

Chronique

de la recherche sur le web

par Marc Duval

25 juin 2003, vol. 3, n°6

Google revisité

Google est l'automate de recherche le plus connu. Il a étendu ses tentacules à travers le cyberspace prenant diverses apparences selon la particularité du partenaire. C'est sous l'angle de ses multiples visages qu'il sera abordé à travers le formulaire de recherche.

Trois sites utilisant Google ont été choisis pour illustrer les différences dans les formulaires: [AOL US](#), [HotBot](#) et Yahoo! ([Yahoo! US](#) et [Yahoo! Canada en français](#)). [Google US](#) et [Google Canada en français](#) sont les références. Parmi les quatre, AOL US n'a pas de formulaire. Son usage est d'ordre général en étant un complément au portail. Les deux sites HotBot et Yahoo! ont des particularités intéressantes que n'a pas Google à moins d'utiliser les requêtes complexes. Cet article aborde les opérateurs de recherche, le nombre de références par pages, la recherche par langue et par pays, le format de document, la recherche par dates, la recherche dans les parties d'une page web, les pages hyperliées et reliées, et le filtre parental (voir le [Tableau des opérateurs et des champs de Google selon le site de recherche](#) à la fin de la chronique).

Les opérateurs de recherche

Hormis AOL, les trois autres sites ont les opérateurs logiques et de proximité traditionnels. HotBot en a un de plus: «l'exclusion de l'expression». De ce point de vue, HotBot est supérieur à ses concurrents pour une recherche simple.

Le nombre de références par pages

Le nombre de références par pages permet d'avoir une vue d'ensemble. Dans la plupart des

cas, une page pertinente se trouve parmi les 30 premières réponses, la probabilité décroissante. Google et Yahoo! ont cette option dans leur formulaire. HotBot a choisi de le mettre dans ses préférences. A moins de connaître ce fait, cette option demeure masquée aux internautes. La haute vitesse aidant, il devient injustifié de se priver de ces 30 adresses URL. Sinon, Yahoo! coupe la poire en deux avec ses 15 références par page.

La langue et le pays

Tous les trois automates ont la recherche par langue et le même nombre de langues, soit 35. L'interface de Yahoo! Canada en a 33 - l'indonésien et le serbe manquant à l'appelé. Cependant, seul Yahoo! a la recherche par pays. Il est vrai que Google a cette recherche dans la page d'outils linguistiques. Toutefois, cette division des options de recherche entre deux pages est malaisée.

Le format de documents

L'internaute qui recherche un format spécifique de documents devra se rendre sur le formulaire de Google, seul à délivrer cette information.

La recherche d'archives

Google, comme tous les autres automates, indexe périodiquement une partie de sa banque de données selon ce que son «robot» aura perçu comme actif. Google laisse ainsi de côté plusieurs milliers de pages. Ces pages auront la date du passage du robot et non de leur création ou de leur d'enregistrement sur leur serveur. Certaines d'entre elles n'auront été visitées qu'une fois par semaine, par mois ou plus longtemps. L'intervalle des visites est notée par 3 mois, 6 mois et 12 mois chez Google et Yahoo! Seul HotBot permet une recherche au-delà d'un an, soit jusqu'en 1994. Le Google d'HotBot va jusqu'à raffiner sa recherche en utilisant le quantième et le mois. Plus besoin

d'utiliser le rébarbatif préfixe de recherche par intervalle de dates «**daterange:**» et de faire la conversion dans le calendrier Julien ou d'utiliser des formulaires ad'hoc comme celui de [Google Ultimate Interface](#).

Google met à la disposition des internautes les pages indexées sous l'expression «**cached** » ou « **en cache** ». Seul Yahoo! US a cette option.

La recherche dans la page

Les automates de recherche divisent une page web afin de raffiner une recherche. Le titre, le texte, l'URL sont les trois principales parties. C'est une lapalissade que d'affirmer que tous font la recherche dans la page entière. C'est le seul lien qui unissent les quatre sites.

