

INTERNET MODE D'EMPLOI

Les moteurs de recherche

JACQUES RIOT ^[1]

Voici le premier volet d'une série dont l'objectif est de vous aider à optimiser vos recherches sur internet. Commençons par le commencement, c'est-à-dire par un petit tour d'horizon, sous forme de fiches, des différents moteurs.

Les principaux moteurs de recherche

La bataille a fait rage sur le front des moteurs de recherche, et les rescapés sont de moins en nombreux : ainsi, AltaVista, AllTheWeb sont maintenant des filiales de Yahoo!, et utilisent donc son moteur... lequel n'en a plus pour longtemps ! En effet, en juillet 2009, Microsoft et Yahoo! ont conclu un partenariat, aux termes duquel Bing, le moteur de recherche du numéro un du *soft*, va devenir le moteur des portails de Yahoo!. L'objectif est bien évidemment, en mariant la puissance technologique de l'un à la force de frappe commerciale de l'autre, de s'attaquer à l'hégémonie du géant Google, dont le moteur est choisi en France pour plus de 9 recherches sur 10 (90,2 %), contre 3 % pour Bing et 2,1 % pour Yahoo! (mai 2010).

Pourtant, quelques courageux ont à cœur d'élargir l'offre et de ne pas laisser le Net aux mains des deux supergroupes, à l'instar d'Exalead. Cette petite société française très innovante est plutôt tournée vers les solutions de « B2B » (*Business to Business*), et son moteur reste très confidentiel. Notons toutefois qu'elle a développé une technologie que ce dernier est le seul à proposer pour le moment, Voxlead, qui permet une recherche de mots dans les vidéos (une option dont elle a également

Mots-clés

informatique, internet, média, outil et méthode

[1] Professeur de construction mécanique au lycée Jean-Macé de Vitry-sur-Seine (94).

Google

Fondée le 27 septembre 1998 dans la *Silicon Valley*, en Californie, par Larry Page et Sergey Brin, Google est principalement connue pour son moteur de recherche à la situation monopolistique, mais également par quelques-uns de ses logiciels emblématiques, tels que Google Earth.

Le terme mathématique *googol* (« gogol » en français, signifiant 10^{100}), est à l'origine de son nom, qui serait devenu Google à la suite d'une erreur dans l'enregistrement du nom de domaine.

En 2010, Google possède un parc de 2 millions de serveurs, le plus important au monde (2% du nombre total de machines). En 2008, son moteur de recherche a indexé plus de 1 000 milliards de pages Web. Et, de façon générale, pour une même requête, il retourne davantage de résultats que les autres moteurs.

Google propose la recherche d'images, de vidéos, dans l'actualité, sur une carte (Google Maps), de livres, de blogs, et beaucoup d'autres possibilités. Par exemple, en bas de l'écran de résultats, « Rechercher dans ces résultats » pour affiner sa requête ainsi que l'accès à des suggestions de « recherches associées ».

Son interface, sous son apparent dépouillement, recèle d'innombrables options (voir ci-contre).

Il est également possible de créer gratuitement un compte Google (voir « À la recherche du temps perdu », *Technologie* n° 165, janvier-février 2010), qui permet notamment d'accéder à son historique et ses favoris depuis n'importe quel ordinateur, ou de partager des photos en ligne, grâce au logiciel Picasa (pour le télécharger, aller dans « plus » en haut à gauche de la page d'accueil de Google, puis faire « et encore plus »). La barre d'outils Google téléchargeable une fois le compte créé donne également de nouvelles options, notamment la possibilité de laisser un commentaire sur la page en cours ou de lire ceux laissés par des internautes (lire l'encadré « SearchWiki, c'est fini ! »).

www.google.fr

SearchWiki, c'est fini !

