

ZONE ABONNÉS : Archives de la lettre "R&R"

Novembre 2000

■ Au coeur de l'annuaire de Lycos France

Nota : merci à l'équipe de l'annuaire de Lycos France pour avoir donné, annoté, validé ces informations et corrigé les erreurs éventuelles qui s'étaient glissés dans l'article.

Organisation / Acceptation / Refus des sites

Quelques chiffres de fréquentation sur Lycos Europe (Lycos France ne communique pas ses chiffres propres, uniquement ceux de l'entité européenne du portail) :

Nombre de pages vues par jour : 16,5 millions.

En comptant Spray et Multmania: 1,1 milliard de pages vues par mois 21 millions de visiteurs uniques.

L'annuaire de Lycos France comprend à l'heure actuelle 64 000 sites web inscrits dans 5 000 catégories.

Les serveurs se situent en Allemagne, à Gütersloh, siège de Lycos Europe. Puisqu'on en est au chapitre de l'étranger, le portail Lycos France, et notamment son annuaire, ne recensent que des sites en français. Si votre site est disponible dans une autre langue, référez-le dans la version linguistique adéquate du réseau Lycos.

L'équipe qui s'occupe de l'annuaire (10 personnes, toutes présentes dans les locaux de Lycos, pas de télétravailleurs) traite en règle générale 3 000 soumissions par mois. Ce chiffre est bien inférieur à celui de Yahoo! France, par exemple (1 000 soumissions par jour). Cela est certainement dû à la jeunesse de l'outil (il n'était pas possible, auparavant, de proposer un site à l'annuaire de Lycos France puisque... il n'y avait pas d'annuaire) et au fait que les gens ont surtout pris l'habitude de soumettre leur site sur la partie "moteur". Or, les deux soumissions sont totalement différenciées (notons que Lycos va très prochainement basculer l'index de son moteur sur celui de Fast, plus d'infos très bientôt sur Abondance).

Seulement 10% du trafic total de Lycos France est généré par l'annuaire. Là aussi, la jeunesse de l'outil en semble la cause. Ce chiffre pourrait augmenter fortement dans les mois qui viennent, notamment lorsque la branche "entreprises" sera en ligne (voir plus loin).

Les netsurfeurs - qui sont spécialisés par thème, comme chez Yahoo! et Voila - reçoivent les soumissions de sites par mail. Ils utilisent un outil spécifique, disponible directement sur le Web, pour évaluer et rentrer les sites dans la base de données. Cet outil, venu de Lycos US, devrait être largement modifié dans les prochains mois. Que le site soit accepté ou refusé, s'il est déjà présent ou s'il n'existe pas de catégorie pertinente pour le classer, un e-mail est envoyé à l'adresse fournie lors de la soumission pour notifier l'acceptation ou le refus. Environ 60% des sites sont acceptés.

Les sites "Grand Public" sont classés et pris en compte, en règle générale, en une semaine. La grande préoccupation actuelle vient des sites d'entreprise qui sont encore très peu présents sur l'annuaire, car une arborescence spécifique, qui leur sera allouée, est en train d'être mise en place.

Si vous avez soumis un site appartenant à une entreprise, sachez donc que toutes les soumissions sont "gardées au chaud" en attendant la mise en ligne de cette branche de l'arborescence, d'ici 2 à 3 mois (logiquement, début 2001).

Ceci dit, si le site que vous avez soumis n'est pas présent dans l'annuaire lorsque la branche "Entreprises" sera en ligne, resoumettez-le pour être sûr qu'il soit bien pris en compte à cette époque-là.

En général, l'arborescence est toujours en effervescence (jeunesse oblige), ce qui explique que certains webmestres ont bien du mal à trouver une catégorie qui corresponde à leur activité. Patience, cela ne saurait tarder. Rappelons que l'annuaire est encore très jeune (quelques mois à peine) ! Si vous avez à soumettre un site sur l'annuaire et que vous ne trouvez pas de rubrique qui colle au mieux à la source d'information en question, tentez de trouver une catégorie la plus proche possible du thème désiré, les netsurfeurs garderont certainement votre site "au chaud" en attendant que la rubrique en question soit créée.

