour les quatre principaux moteurs de recherche (Google, Yahoo, MSN, Ask). La recherche la plus importante qui ait utilisé une technique d'échantillonnage aléatoire pour déterminer la taille totale du Web est la seconde étude de Steve Lawrence et C. Lee Giles (1999). Elle s'appuie sur un ensemble d'adresses IP générées de manière aléatoire dont ils testent la disponibilité. Les auteurs évaluent la taille du Web indexable à environ 800 millions de pages. Ils testent la couverture des moteurs de recherche grâce à 1 050 requêtes. NorthernLight, le moteur de recherche qui enregistrait les meilleures performances, couvrait uniquement 16 % du Web indexable à l'époque. Les moteurs soumis à cette évaluation couvraient, tous réunis, à peine 42 % du Web indexable.

Toutes les études portant sur la taille du Web et la couverture des moteurs de recherche que nous avons mentionnées ont ceci en commun qu'elles se concentrent sur la partie indexable du Web, appelée Web de surface. Mais il ne s'agit que d'une infime partie du Web, le reste formant ce qu'on appelle le Web invisible ou Web profond. Pour résumer; le Web invisible est la partie du Web que les moteurs de recherche n'intègrent pas à leurs index. Michael K. Bergman (2001) estime la taille du Web Invisible à 550 fois celle du Web de Surface. Toutefois, comme l'ont montré Dirk Lewandowski et Philipp Mayr (2006), ces estimations de taille sont bien trop élevées, en raison d'erreurs statistiques et fondamentales. Même si cette estimation repose sur des données de 2001, il semble que le taux de croissance propre des tailles des bases de données (Williams, 2005) n'affecte pas la taille totale de manière considérable. Il est nécessaire de développer d'autres recherches pour distinguer le Web visible et le Web invisible. Au cours des dernières années, nous avons observé la conversion de grandes bases de données en pages HTML afin de permettre leur indexation par les principaux moteurs de recherche en ligne. Bien que cette pratique soit principalement le fait d'acteurs commerciaux, d'autres fournisseurs - telles les bibliothèques - ont suivi la même approche avec des degrés de réussite divers (Lewandowski, 2006b).

Biais liés au pays

Même s'il est important pour un moteur d'être le plus complet possible, un moteur de recherche doté d'une bonne couverture générale n'est pas nécessairement ce qu'il y a de mieux pour chaque domaine du Web.

en français ou seulement très peu. Dans le processus de collecte des contenus du Web, l'index diffèrera d'une part selon les points de départ choisis et, d'autre part, selon la structure du Web. Les pages qui sont la cible de nombreux liens ont de meilleures chances d'être retrouvées par les moteurs que les pages situées en « périphérie » du Web. Andrei Broder *et al.* (2000) ont modélisé la structure du Web sous la forme d'un nœud papillon. Toutefois, les pages au centre du Web (le noyau de celles qui sont fortement connectées entre elles) sont probablement les plus anciennes et, si l'on considère la structure évolutive du Web, en provenance des États-Unis (Vaughan, Thelwall, 2004). Il est surprenant de constater qu'il existe (du moins à notre connaissance) à peine deux études portant sur les biais liés au pays. Notamment dans le contexte européen, vu les nombreuses langues qui y sont parlées, ce sujet devrait susciter un intérêt plus vif.

Lewen Vaughan et Mike Thelwall (2004) comparent la couverture des sites Web aux États-Unis, en Chine, à Singapour et Taiwan. Ils notent que les sites américains sont mieux couverts tant en proportion qu'en termes de profondeur d'indexation. Ils en concluent que le désavantage en faveur de ces sites doit provenir de la structure des liens du Web et non des différentes langues. Cette étude a été reproduite et étendue en 2005 (Vaughan et Zhang, 2007). Les résultats obtenus ne diffèrent pas vraiment de ceux de la première étude. Ces études indiquent qu'il existe des biais importants liés aux pays dans les index des moteurs de recherche. Nous pensons qu'il est important de conduire des études similaires. Si l'on revient à la question de savoir s'il faut envisager la construction d'un moteur de recherche purement européen pour concurrencer les moteurs de recherche américains dominants et à l'analyse de l'utilité de moteurs de recherche spécifiques par pays, nous pensons qu'il est urgent de mener des études sur la question des biais liés au pays, au moins pour certains pays européens.

