Mai 2000

- -> Nouveaux index et annuaire pour Altavista

Altavista a fait de nouvelles annonces concernant son moteur de recherche :
- Il récupère ("crawle") sur le Web aujourd'hui près de 1,2 milliards de pages Web pour n'en garder que 350 millions, ce qui en fait le plus gros index actuel devant celui d'All The Web. Il propose également 30 millions de fichiers multimédia (sons, images, vidéos).
- Un nouvel algorithme de classement des pages est également mis en place (un article spécial sur ce sujet est disponible dans la zone"abonnés" du site Abondance à l'adresse http://abonnes.abondance.com/articles/criteres-av.html).
- Un nouvel annuaire, dérivé de Looksmart et de l'Open Directory, proposant plus de 2 millions de sites Web.
- Altavista a également indiqué que son moteur répondait à 40 millions de requêtes par jour.

Source :
Altavista

- Altavista recherche les produits

Altavista (version américaine) a rajouté, en collaboration avec son site Shopping.com (8 millions de produits chez 700 vendeurs), une possibilité de recherche sur les produits, au travers du bouton radio "Products" sur sa page d'accueil. Cette possibilité est doublée du programme "AltaVista Rewards" (les "récompenses d'Altavista") qui vous fait gagner des points pour chaque achat ou recherche effectués sur le site. Un grand concours permettant de gagner notamment une Mercedes SLK 230 est également mis en place pour fêter ces nouvelles possibilités.
Source :
Altavista

- Lycos abandonne son index et passe chez Inktomi

Il était de notoriété publique que Lycos USA (comme Excite, d'ailleurs) n'avait quasiment plus indexé un nouveau site dans son index depuis octobre 1999. L'explication est peut-être connue aujourd'hui : Lycos vient d'abandonner son "vieil" index pour passer chez Inktomi. Les résultats de type "Web sites" sur le moteur américain seront donc proposés maintenant par Inktomi et non plus au travers de l'index propre à Lycos qui semble cependant encore utilisé par Lycos France (à la fois dans ses recherches "web français" et "web mondial"). Plus pour longtemps, certainement. Les résultats de Lycos USA semblent en tout cas encore instables au niveau des index employés. Il semble bien que cela soit celui d'Inktomi pour certaines requêtes, mais pas pour d'autres... Après plusieurs tests, il semblerait que les résultats d'Alltheweb (qui fournissait déjà les résultats de la recherche avancée de Lycos) et d'Inktomi soient mélangés. Nous continuons nos investigations. Rendez-vous le mois prochain pour une analyse de ces phénomènes étranges dignes de X-Files.

- Filtre familial pour Google L'annonce était attendue, c'est maintenant chose faite : Google va proposer un filtre familial en collaboration avec SurfWatch, l'un des leaders du domaine. La technologie de SurfWatch permettra d'éliminer des résultats de Google les pages traitant de pornographie, haine, racisme, violence, alcool et drogues si l'internaute le désire.
http://www.google.com/safesearch_help.html
D'autre part, il est possible, par la fonction cache:, d'obtenir la version d'une page dans la version "sauvegardée" par Google, au moment où il a indexé le document, et non pas dans sa version actuelle. La date de dernière indexation est également fournie. Essayez cache:www.abondance.com, par exemple :
http://www.google.com/search?q=cache%3Awww.abondance.com&meta=lr%3D%26hl%3Den
Enfin, Google a signé un accord avec Yahoo! Maps pour proposer des cartes lorsque le nom d'une rue d'une ville américaine est demandé. essayez la requête "100 independence avenue washington dc" par exemple :
http://www.google.com/search?q=100+independence+avenue+washington+dc&num=10&meta=hl%3Den%26lr%3D&safe=off&btnG=Google+Search

- Noeud papillon

Altavista, Compaq et IBM ont réalisé une étude conjointe sur le "web déconnecté". Les scientifiques des centres de recherche IBM, Compaq et altaVista ont donc achevé la représentation graphique d'une carte topographique complète du Web mondial, après le webmap d'Inktomi (l'annonce successive des deux études n'est peut-être pas d'ailleurs un hasard) et ont découvert l'existence de division entre différentes zones d'Internet, pouvant rendre la navigation sur le Web difficile, voire impraticable.
Les recherches qui avaient été effectuées auparavant, basées sur de simples

