

Chronique

de la recherche sur le web

par Marc Duval

27 juin 2002, vol.2, n°6

MSN Search, «le tigre dans votre moteur»

MSN Web Search, ou MSN Search de son diminutif, est apparu tardivement en 1998. Il correspond à l'entrée en scène avec force de Microsoft qui était en latence depuis 1996. A cette époque, la recherche dans le web était faite en collaboration avec AltaVista, Infoseek, Lycos et Snap-Inktomi, chacun ayant un automate distinct. De ces quatre compères, seul demeure AltaVista. Infoseek a disparu, Lycos s'est départi de sa base de données et Snap a été absorbé par NBCi. Cette association a duré peu de temps. Le 7 octobre 1998, Inktomi prenait la relève, puis MSN est revenu à AltaVista pour enfin reprendre Inktomi. Depuis, MSN Search a essaimé sur le Globe.

MSN se positionne comme étant un portail d'information généraliste. Il sert à centraliser les services offerts par Microsoft (MSN Automobile, MSN Money, MSN Shopping, GamingZone, HomeAdvisor, MSNBC Health, Encarta, Auctions MSN, Music Picture It! On MSN, et MSN Hotmail, sans oublier la promotion de ses logiciels). Tous ces services sont disponibles sur ses plateformes de langue anglaise. Certains ont été adaptés au pays. Ainsi, MSN Money n'est pas sur la page canadienne française, mais il est adapté au marché français. Le site n'a ni de recherche d'actualités, ni d'images, ni de documents multimédias. Microsoft a la seule prétention de livrer de l'information à travers ses sites reliés par une thématique appelée

«chaîne». Chacun de ses sites nationaux est associé à des services d'informations locaux comme Radio-Canada, Branchez-vous, Cyberpresse... pour le Canada français; les agences de fil de presse AFP et Reuters pour la France. MSN inclut ses services dans le résultat d'une recherche.

Le 26 octobre 2000, MSN France avait introduit le service de référence « Nous cherchons, vous trouvez ». Quinze mois plus tard, Google inaugurerait le sien et Yahoo! s'associait à LiveAdvice. Le service de MSN France a été interrompu le 1^{er} juillet 2001. L'expérience de MSN se répétera-t-elle avec Google?

MSN Web Search

MSN Web Search n'est pas un automate de recherche ou plus familièrement un moteur de recherche. MSN Search est un site de recherche composé d'un automate tiers et d'un répertoire. L'automate de recherche est fourni par Inktomi et le répertoire - appelé Guide Web - par Looksmart, par des rédacteurs de MSN et par la société La Mine pour le site de MSN France.

Bien qu'Inktomi ait un bon produit dont profite HotBot, MSN ne livre pas toutes les promesses révélées dans son aide. Ainsi, les actualités ne sont pas affichées dans le résultat de recherche. Cependant, MSN a adapté le service d'Inktomi à ses clientèles nationales. Cette configuration ciblée est

profitable aux internautes parce qu'ils y trouvent des sites de « proximité ». Par exemple, le site de DSI avec la requête « dsi » uniquement se retrouvera 1^{er} sur MSN Canada, mais loin sur MSN France. Une recherche avec la requête « pascal » dans la case de recherche d'accueil donnera un résultat différent sur MSN Canada et sur MSN USA. Toutefois, si l'on utilise les mêmes paramètres de recherche du formulaire pour une même requête, les réponses seront identiques.

Les faiblesses et les forces

MSN Web Search a cinq faiblesses et quatre forces. Les faiblesses sont la troncature, les requêtes à opérateurs multiples, la personnalisation, la restriction des formats de documents et l'agrégation des sources d'informations. Ses forces sont la base de données de Inktomi, le tri des références, la recherche par la profondeur d'un répertoire informatique et le langage de recherche développé.

Les faiblesses

- La troncature
- Les requêtes à opérateurs multiples
- La personnalisation
- Restriction des formats de documents
- L'agrégation des sources d'informations

La troncature

Selon l'aide, la troncature est permise à la fin d'un mot pour rechercher ses déclinaisons et à l'intérieur du mot pour remplacer une lettre. Par exemple, « év*nement » trouverait « évènement » et « événement ». Malheureusement, elle est inefficace autant sur le formulaire qu'en utilisant le symbole « * ».

Les requêtes à opérateurs multiples.

Contrairement aux autres logiciels de recherche, MSN est configuré pour n'accepter que l'inclusion (ET) des mots. L'union (OU) et l'exclusion (SAUF) des mots d'un résultat de recherche est inefficace. Pour les activer, il faut utiliser la commande « Expression booléenne » de la liste déroulante du formulaire. L'internaute ne connaissant pas cette particularité ne pourra profiter de la puissance d'Inktomi.

Les opérateurs booléens sont les signes « + » et « - », ou « AND », « AND NOT » et « OR » en plus de libellés de la liste déroulante. Bien que Inktomi accepte les minuscules, le nombre de références trouvées sera plus élevé avec les majuscules.

Notons que la case de recherche accepte 150 caractères, ce qui est suffisant pour une recherche courante mais peut être restreignant pour une recherche complexe. De plus, cette case en page d'accueil n'est accompagnée d'aucun conseil ou de recherche d'expression. L'internaute doit savoir qu'il faut utiliser les « guillemets anglais » pour rechercher un groupe de mots.

La personnalisation

MSN Search offre une personnalisation minimale comprenant le choix du nombre de références par page et le choix du format d'affichage par titre ou par titre et description. Sur les sites en anglais, MSN a ajouté le choix d'activer le correcteur orthographique.