Fraîcheur de l'index

Le troisième facteur important de la qualité des index est leur actualisation. Le Web est dans un flux constant (Ntoulas, Cho, Oison, 2004) : de nouveaux documents sont ajoutés, d'anciens disparaissent et d'autres sont mis à jour. Comme on peut le constater d'après l'étude de Nadine Schmidt-Maenz et Martina Koch (2006), la plupart des utilisateurs font des recherches sur des faits actuels et des actualités récentes. De plus, le nombre de liens entrants change de manière similaire. Les pages

D'abord, il est important de maintenir l'index pour qu'il soit récent. En second lieu, ces facteurs sont utilisés dans le classement des pages Web (Acharya et *al.*, 2005 ; Lewandowski, 2006a). Enfin, ces facteurs pourraient jouer un rôle important dans la recherche scientifique sur le Web (Lewandowski, 2004b).

Nous avons comparé l'actualité des index des principaux moteurs de recherche (Lewandowski, Wahlig, Meyer-Bautor; 2006 ; Lewandowski, 2008a). Les résultats montrent que, du point de vue de la mise à jour; le meilleur moteur de recherche change au fil des années et qu'aucun des moteurs n'a de solution idéale pour préserver la fraîcheur de son index. Les distributions de fréquence concernant l'ancienneté des pages sont faussées, ce qui signifie que les moteurs de recherche font bien la différence entre les pages selon qu'elles sont mises à jour fréquemment ou rarement. Nous avons identifié un problème majeur quant au délai entre le moment où les pages sont collectées et celui où elles sont disponibles pour une recherche en ligne, qui diffère d'un moteur à un autre.

Efficacité de la recherche

Si l'on se penche sur la qualité des résultats, il existe quantité de littérature scientifique sur l'évaluation de l'efficacité de la recherche des moteurs (Leighton, Srivastava, 1999 ; Wolff, 2000 ; Singhal, Kaszkiel, 2001 ; Ford, Miller Moss, 2002 ; Griesbaum, Rittberger Bekavac, 2002 ; Griesbaum, 2004 ; Machill et *al.*, 2004 ; Véronis, 2006). Du fait que le paysage des moteurs de recherche change en permanence, les études les plus anciennes sont surtout intéressantes pour leurs méthodes mais elles ne sont que d'une utilité limitée quant à leurs résultats. Pour les besoins de cette contribution, nous examinons les résultats généraux de la littérature scientifique sur l'efficacité de la recherche d'information ainsi que les résultats de notre propre étude (Lewandowski, 2008b), qui fait également un état de l'art approfondi.

Le nombre de requêtes utilisées dans les études varie considérablement. En particulier les études les plus anciennes (Chu, Rosenthal, 1996 ; Ding, Marchionini, 1996 ; Leighton, Srivastava, 1999) utilisent quelques requêtes seulement (5 à 15) et sont, par conséquent, d'un intérêt limité (Buckley, Voorhees, 2000). Les plus récentes utilisent au moins 25 requêtes, certaines 50, voire plus. Dans les études plus anciennes, les requêtes sont généralement inspirées de questions de référence ou des systèmes commerciaux en ligne, alors que les études récentes se concentrent

Gordon, Pathak, 1999), mais nous constatons une tendance à se focaliser sur l'internaute ordinaire dans les tests de moteurs.