Je vous avais présenté cet outil dans le numéro 165 de janvier-février, mais, le 4 mars, Cédric Dupont, chef de produit chez Google, qui l'avait inauguré il y a près de deux ans, a annoncé la fin de SearchWiki. C'est l'intégration de favoris Google dans les pages de résultats qui le remplace. Des étoiles sont désormais présentes à côté de chaque résultat – à condition d'être connecté à son compte Google. Lorsque vous effectuez une recherche, un encart situé sous la barre de recherche s'affiche si des sites marqués par vos soins correspondent à votre requête. Voici les actions que vous pouvez effectuer actuellement – à condition, toujours, d'être connecté à votre compte Google :

- **Pour ajouter un résultat aux favoris :** Cliquez sur l'icône  en regard d'un résultat de recherche Google. Ce résultat apparaîtra en haut de page à chaque fois que vous effectuerez la même recherche ou une recherche similaire. Il sera également enregistré dans vos favoris Google.
- **Pour afficher les résultats favoris :** Si, lors d'une recherche, une page que vous aviez auparavant ajoutée à vos favoris s'avère pertinente, elle sera affichée dans une section « Résultats dans les favoris » en haut de la page de résultats. Vous pouvez également afficher la liste complète de vos favoris : <https://www.google.com/bookmarks/?hl=fr>

- **Pour afficher les résultats masqués :** Si vous avez supprimé un résultat, ce dernier restera masqué chaque fois que vous effectuerez la même recherche. Les résultats masqués sont répertoriés dans la section « Résultats supprimés », en bas de la page.
- **Pour afficher vos commentaires :** Vous pouvez attribuer un commentaire à un résultat, ou le supprimer, et le consulter chaque fois que ce dernier apparaîtra.

Yahoo!

Yahoo! est créé par David Filo et Jerry Yang en janvier 1994 à la Leland Stanford Junior University (*Silicon Valley*), puis devient une entreprise en mars 1995. Son nom aurait été choisi en référence aux Yahoos, créatures des *Voyages de Gulliver* de Jonathan Swift. Yahoo!, site Web le plus visité au monde en 2004, lance Yahoo! France en septembre 1996 et prend, en un mois, la première place des sites visités en France.

Annuaire Web à l'origine, il s'est transformé en portail où l'on trouve le service Yahoo! Mail (stockage illimité, sous réserve de ne pas avoir d'inactivité de plus de quatre mois), une messagerie instantanée, l'hébergement de listes de diffusion, des jeux en ligne, des *chats*, des portails d'information variés (économie, actualités, sports, etc.)

À l'instar de Google, l'interface de Yahoo! propose la recherche d'images, de vidéos, dans l'actualité, sur une carte..., et donne accès à un certain nombre d'options. Moins nombreuses cependant que chez son concurrent, et parfois moins performantes : ainsi, le moteur ne fournit des suggestions orthographiques que s'il parvient à définir le contexte de la requête, qui doit donc contenir plusieurs mots. Signalons toutefois une application intéressante, comparable au feu SearchWiki de Google : Search Pad (accès en haut à gauche de l'écran de résultats). Elle permet de créer une liste où l'on peut ordonner les sites que l'on a visités et de prendre des notes au fil des recherches. Ces listes et notes peuvent être enregistrées, sous réserve d'avoir un compte Yahoo!, gratuit, et d'y

être connecté. Cette application, très utile, peut servir en classe pour cibler et limiter le domaine de recherche des élèves (voir ci-dessous).

<http://fr.yahoo.com>



■ Le Search Pad

Exalead

Fondée en 2000 par Patrice Bertin et François Bourdoncle, ouvert au public en octobre 2006, la française Exalead propose un moteur qui permet la recherche d'images, de vidéos, de blogs, dans les forums, dans Wikipédia. Ce moteur a indexé actuellement plus de 8 milliards de pages.

L'interface d'Exalead, très dépouillée, propose des miniatures de pages que l'on peut personnaliser et redimensionner. Elle affiche un aperçu des pages pour chaque résultat. Une « zapette » (voir ci-contre) permet différents filtres tels que le choix du type de fichiers recherchés, des termes associés...