Yahoo! travaille sur le titre et sur l'URL, mais non sur le texte uniquement. Pour cela, il faut utiliser Google et HotBot, tout comme la recherche dans le texte de l'hyperlien.

La flexibilité d'un formulaire est un atout. Elle se retrouve dans la possibilité de multiplier la recherche selon l'emplacement des mots que l'internaute choisit. Le formulaire de Google n'a pas cette flexibilité contrairement à ceux de HotBot et Yahoo! HotBot va plus loin en proposant d'ajouter des cases de recherche liées aux différentes parties d'une page web.

Bien que les trois offrent la recherche par nom de domaine, c'est Yahoo! qui en facilite l'accès en en donnant une liste.

Les pages liées et les pages reliées

La recherche de pages liées et reliées donne une autre dimension de la recherche.

La première proposition élargie la recherche vers des pages qui suggèrent le lien de celle considérée pertinente.

La seconde proposition fait intervenir le contenu des pages dans une amorce de recherche sémantique ou contextuelle. Elle est souvent peu concluante. En raison de leur faible efficacité, HotBot les a exclus de son formulaire. Toutefois, seuls Google, Yahoo! US et Yahoo! France ont cette option de recherche. Le jour n'est pas loin où son efficacité s'accroîtra.

Le filtre parental

Le filtre parental est un rempart contre les pages licencieuses. Un filtre parental peut comporter divers degrés de tolérance. AOL n'a aucun filtre. HotBot module son offre entre l'exclusion, la complaisance et l'inclusion. Le formulaire de Yahoo et de Google ont l'option d'inclure ou d'exclure les pages « **offensantes** ». Yahoo! Canada en français a cette option contrairement au formulaire des interfaces en français de Google. Ni Yahoo! France, ni Google France n'ont le filtre parental. Si Yahoo! reconnaît la faiblesse du filtre, Google donne un hyperlien vers une explication de son efficacité.

Conclusion

Que peut-on retenir des avatars du formulaire de Google?

Premièrement, Google Technology inc. a retenu un formulaire restrictif pour son site et peu détaillé. La société complique même la recherche en la divisant entre le formulaire général et la recherche dans les préférences linguistiques. Bref, Google a un formulaire bien peu convivial

Deuxièmement, Google Technology inc. a distribué des avantages à HotBot et à Yahoo! (Yahoo! US et Yahoo! Canada). Tous les deux offrent des champs de recherche multiples pour les associer à chacun des termes de recherche. On retrouve en exclusivité «l'exclusion d'une expression» et la recherche étendue par dates chez HotBot . Du côté de Yahoo!, c'est la liste des domaines.

Troisièmement, aucun des sites n'offre la puissance de Google. L'entreprise éponyme a décidé de restreindre la recherche en ne proposant pas directement sur son site un espace pour la recherche libre comme le fait ses concurrents AltaVista et AlltheWeb. Seuls les internautes avertis sont en mesure d'exploiter une partie du logiciel, soit en utilisant les préfixes de recherche ou encore en voyageant d'un formulaire à l'autre selon la particularité de la requête.

Quatrièmement, pour un usage général, le formulaire de Yahoo! est une offre mitoyenne et supérieure à celui du site de Google tant dans sa conception que dans le résultat de recherche où l'on

retrouve le fameux «cached», un correcteur d'orthographe et l'information boursière. L'intégration de différents services de Yahoo! est encore un atout malgré qu'elle ne soit pas encore aussi complète que souhaitée comme l'absence de cartes routières offerts par Google. Encore faut-il le savoir! Lorsque Yahoo! mettra en oeuvre Inktomi sur son site, Google perdra un formulaire plus performant que le sien.

Cinquièmement, pour une recherche dans les archives de Google, autant s'en remettre à HotBot.

Et sixièmement, les parents seront ravis d'apprendre qu'il faut éviter AOL qui n'a pas de filtre pour contrer le contenu pornographique. Les parents francophones le seront davantage en se rendant sur Yahoo! Canada en français qui dispose d'un tel filtre.