En ce qui concerne le nombre de résultats pris en compte, la plupart des études ne tiennent compte que des 10 ou 20 premiers résultats. Ceci, principalement en raison de la quantité de travail que cela représente pour les évaluateurs, mais également au vu du comportement général des utilisateurs des moteurs de recherche. Ces utilisateurs voient rarement plus que la première ou la deuxième page de résultats (avec généralement 10 résultats par page). Habituellement, ce sont les moteurs de recherche les plus importants et les plus populaires qui sont testés, et ils sont parfois comparés avec des moteurs de recherche nouveaux ou spécifiques à une langue donnée (e.g. Véronis, 2006). Une question importante est de savoir comment on devrait mesurer les résultats. La plupart des études utilisent des échelles de pertinence (de trois à six points). Les analyses de Joachim Griesbaum (2004 ; Griesbaum et *al.*, 2002) s'appuient sur des jugements de pertinence binaires avec une exception : les résultats peuvent aussi être jugés comme « renvoyant vers un document pertinent » (c'est-à-dire que la page elle-même n'est pas pertinente mais contient un hyperlien vers une page pertinente). Tandis que la plupart des études récentes assurent l'anonymat des résultats (le moteur de recherche n'est pas connu), nous avons trouvé seulement deux études qui rendent les listes de résultats aléatoires (Gordon, Pathak, 1999 ; Véronis, 2006). Les études récentes montrent en général que, dans l'ensemble, les résultats des moteurs de recherche ne sont pas très bons, mais ils fournissent quelques bons résultats sur la première page de résultats. Il existe des différences de performance entre les moteurs mais - au moins pour les moteurs de recherche les plus performants Google et Yahoo - les résultats ne diffèrent pas au point de justifier l'énorme préférence des utilisateurs pour Google,

Spécificité des résultats de la recherche

En ce qui concerne la spécificité des moteurs de recherche, nous devons faire la distinction entre le caractère unique des bases de données (défini par leur recouvrement) et le caractère unique des résultats de la recherche (jusqu'à un certain seuil). En effet, deux moteurs de recherche exploitant un index strictement identique peuvent donner les résultats dans un ordre totalement différent en fonction de leurs algorithmes de classement. Des études analysant le recouvrement des résultats de recherche de différents moteurs ont été réalisées à large échelle (e.g. Schwartz, 1998 ; Chignell,

couverts par cette étude sont Ask Jeeves, Google, MSN et Yahoo. Pour chaque moteur de recherche, les 10 premiers résultats ont été pris en compte. L'étude repose sur deux ensembles de requêtes, soumises à partir d'avril 2005 (10 316 requêtes) et en juillet 2006 (12 570 requêtes). Pour chacune, les 10 premiers résultats du moteur ont été téléchargés. Le recoupement est fait de manière automatique par comparaison directe des URL. Cette approche est quelque peu problématique car des contenus identiques peuvent correspondre à des URL différents et les moteurs de recherche ne conservent qu'un URL pour les contenus dupliqués (Bharat, Broder 1998).

Amanda Spink *et al.* (2006) constatent qu'un seul moteur de recherche liste 84,9 % de l'ensemble des résultats, deux moteurs en listent 14 %, trois 2,6 % et tous les moteurs étudiés n'en listent que 1,1%. Les auteurs concluent qu'« en utilisant un seul moteur de recherche Web pour une requête, l'utilisateur se prive de découvrir une gamme de sites Web bien classés qui apparaissent sur la première page de résultats pour n'importe quelle requête » (*ibid.* : 1385).

Qualité des fonctionnalités de recherche

Les comparaisons des fonctionnalités et des commandes des moteurs de recherche ne manquent pas (Ojala, 2002 ; Hock, 2004 ; Notess, 2006). Dirk Lewandowski (2004a) analyse les fonctionnalités de recherche que devraient proposer les moteurs et le degré auquel les principaux moteurs de recherche les appliquent. Malheureusement, ces comparaisons des moteurs sont irrémédiablement dépassées. Nous renvoyons le lecteur ici à la compilation de Greg R. Notess (2007). Dans une comparaison des moteurs de recherche et bases de données en ligne, Roslina Othman et Nor Sahlawaty Halim (2004) montrent à quel point les fonctionnalités de recherche des moteurs sont généralement limitées. Le problème est que la fiabilité opérationnelle des fonctionnalités de recherche est souvent surestimée. Alors que certaines fonctions ne présentent manifestement aucun problème potentiel (comme la restriction au domaine de premier niveau), d'autres fonctions pourtant relativement simples à appliquer ne fonctionnent pas convenablement sur certains des principaux moteurs, par exemple l'opérateur booléen OR sur Google (voir Notess, 2000). Le niveau de performance d'autres fonctions, comme la restriction à un certain langage, la recherche de pages apparentées, les filtres sur le contenu et la restriction par date, n'est pas clair