Exalead a également mis au point un moteur de recherche pour Wikipedia qui permet, entre autres, la navigation dans l'encyclopédie grâce à un « nuage » de mots-clés (catégories, personnes, lieux, organisations, et termes associés).

www.exalead.com



■ La page d'accueil

Bing

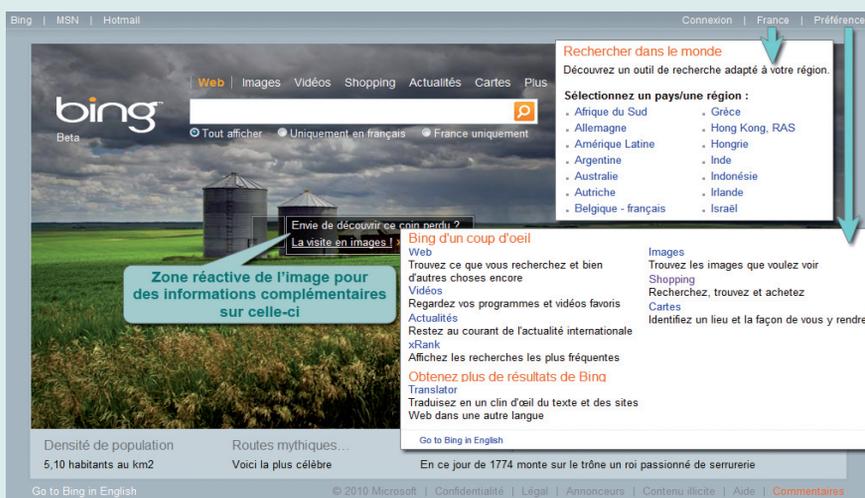
Anciennement Live Search, Windows Live Search puis MSN Search, Bing est un moteur de recherche développé par Microsoft. Il a été rendu public le 1^{er} juin 2009. Microsoft aurait choisi ce nom, simple onomatopée, parce qu'il est « mémorable, court, facile à épeler et peut fonctionner comme une URL partout dans le monde ».

Bing propose lui aussi la recherche d'images, de vidéos, dans l'actualité, sur une carte (Bing Maps), et la comparaison de prix via Ciao.

L'interface de Bing propose peu d'options, ciblant davantage le grand public. Elle présente chaque jour une nouvelle image comportant des zones masquées, réactives au passage de la souris, qui vous mènent à des informations complémentaires concernant cette image.

À noter que les suggestions de recherche ne sont basées que sur les recherches effectuées antérieurement.

www.bing.com



■ La page d'accueil

doté le site de la présidence de la République). Dernier rebondissement, on vient d'apprendre que Dassault systèmes a fait l'acquisition d'Exalead ce 9 juin.

Leur fonctionnement

Quels sont les principaux paramètres pris en compte par les moteurs pour trouver des résultats à une requête parmi les milliers de milliards de pages d'internet (1 000 milliards référencées par Google en juillet 2008) et les hiérarchiser ?

- La présence du mot-clé dans la page. C'est le premier critère. Le classement diffère selon la présence du mot dans l'URL (adresse internet), dans le titre de la page, selon sa position dans la page mais aussi sa mise en avant (souligné, en gras) et sa récurrence (répétition).

- L'ordre des mots-clés dans la page, respectant celui de la requête, et leur proximité.

- La présence du mot-clé dans les balises *meta*, des champs dans le code de la page invisibles au lecteur. À la suite de certains abus de référenceurs de sites, cette méthode est de moins en moins prise en compte par les robots qui scrutent et indexent les pages internet.

- La popularité de la page. Le robot analyse également le nombre de liens d'autres sites pointant vers la page. Plus il est important, plus cette page sera jugée populaire, et meilleur sera son classement.

Afin d'affiner le classement, les robots utilisent également des critères qui prennent en compte l'usage fait par les internautes de la page : le nombre de fois où cette page est choisie dans les résultats de recherche, le temps passé par les utilisateurs sur celle-ci, leur retour ou non vers la page de résultats...

Les résultats proposés par le moteur sont donc issus d'une équation, tenue secrète et constamment modifiée, intégrant tous ces paramètres et bien d'autres. Une note est attribuée à la page. Plus elle est élevée, plus ce résultat se placera en tête du classement.

Mais les moteurs ont aussi leurs limites, que nous verrons dans un prochain article, qui donnera également des conseils, simples mais méconnus, pour formuler les requêtes les plus efficaces. ■



■ Une page de résultats