échantillonnages du Web, avaient permis de conclure à un haut degré de connectivité entre les sites.
Cependant, la recherche effectuée par IBM, Compaq et Altavista sur l'analyse de plus de 200 millions de pages Web, prouve (contrairement à ce que l'on croyait) que le Web Mondial est fondamentalement divisé en quatre grandes zones, chacune comprenant approximativement le même nombre de pages, environ 50 millions. On a pu constater de même qu'un nombre impressionnant de sites Web était inaccessible par le biais des liens hypertextes. Or, ces liens sont ce qu'un internaute utilise le plus au cours de ses navigations sur le reseau. La théorie du "noeud papillon" permet d'appréhender la dynamique comportementale du Web et son organisation complexe. La theorie du "noeud papillon" et les quatre zones du Web : C'est au fur et a mesure des recherches que la representation du Web s'est profilée en forme de noeud papillon : 90% du Web environ se divise en quatre grandes zones, les 10% restants se trouvant totalement déconnectés du "noeud papillon" en question.
Le " noeud " est constitue du "noyau ultra connecté" et contient à peu près à 56 millions de pages. Les internautes peuvent aisément naviguer entre ces sites, via les liens hypertextes. Ce noyau compact constitue le coeur du réseau Internet. La partie gauche du "noeud papillon" contient les pages "de création" et représente environ un quart du réseau. Elles permettent l'accès au coeur du Web (le noyau hyper-connecté) mais l'inverse n'est pas possible (le "noyau dur" n'a pas de liens vers elles). La partie droite du "noeud papillon" représente environ un cinquième du Web et est le contraire de l'aile droite. Les pages de destination sont accessibles depuis le noyau ultra connecté, mais aucun retour vers le noyau n'est possible ; c'est par exemple le cas des sites institutionnel d'entreprise qui reçoivent beaucoup de liens mais qui n'en offrent que très rarement. Des culs de sac du Web en quelque sorte.
La quatrième et dernière zone contient des pages "déconnectées", qui représentent environ un cinquième également du Web. Les pages déconnectées sont accessibles mais ne donnent pas acces au noyau ultra connecté et ne sont "pointées" par aucune page du web.
Cette etude, la plus vaste jamais realisée sur la topographie du Web, fait partie d'un projet de collaboration entre AltaVista, Compaq et IBM. Les chercheurs espèrent pouvoir mettre régulièrement à jour l'étude menée, sur une base réguliere de données collectées au moyen du moteur de recherche AltaVista et d'un logiciel serveur de connectivité avancée avec le système alphaServer de Compaq. Les centres de recherche d IBM analysent les données et contribuent au développement de la théorie du "noeud papillon".
Adresses :
http://websearch.about.com/internet/websearch/library/weekly/aa051600a.htm
http://www.almaden.ibm.com/cs/k53/www9.final/
http://www.liberation.fr/multi/actu/20000515/20000517.html
http://www.zdnet.fr/actu/inte/a0014297.html?nl
http://www.research.ibm.com/resources/news/20000511_bowtie.html
http://www.parc.xerox.com/istl/groups/iea/www/growth.html
http://tech.altavista.com/scripts/editorial.dll?fromspage=hb/t2.htm&categoryid=&only=y&efi=239&ei=1785926&ern=y

- Etude : les internautes naviguent beaucoup et trouvent peu !

Une étude, menée pour le compte de RealNames par la société Berrier Associates sur 1 000 personnes habituées de l'Internet, agées de 18 à 49 ans et interrogées par téléphone, a donné les principaux résultats suivants :
- 75% des internautes utilisent un outil de recherche sur le web.
- Mais 60% utiliseraient un outil de recherche plus efficace que ceux d'aujourd'hui si il leur était proposé.
- 70% des gens interrogés savent exactement ce qu'ils veulent lorsqu'ils font leurs recherches.
- La moitié des utilisateurs d'Internet passent plus de 70% de leur temps sur le web à chercher de l'information.
- 44% des internautes sont frustrés de leur navigation et des résultats trouvés.
- Quand ils ne trouvent pas ce qu'ils désirent, la plupart des internautes testent un autre outil de recherche, mais 20% abandonnent sans aller sur un autre outil.

Source :
Individual.com

- All The Web : syntaxe avancée.

Il est possible d'utiliser les fonctions suivantes sur le moteur All The Web (http://www.alltheweb.com/) :
t: recherche sur le titre des pages (équivalent du title: d'Altavista)
u: recherche sur l'url (équivalent du url: d'Altavista)
h: recherche sur le nom de domaine (équivalent du host: d'Altavista)
ml: reherche sur les liens vers le site (équivalent du link: d'Altavista)