exploitons les résultats de l'étude que nous avons menée en ligne en 2003 (Schmidt-Maenz, Bomhardt, 2005) ainsi que d'autres études sur les utilisateurs des moteurs de recherche. En outre, nous avons observé les requêtes saisies en temps réel sur trois moteurs différents - Fireball (FB), Lycos (L) et Metaspinner (MS) - depuis l'été 2004 (Schmidt-Maenz, Koch, 2005, 2006). Dans les requêtes en temps réel que nous avons observées, la liste a pu être automatiquement mise à jour à l'aide d'un programme de rafraîchissement des pages. Grâce à ce dispositif, nous avons collecté une liste pratiquement complète des requêtes effectuées sur ces moteurs au cours de notre période d'observation. Certains résultats sont listés dans le tableau I.

Tableau I : Résultats empiriques de l'observation de quatre saisies de requêtes différentes.

ID	Année	jours	Nb de requêtes ;	Longueur moyenne	Requêtes à un seul terme	Requêtes complexes	Recherche d'expression exacte	Fonctionnalité de recherche
FB	2004	399	132 833 007	1,8	50,1 %	<3,0%	21 %	65,8%
L	2004	403	189 930 859	1,7	51,9%	<3,0%	2,4%	-
MS	2004	314	4 089 731	1,8	48,4%	<3,0%	2,5%	87,9%

Conception de l'interface

Les interfaces des moteurs de recherche ont une seule dimension mais il existe différents groupes d'utilisateurs de moteurs de recherche avec des besoins différents (Hotchkiss, Garrison.Jensen, 2004). La plupart des internautes qui effectuent des recherches d'information évaluent la liste des résultats très rapidement avant de sélectionner une ou deux pages Web recommandées (Hotchkiss *et al.*, 2004 ; Spink, Jansen, 2004). Sur Google, la fenêtre de saisie de requête est très claire alors que celle de Yahoo est surchargé de messages publicitaires et d'actualités (Geoghegan, 2004). Les liens provenant de la publicité payante ne sont généralement pas clairement séparés des résultats naturels. Ces liens sont signalés par des couleurs de fond très pâles (e.g. sur Google) ou par une indication textuelle en très petits caractères de couleur claire (e.g. Altavista). C'est pourquoi, souvent, les utilisateurs ne peuvent différencier les deux ou ont le sentiment que le lien sur lequel ils ont cliqué pourrait être un lien publicitaire payant. De plus, il est important de ne présenter que quelques résultats (10 à 15) car les utilisateurs des moteurs n'aiment pas faire défiler la liste (Hotchkiss *et al.*, 2004). En outre, les moteurs de recherche doivent fournir des fonctionnalités pour aider les utilisateurs à affiner leurs requêtes. Seuls les utilisateurs aguerris emploient les opérateurs

Adoption des fonctionnalités et des opérateurs de recherche

Les requêtes des internautes sont très courtes et ce paramètre ne montre aucune variation dans le temps. Les requêtes en allemand sont en moyenne légèrement plus courtes que les requêtes en anglais, car l'allemand utilise des mots composés plutôt que des chaînes de mots. Près de la moitié des requêtes ne contiennent qu'un seul terme. En ce qui concerne les termes employés pratiquement quotidiennement (Schmidt-Maenz, Koch, 2006), on constate l'utilisation inadéquate de nombreux opérateurs et des prépositions comme « dans » ou « pour ». Cela montre combien les internautes formulent leurs requêtes en ligne de manière intuitive. L'usage d'opérateurs représente moins de 3 % de toutes les requêtes observées. La recherche sur des expressions exactes est la forme la plus fréquente de formulation complexe de requêtes. Dans cette étude, les requêtes sur des expressions représentaient 2,1 % pour Fireball, 2,4 % pour Lycos et 2,5 % pour Metaspinner (Schmidt-Maenz, Koch, 2005). Dans leurs enregistrements d'expressions de requêtes en temps réel (*live ticker*), les moteurs de recherche allemands comme Fireball et Metaspinner montrent aussi la région de recherche sélectionnée. Seule la recherche de pages en allemand est fréquemment sélectionnée. Ceci tient au fait que la région géographique est un paramètre par défaut des deux moteurs de recherche. Dans plus de deux tiers de l'ensemble des requêtes, les utilisateurs n'emploient pas ces fonctionnalités pour personnaliser leur recherche. Les personnes interrogées dans le cadre de notre enquête internet ont également reconnu qu'elles ne personnalisent pas leurs moteurs de recherche favoris en fonction de leurs besoins. En fin de compte, cela signifie que les fonctionnalités telles que les opérateurs de recherche ne sont pas adoptées par les internautes.