La restriction de formats de documents

Inktomi donne accès à plusieurs formats de documents parmi lesquels se trouve le format « pdf » mais il ne les affiche pas. Il donne l'adresse web de la page dans laquelle le document se trouve. Sa pertinence n'est pas prise en défaut, mais la présentation

du résultat. Il n'a pas suivi l'évolution des autres automates qui ont des recherches d'images et de documents audio-visuels explicites.

L'agrégation des sources d'informations

Le résultat de recherche est divisé en plusieurs sections selon la requête et le site national. Sur le site amiral, le résultat est divisé en Sites présentés, en Extraits d'actualité, en Sites commandités, en Résultat de recherche général et en «Sites Web corrélatifs».

Les sites présentés regroupent un peu de tout sans distinction provenant des services de MSN comme Encarta, de filiales, de partenaires, d'annonceurs.

Les «Extraits d'actualité» semblent avoir été retirés alors que les autres sites de recherche en fait un élément important.

Les sites commandités proviennent de la société Overture. Ils sont intégrés au résultat de recherche. Aucune inscription ne relève la présence d'annonceurs, mais les premiers sites de plusieurs recherches sont ceux de sociétés (domaine « .com »).

Le résultat de recherche général est la plus «subversive». Elle comprend des liens commandités d'Overture, des pages du guide Web de MSN et de Looksmart, les «10 sites les plus populaires» de Ask Jeeves, sans oublier les références de Inktomi qui viennent en dernier.

Les pages provenant du guide Web peuvent dépasser les 2 pages. Or, la plupart des internautes ne consultent pas au-delà de 2 pages. Ils ne peuvent donc voir le résultat obtenu par Inktomi.

Enfin, si la référence mène à une page d'erreur, MSN suggère des adresses qui

pourraient correspondre au site introuvable. Tous ses liens ne sont pas identifiés. Ils peuvent tromper l'internaute.

Ses forces

MSN compte 4 forces:

- La base de données de Inktomi
- Le tri des références
- La recherche par la profondeur du répertoire
- Le langage de recherche développé

La base de données de Inktomi

La base de données de Inktomi est l'une des plus grandes. Elle rivalise avec celle d'AltaVista, mais non avec Google et AlltheWeb qui ont atteint les 2 milliards de documents. Inktomi a choisi de privilégier la qualité plutôt que la quantité en constituant une base de données sélective.

Le tri des références

MSN offre la fonction de tri qui ne se retrouve sur aucun autre automate. Quatre choix sont disponibles: par pertinence (tri par défaut), par titre, par date et par structure (répertoire d'une adresse URL). Le tri est actif sur tout le résultat, mais au lieu d'être étendu sur toutes les références, chaque page est triée. Selon la recherche, il peut apporter une nouvelle dimension au résultat et déjouer la présentation par défaut dans lequel les pages commanditées sont cachées.

La recherche par la profondeur du répertoire

La recherche par la profondeur du répertoire donne un peu plus de contrôle au résultat. Nous pouvons supposer que l'information qui sera le plus près de la racine du répertoire (la page d'accueil) sera la plus

intéressante. En pratique, elle permet de scinder le résultat d'une recherche pour essayer de trouver l'aiguille dans la meule de foin.

Le langage de recherche développé

Inktomi a le langage de recherche le plus développé. Ce langage se retrouve chez HotBot et dans une moindre mesure chez MSN. La différence entre les deux sites tient du formulaire ; MSN n'a retenu que l'essentiel. D'ailleurs, ce formulaire répond aux principaux critères de recherche.

Une variante existe aussi entre HotBot et MSN: la recherche par domaine. Chez HotBot, le préfixe «**domain:**» donne toutes les pages d'un site - la référence est accompagnée de «**See results from this site only**» / «**voir les résultats de ce site uniquement**». Ce préfixe est inopérant chez MSN. Il faut utiliser la recherche par domaine du formulaire.

Présente sur HotBot, la recherche par date n'existe pas chez MSN Search. Cependant, la recherche par préfixes de date est recensée dans la plupart des descriptions de MSN ou de HotBot. Ces préfixes sont «after», «before» et «within». Les auteurs font suivre ces préfixes par des «:» tel que recommandé par le guide de HotBot. En fait, ces préfixes de date sont inopérants écrits ainsi. Il faut changer les deux-points par le

signe «=». La formule devient: «**after=jj/mm/aa**», «**before=jj/mm/aa**» et «**within=jj/mm/aa**». La recherche par préfixes n'est pas aussi efficace que l'utilisation du formulaire de HotBot.

Conclusion

MSN Search est un site de recherche qui convient à la plupart des internautes. Au cours de son histoire, il a «évolué» vers un contenu plus commercial, délaissant des fonctions reprises par d'autres comme l'affichage de l'actualité dans le résultat de recherche. De plus, ses caractéristiques indéniables ont été occultés par l'agrégation de ses sources anonymes. Si MSN demeure parmi les sites les plus fréquentés, ce n'est pas en raison de la performance de son automate de recherche, mais de ses guides et de ses sites associés, des produits de Microsoft, de la notoriété de la société informatique et dans une certaine mesure d'une clientèle «fidèle».

Marc Duval
Service de recherche documentaire DSI
1280-4, bd de Montarville.
Longueuil. Québec. Canada
J4B 8B4
marcduval@dsi-info.ca
<http://www.dsi-info.ca>