

Première analyse du futur moteur de Microsoft

[Retour au sommaire de la lettre](#)

Microsoft, via sa filiale "search" MSN, a lancé il y a quelques jours de cela, une version "alpha" de son futur moteur de recherche. Une bonne occasion, pour effectuer quelques tests pour voir à quoi ressemble cette première mouture, tout en gardant en mémoire qu'il ne s'agit là que d'une version très provisoire... Ces tests nous fournissent cependant quelques données intéressantes quant à la future version finale du moteur.

Le 1er juillet dernier, Microsoft/MSN a sorti une version "beta preview" de son futur moteur de recherche en France (<http://techpreview.search.msn.fr/>) et dans 28 pays (<http://techpreview.search.msn.com/worldwide.aspx>).

Bien entendu, il ne s'agit que d'une première version, peut-être plus destinée à tester le robot (*spider*) du moteur que l'algorithme de pertinence lui-même. Il nous cependant semblé intéressant de commencer à étudier cet outil, ne serait-ce que pour mieux prendre en compte son évolution dans les mois qui viennent.

Nous avons donc examiné plusieurs critères d'étude : comment les requêtes sont-elles prises en compte (minuscules/majuscules, troncature, etc.) et les documents indexés, syntaxe de recherche (ET, OU, SAUF, syntaxe avancée, notamment par rapport à celle de Google, etc.), présentation des résultats, etc.

Prenons, dans un premier temps, en compte, la façon dont sont recherchés les mots clés : minuscules ? Majuscules ? Avec troncature ? Le moteur effectue-t-il ses recherches dans les balises meta et prend-il en compte les fichiers PDF et les pages dynamiques ? Réponses :

Prise en compte des requêtes et indexation

	
Minuscules / majuscules	Idem. IBM = Ibm = ibm
Accentuation	Importante. Le moteur recherche l'occurrence exacte du mot demandé. électricité recherche électricité et rien d'autre. electricite recherche electricite et rien d'autre.
Ordre des mots	Important. dakar paris donne un résultat différent de paris dakar .
Troncature	Pas de troncature par défaut. Seul le mot demandé est recherché (abonda ne trouve pas abondance).
Recherche dans les balises Meta Description des pages	NON
Recherche dans les balises Meta Keywords des pages	NON
Recherche dans les commentaires des codes HTML	NON
Indexation des documents PDF	OUI (peu de documents à ce format semblent cependant disponibles)
Indexation des sites dynamiques	OUI (nombreuses pages contenant des signes "?" et "&" dans l'url)

Très important : la syntaxe d'interrogation simple et avancée. Quelles sont les possibilités que MSN met à notre disposition pour effectuer des requêtes plus approfondies et efficaces?

Syntaxe de recherche	
	
ET	Espace par défaut. La requête moteur recherche est équivalente à "moteur ET recherche". Fonctionnent également : moteur AND recherche moteur and recherche moteur +recherche +moteur +recherche
OU	OR (en majuscules obligatoirement). Exemple : moteur OR recherche
SAUF	Signe - . Exemple : moteur recherche -google .
Expressions	Guillemets. Exemple : " moteur de recherche ".
Troncature forcée	NON. L'astérisque (*) ne fonctionne pas. rech* ne trouve pas "tout ce qui commence par rech".
Parenthèses	Semble fonctionner. Par exemple, la requête (" sergey brin " OR " larry page ") -(google OR backrub) semble donner des résultats cohérents.
Recherche sur un site	site: ou domain: site:www.abondance.com ou domain:www.abondance.com renvoient toutes les pages du site www.abondance.com présentes dans l'index du moteur. - Fonctionne également pour une recherche de sous-domaines : site:abondance.com ou domain:abondance.com renvoient les pages des sous-domaines du site (actu.abondance.com, lettres.abondance.com, etc.) - Fonctionne également avec un mot clé : google site:www.abondance.com
Recherche de backlinks	Non. La syntaxe link: ne fonctionne pas.
Recherche de sites similaires	Non. La syntaxe related: ne fonctionne pas.
Recherche de la version en cache de la page	Non. La syntaxe cache: ne fonctionne pas.
Recherche d'informations sur une page.	Non. La syntaxe info: ne fonctionne pas.
Recherche sur le titre des pages	Non. Les syntaxes title: , intitle: et allintitle: ne fonctionnent pas.
Recherche dans l'adresse des pages	Non. Les syntaxes url: , inurl: et allinurl: ne fonctionnent pas.
Recherche sur le format des fichiers	Non. La syntaxe filetype: ne fonctionne pas.

Enfin, au niveau de la présentation des résultats, quelles sont les grandes différences entre la version "alpha" du moteur de MSN et Google ?

Présentation des résultats		
		
Description pour chaque lien proposé		
Titre de la page	60 caractères maximum Titre issu de l'annuaire de MSN lorsque le site y est présent.	70 caractères maximum
Résumé de la page	200 caractères maximum. Résumé issu de l'annuaire de MSN lorsque le site y est présent. "Snippet" (extrait de la page contenant la requête demandée) sinon. Pas d'utilisation de la balise Meta Description.	160 caractères maximum. Contenu de la balise Meta Description et/ou "Snippet".
URL de la page	OUI	OUI
Lien vers un formulaire de "feedback"	OUI : 	NON
Taille de la page	NON	OUI
Date de dernière modification de la page	NON	Parfois
Autres possibilités	NON	Liens vers : - Version en cache de la page - Pages similaires - Autres résultats pour le site (clustering) - Traduction
Autres informations		
Affichage du nombre de résultats	NON	OUI
Affichage du temps de réponse	NON	OUI
Pages suivantes	Uniquement suivante + précédente	Possibilité de navigation dans une "fourchette" de 10 pages (suivantes ou précédentes)
Lien général vers un formulaire de "feedback"	OUI	OUI

Conclusion

Nous n'avons pas voulu, dans cette première analyse, comparer de façon quantitative et qualitative les résultats de Google et ceux de MSN "nouvelle version". En effet, il serait difficile de comparer un moteur qui existe depuis 1998 (Google) avec une version alpha d'une technologie naissante. Laissons cette tâche aux lettres R&R qui paraîtront dans les mois qui viennent :-)

En revanche, on peut déjà tirer de cette première étude quelques enseignements intéressants :

- Le moteur de MSN ne tient pas compte des balises Meta Description et Keywords.
- Il indexe les sites dynamiques et les documents au format PDF. Nous n'avons, en revanche, pas pu identifier dans l'index de fichiers aux formats Word, Excel ou Powerpoint...
- Sa syntaxe d'interrogation est pour l'instant rudimentaire.

Alors, bien entendu, cette version alpha souffre encore de quelques défauts non négligeables (lenteur, plantages, pertinence parfois hésitante des résultats, index quantitativement faible encore, etc.). Mais n'oublions pas qu'il s'agit là d'une version "alpha" qui doit s'améliorer dans les mois qui viennent.

Bref, rendez-vous à la rentrée et, quoi qu'il en soit, avant la fin de l'année 2004 pour l'apparition du nouveau moteur de recherche de MSN/Microsoft, en version finalisée, qui nous permettra de mettre en place une seconde batterie de tests plus affinés...