Performance des moteurs de recherche

Le facteur le plus critiqué dans les listes de résultats des moteurs de recherche est l'optimisation des pages Web uniquement dans le but de les faire apparaître en tête de classement, ce qui présente peu d'intérêt pour l'utilisateur. C'est également le cas pour les pages qui ne correspondent pas à la requête (24,4 % de 2014) et les pages de publicité (21,4 % de 2014). Nous pensons que, souvent, les internautes ne savent pas s'ils cliquent sur des résultats dont le référencement est payant ou naturel.

interrogées (76,6 % de 6133) ne pensent pas aux possibilités de personnalisation de leurs moteurs préférés. Ces résultats montrent qu'il est possible d'évaluer l'utilisabilité du moteur de recherche par le biais d'enquêtes auprès des utilisateurs. Les personnes interrogées déclarent également trouver ce qu'elles recherchent. Toutefois la qualité des résultats obtenus n'est pas claire. 70,8 % des 6722 répondants retournent très souvent sur le moteur; immédiatement, dès qu'ils ne trouvent pas ce qu'ils souhaitent sur une page Web recommandée.

Aide à l'utilisateur

Généralement, les internautes ne savent pas comment fonctionnent les moteurs de recherche. Nous leur avons posé cinq questions générales sur les moteurs, telles que « L'affirmation suivante est-elle exacte ? Les métamoteurs possèdent leur propre index ». Seulement 44,2 % sur 5944 personnes interrogées ont pu répondre correctement à quatre ou cinq de ces questions. Nous avons également constaté que les utilisateurs qui ont donné le plus de réponses exactes utilisent beaucoup plus les opérateurs (Schmidt-Maenz, Bomhardt, 2005). Ces résultats permettent de montrer qu'en général les utilisateurs ne comprennent pas les moteurs de recherche. Compte tenu de cela, il est important que l'interface d'un moteur de recherche soit claire et simple pour en améliorer l'utilisabilité.

La plupart des moteurs de recherche proposent des fonctions d'aide mais c'est toujours via un bouton minuscule (Google, Yahoo). Jens Fauldrath et Ame Kunisch (2005) ont constaté que seulement 57 % des moteurs de recherche qu'ils ont analysés proposent une page d'aide qui soit facile à trouver. En outre, il manque une description générale de ce que font réellement les moteurs de recherche. Seulement 71 % des principaux moteurs de recherche fournissent une aide sur la manière de procéder à une session de recherche. Un autre point pour améliorer les conseils aux utilisateurs serait de fournir des informations supplémentaires sur les pages classées. Sur ce point, le titre des documents, une brève description et l'URL sont utiles. Tous les moteurs de recherche fournissent cette information. Il est pourtant tout aussi intéressant pour l'utilisateur de voir quand les dernières modifications ont été apportées à la page recommandée, ou de se voir suggérer des termes de recherche similaires. 71 % des principaux moteurs de recherche donnent des informations temporelles et seulement 29 % suggèrent des termes similaires (Fauldrath, Kunisch, 2005).

, »^j^VJI VJ I M , ^UIH idIL IUI !! Iai ILLJ I———>I" >

moteurs de recherche. Toutefois, grâce aux mesures que nous proposons, il est possible de stimuler le débat sur ce qui est réellement important pour l'évaluation de la qualité d'un moteur de recherche compte tenu de la concentration du marché dans ce domaine. Cela permettrait aussi d'en améliorer la transparence. Nous avons montré qu'il existe manifestement un écart entre la performance des moteurs de recherche et les besoins de l'utilisateur du point de vue des fonctionnalités proposées. Si l'on tient compte du comportement de recherche de l'utilisateur plusieurs aspects pourraient être améliorés. Notre hypothèse est que les utilisateurs ne savent pas comment interagir au mieux avec les moteurs de recherche. C'est pourquoi des fonctions d'aide doivent être proposées de manière à ce que les utilisateurs qui recherchent de manière intuitive apprennent aussi à utiliser les moteurs de recherche internet.