Évidemment, tout cela vaut pour un

anglophone. Pour les francophones, nous devons faire mauvaise fortune bon coeur et attendre le bon vouloir des sociétés. Google a omis d'inclure son filtre parental dans ses formulaires, HotBot France a uniquement Inktomi, Yahoo! France a un formulaire amputé. La bonne nouvelle vient de Yahoo! Canada en français qui a mis en ligne un nouveau formulaire semblable à celui de Yahoo! US. Quant à AOL France, son offre est minimale et AOL Canada n'a pas de formulaire en français.

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

Mots clés : Bibliothèques (vedettes matières) : recherche sur Internet ; recherche de l'information ; recherche documentaire ; ouvrages de références ; gestion de l'information / Nomade : recherche sur le Web ; moteurs de recherche / ODP : recherche sur le Net ; moteurs de recherche / Toile du Québec : gestion de l'information ; Internet - Guides ; guides perfectionnement à la recherche ; moteurs de recherche / Voila : internet - recherche d'info sur le web ; moteurs de recherche / Yahoo : recherche sur le Net ; recherche sur le Web ; recherche et traitement de l'information ; internet - information et documentation ; moteurs de recherche / Pages jaunes : renseignements-service Longueuil ; Identificateurs: formulaire de recherche

[\(Retour au début de la page\)](#)

Tableau des opérateurs et des champs de Google selon le site de recherche

Opérateurs/Champs/ Fonctions	AOL	HOTBOT	GOOGLE	GOOGLE CF	YAHOO! CF	YAHOO!	
ET	Non	Oui	Oui	Oui	Oui	Oui	
OU	Non	Oui	Oui	Oui	Oui	Oui	
SAUF	Non	Oui	Oui	Oui	Oui	Oui	
EXPRESSION	Non	Oui	Oui	Oui	Oui	Oui	
SAUF EXPRESSION	Non	Oui	Non	Non	Non	Non	
Nombre de références par pages	Non	Non / Voir Preferences	Oui / 10, 20, 30, 50, 100	Oui / 10, 20, 30, 50, 100	Oui / 10, 15, 20, 30, 40, 100	Oui / 10, 15, 20, 30, 40, 100	
Langues	Non	Oui / 35	Oui / 35	Oui / 35	Oui / 33	Oui / 35	
Pays	Non	Non	Non / Voir Languages Tool	Non / Voir outils linguistiques	Oui	Oui	
Format de document	Non	Non	Oui / pdf, Adobe postscript, word, Excel, Powerpoint, RTF	Oui / pdf, Adobe postscript, word, Excel, Powerpoint, RTF	Non	Non	
Date	Non	Oui / dernière semaine, 2 sem., 1 mois, 3 mois, 6 mois, un an, deux ans; intervalle de dates	Oui / Tout, 3 mois, 6 mois, 12 mois	Oui / Tout, 3 mois, 6 mois, 12 mois	Oui / Tout, 3 mois, 6 mois, 12 mois	Oui / Tout, 3 mois, 6 mois, 12 mois	
Recherche générale	Oui	Oui	Oui	Oui	Oui	Oui	
Recherche par titre	Non	Oui	Oui	Oui	Oui	Oui	
Recherche dans le texte	Non	Oui	Oui	Oui	Non	Non	
Recherche dans l'URL	Non	Oui	Oui	Oui	Oui	Oui	
Recherche dans le texte de l'hyperlien	Non	Oui	Oui	Oui	Non	Non	
Champs de recherche fixe	Non	Non	Oui	Oui	Non	Non	
Champs de recherche multiple	Non	Oui	Non	Non	Oui	Oui	
Recherche par nom de domaine	<i>sans domaine</i>	Non	Oui	Oui	Oui	Non	Non
	<i>avec domaines</i>	Non	Non	Non	Non	Oui	Oui
Recherche par pages liées	Non	Non	Oui	Oui	Non	Oui	
Recherche par pages reliées	Non	Non	Oui	Oui	Non	Oui	
Filtre parental	Non	Oui	Oui	Non	Oui	Oui	