Les paramètres d'évaluation d'un moteur de recherche que nous proposons aideront à mener des études de qualité pour comparer différents moteurs de recherche selon les mêmes mesures. De même, cela aidera les utilisateurs à décider quels moteurs de recherche ils préféreront utiliser. La plupart des recherches scientifiques utilisent des paramètres très spécifiques pour mesurer la qualité d'un moteur de recherche et le comportement de l'utilisateur est souvent totalement occulté. Dans notre cas, nous avons introduit une approche globale pour mesurer la qualité d'un moteur de recherche, prenant en ligne de compte à la fois le plan technique et le point de vue de l'utilisateur. Le fait de changer de moteur de recherche permet d'obtenir des résultats différents. C'est une bonne option lorsqu'on ne trouve pas ce que l'on cherche. De plus, la question de la différence des résultats est extrêmement importante dans le débat sur les problèmes de monopole (ou d'oligopole) du marché des moteurs de recherche.

Traduction : Brigitte Simonnot
(université Paul Verlaine-Metz, Centre de recherche sur les médiations)

Références

- Acharya A., Cutts M., Dean J., Haahr R., Henzinger M., Hoelzle U. *et al*, 2005, « Information retrieval based on historical data », *Us patent & trademark Office*, 20050071741, USA
- Beitzel S. M., Jensen E. C., Chowdhury A., Grossman D., Frieder O., 2004, « Hourly analysis of a very large topically categorized Web query log », pp. 321 -328, *in : Proceedings of Sheffield SIGIR - Twenty-Seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

- Bharat K., Broder A., 1998, « A technique for measuring the relative size and overlap of public Web search engines », *Computer Networks and is ON Systems*, 30(1-7), pp. 379-388.
- Broder A., Kumar R, Maghoul F, Raghavan R, Rajagopalan S,, Stata R. et al., 2000, « Graph Structure in the Web », *Computer Networks*, 33, pp. 309-320, <http://hsd.soc.cornell.edu/curricular/WebStructure.pdf> (consulté le 14/04/08).
- Buckley C.Voorhees E. M., 2000, « Evaluating evaluation measure stability », *SGR Forum*, ACM Special Interest Group on Information Retrieval, pp. 33-40.
- Cacheda F, Viña A., 2001, « Understanding How People Use Search Engines : A Statistical Analysis for E-Business », pp. 319-325, in *The e-2001 E-Business and E-Work Conference and Exhibition*, vol. I.
- Chignell M. H., Gwizdka J., Bodner R. C., 1999, « Discriminating meta-search : a framework for evaluation », *Information Processing & Management*, 35(3), pp. 337-362.
- Chu H., Rosenthal M., 1996, « Search Engines for the World Wide Web : A Comparative Study and Evaluation Methodology », pp. 127-135 in : *Proceedings of the ASIS Annual Meeting*,
- Ding W., Marchionini G., 1996, « A comparative study of web search service performance », pp. 136-142 in : *Proceedings of the 59th American Society for Information Science Annual Meeting : Learned Information*, 33.
- Ford N., Miller D., Moss N., 2002, « Web search strategies and retrieval effectiveness: an empirical study » *Journal of Documentation*, 58(1), pp. 30-48.
- Gordon M., Pathak, R, 1999; « Finding information on the World Wide Web: the retrieval effectiveness of search engines », *Information Processing & Management*, 35(2), pp. 141-180.
- Griesbaum J., 2004, « Evaluation of three German search engines : Altavista.de, Google.de and Lycos.de », *Information Research*, 9(4), <http://informationnet/ir/9-4/paper189.html>, consulté le 25/08/08.
- Griesbaum J., Rittberger M., Bekavac B., 2002, « Deutsche Suchmaschinen im Vergleich: AltaVista.de, Fireball.de, Google.de und Lycos.de », pp. 201-223, in : Hammwöhner R., Wolff C., Womser-Hacker C., eds, *The Information und Mobilität. Optimierung und Vermeidung von Mobilität durch Information*, 8^e Internationalen Symposium für Informationswissenschaft, UVK
- Gulli A., Signorini A., 2005, « The indexable Web is more than 1 15 billion pages », pp. 902-903, in : *Special Interest Tracks and Posters of the 14th International Conference on World Wide Web*, Chiba, Japan.
- Hock R., 2004, « The latest field trip : an update on field searching in web search engines », *Online*, 28(5), pp. 15-21.
- Höchstötter N., Koch M., 2008, « Standard Parameters for Searching Behaviour in Search Engines and their Empirical Evaluation ». *Journal of Information Science*, à paraître.

- Hotchkiss G., 2000, *Inside the mind of a searcher*, <http://www.enquioresearch.com/campaigns/Inside%20the%20Mind%20of%20the%20Searcher.pdf> (consulté le 18/08/08).
- Ingwersen R, Järvelin K., 2005, *The Turn : Integration of Information Seeking and Retrieval in Context*, Dordrecht, Springer
- Jansen B. J., 2000, « The effect of query complexity on Web searching results », *Information Research*, 6(1), <http://informationr.net/ir/6-1/paper87.html>
- Jansen B. J., Spink A., 2006, « How are we searching the World Wide Web ? A comparison of nine search engine transaction logs », *Information Processing & Management*, 42(1), pp. 248-263.
- Lawrence S., Giles C. L., 1999, « Accessibility of Information on the web », *Nature*, 400(8), pp. 107-109.
- Leighton H.V., Srivastava J., 1999, « First 20 Precision among World Wide Web Search Services (Search Engines) » *Journal of the American Society for Information Science*, 50(10), pp. 870-881.
- Lewandowski D., 2004a, « Abfragesprachen und erweiterte Suchfunktionen von www-Suchmaschinen », *Information Wissenschaft und Praxis*, 55(2), pp. 97-102.
- 2004b, « Date-restricted queries in web search engines », *Online Information Review*, 28(6), pp. 420-427.
- 2006a, « Aktualität als erfolgskritischer Faktor bei Suchmaschinen », *Information Wissenschaft und Praxis*, 57(3), pp. 141-148.
- 2006b, « Suchmaschinen als Konkurrenten der Bibliothekskataloge : Wie Bibliotheken ihre Angebote durch Suchmaschinentechnologie attraktiver und durch Öffnung für die allgemeinen Suchmaschinen populärer machen können », *Zeitschrift für Bibliothekswesen und Bibliographie*, 53(2), pp. 71-78.
- 2006c, « Zur Bewertung der Qualität von Suchmaschinen », pp. 195-199, in : Eberspächer J., Holtel S., eds, *Suchen und Finden im Internet*, Heidelberg, Springer
- 2008a, « A three-year study on the freshness of Web search engine databases », *Journal of Information Science*, à paraître.
- 2008b, « The Retrieval Effectiveness of Web Search Engines : Considering Results Descriptions » *Journal of Documentation*, 64, à paraître.
- Lewandowski D., Mayr R., 2006, « Exploring the Academic Invisible Web », *Library HiTech*, 24(4), pp. 529-539.
- Lewandowski D., Wahlig H., Meyer-Bautor G., 2006, « The Freshness of Web search engine databases » *Journal of Information Science*, 32(2), pp. 133-150.
- Machill M., Neuberger C., Schweiger W, Wirth W, 2003, « Wegweiser im Netz : Qualität und Nutzung von Suchmaschinen », pp. 13-490, in : Machill M., Welp C., eds, *Wegweiser im Netz*, Gütersloh, Bertelsmann Stiftung.

- 1 Nijludui l o., ZUUU, « Kaising reliadillty OT vveu bCdruil LUUI icbcdii.ii unuugn replication and chaos theory » *Journal of the American Society for Information Science*, 51(8), pp. 724-729.
- Notess G. R., 2000, « Boolean Searching on Google », 03/1 1/2000, <http://www.searchengineshowdown.com/features/google/googleboolean.html> (consulté le 14/04/08).
- 2007, « Search Engine Features Chart », 01/10/07, <http://www.searchengineshowdown.com/features/> (consulté le 14/04/08).
- Ntoulas A., Cho J., Olston C., 2004, « What's New on the Web ? The Evolution of the Web from a Search Engine Perspective », *Thirteenth WWW Conference*, New York, USA.
- Ojala M., 2002, « Web search engines: Search syntax and features », *Online*, Wilton, Connecticut, 26(5), p. 28.
- Othman R., Halim N. S., 2004, « Retrieval features for online databases : Common, unique, and expected », *Online Information Review*, 28(3), pp. 200-210.
- Parasuraman A., Zeithami V.A., Berry L.L., 1988, « SERVQUAL : a multi-item scale for measuring consumer perceptions of service quality » *Journal of Retailing*, 64(1) : pp. 12-40.
- Risvik K. M., Michelsen R, 2002, « Search engines and Web dynamics », *Computer Networks*, 39(3), pp. 289-302.
- Schmidt-Maenz N., Bomhardt C., 2005, « Wie suchen Onliner im Internet ? », *Science Factory ! Absatzwirtschaft*, 2, pp. 5-8.
- Schmidt-Maenz N., Gaul W., 2005, « Web Mining and Online Visibility », pp. 418-425, in : Weihs C., Gaul W., eds, *Classification - The Ubiquitous Challenge*, Springer:
- Schmidt-Maenz N., Koch M., 2006, « A General Classification of (Search) Queries and Terms », pp. 375-381, in : *The 3rd International Conference on Information Technologies : Next Generations*, Las Vegas, Nevada, USA.
- Schwartz C., 1998, « Web search engines », *Journal of the American Society for Information Science and Technology*, 49(11), pp. 973-982.
- Silverstein C., Henzinger M., Marais H., Moricz M., 1999, « Analysis of a Very Large Web Search Engine Query Log », *SIGIR Forum*, 33(1), pp. 6-12.
- Singhal A., Kaszkiel M., 2001, « A case study in web search using TREC algorithms », pp. 708-716, in : *the 10th international conference on World Wide Web*, Hong Kong.
- Spink A. Jansen B. J., 2004, *Web Search : Public Searching of the Web*, vol. 6, Dordrecht, Boston/London, Kluwer Academic Publishers.
- Spink A., Jansen B. J., Blakely C., Koshman S., 2006, « A study of results overlap and uniqueness among major Web search engines », *Information Processing & Management*, 42(5), pp. 1379-1391.

- Spink A., Ozmutlu S., Ozmutlu J. C, Jansen B. J., 2002, « u.s. versus European Web Searching Trends », *SIGIR Forum*, 36(2), pp. 32-38.
- Spink A., Wolfram D., Jansen B. J, Saracevic T., 2001, « Searching the Web :The Public and Their Queries » *Journal of the American Society for Information Science and Technology*, 52(3), pp. 226-234.
- Sullivan D., 2005, « Search Engine Sizes », SearchEngineWatch.com, <http://searchenginewatch.com/showPage.html?page=2156481> (consulté le 14/04/08).
- Vaughan L., 2004, « New Measurements for Search Engine Evaluation Proposed and Tested », *Information Processing & Management*, 40(4), pp. 677-691.
- Vaughan L., Thelwall M., 2004, « Search Engine Coverage Bias : Evidence and Possible Causes », *Information Processing & Management*, 40(4), pp. 693-707.
- Vaughan L., Zhang Y., 2007, « Equal representation by search engines ? A comparison of websites across countries and domains » *Journal of Computer-Mediated Communication*, 12(3), article 7, <http://jcmc.indiana.edu/vol12/issue3/vaughan.html>.
- Véronis J., 2006, « A comparative study of six search engines », université de Provence, <http://sites.univ-provence.fr/veronis/pdf/2006-comparative-study.pdf> (consulté le 25/08/08).
- Williams M. E., 2005, « The State of Databases Today: 2005 », pp. XV-XXV, in : *Gale Directory of Databases*, vol. 2, Detroit, Mich., Gale Group.
- Wolff C., 2000, « Vergleichende Evaluierung von Such- und Metasuchmaschinen » pp. 31-38, in : *7 Internationales Symposium für Informationswissenschaft (ISI 2000)*, Darmstadt, Germany, Universitätsverlag Konstanz.