En effet, ils inscriront peut-être même votre site dans l'annuaire sans que vous le sachiez, car l'outil de gestion permet de travailler à la fois sur des parties de l'arborescence "publiques" (disponibles en ligne sur le site lycos.fr), mais également sur des parties

"cachées", c'est-à-dire en train d'être construites. Dès qu'une branche est considérée comme étant assez "remplie", elle est mise en ligne automatiquement.

Les sites peuvent être refusés pour les (classiques) motifs suivants :

- Site pas en français ;
- Lien fourni mort ;
- Fotes d'ortograf ;
- Sites "plaquettes" de quelques pages sans contenu réel ;
- Sites pas obligatoirement très "chaud", mais proposant des liens vers des sites X qui, eux, le sont, et demandant à être inscrit dans une rubrique non-X de l'annuaire ;
- Sites en construction ;
- Sites non mis à jour ;
- Activités illégales, sectes, racisme, piratage informatique, et autres joyusetés.

Ce sont les principaux critères de refus. Il peut y en avoir d'autres (voir nos articles précédents sur Yahoo France, Nomade ou le Guide de Voila, les motifs de refus sont toujours quasi identiques d'un annuaire à l'autre).

L'immense majorité des résumés fournis lors de la soumission sont ré-écrits car pas assez descriptifs et trop commerciaux, marketing ("Notre site propose...", "Nous sommes les leaders..").

Il est également possible, une fois que le site est inscrit, de demander une modification de la fiche descriptive du site. Pour cela, il suffit d'envoyer un e-mail au surfeur en charge de la catégorie dans laquelle votre site est présent. La liste et les coordonnées des surfeurs se trouvent à l'adresse <http://www.lycos.fr/service/contact.html> et le nom des personnes est également indiqué dans toutes les catégories de l'annuaire à droite du pavé jaune "Sites web".

Cependant, cette modification ne sera prise en compte que s'il existe une erreur dans le descriptif proposé en ligne ou s'il s'agit de rendre le résumé plus précis. Si la demande vise à rajouter des mots clés importants de façon "innocente", elle ne sera pas prise en compte. Rappelez-vous que les netsurfeurs des annuaires ne sont pas tombés de la dernière pluie...

Les délais de prise en compte d'un site sont d'environ une semaine. La base de données de l'annuaire est remise à jour 2 fois par mois. Dans le pire des cas, un site accepté doit donc attendre 3 semaines avant d'apparaître dans l'annuaire. Dans le meilleur des cas, une semaine environ (l'inscription aura été faite juste avant la mise à jour).

Chaque netsurfeur a comme objectif interne de rentrer dans la base un minimum de 35 sites par jour (qu'ils soient dans l'arborescence "publique" ou "cachée" de l'annuaire). Une liste d'avancement des travaux sur l'annuaire indique les priorités, donc les rubriques et/ou les sites à traiter au plus vite. Pour l'instant, aucune "liste noire" n'est utilisée par l'annuaire de Lycos France pour y classer des sites ayant quelque peu exagéré dans leurs soumissions "pirates". Les netsurfs disent ne pas être énormément gênés par des procédures de spam pour l'instant. Enfin, soulignons que ceux-ci alimentent également la rubrique "Coups de Coeurs" de Lycos qui est elle aussi organisée par catégorie et qui recense en même temps les sites Nouveautés ou les sites insolites.

Autre point important : les netsurfeurs n'ont pas la possibilité de rentrer, comme sur Nomade ou le Guide de Voila, des mots clés de façon complémentaires sur une fiche descriptive de site. En revanche, comme sur Yahoo!, cette possibilité existe sur les noms de rubrique, ce qui permet de générer des réponses sous forme de catégories à de nombreuses requêtes.

Dernier point : Lycos vient de mettre en ligne une nouvelle version de sa chaîne "shopping" (<http://www.fr.lycos.de/webguides/shopping/liste/>) qui comprend un annuaire de sites de commerce électronique rangés par catégories (autour de 1500 sites, 19 catégories principales, plus de 100 sous-catégories). Cet annuaire est promu depuis la chaîne "shopping" et depuis les différentes rubriques de l'annuaire (d'ici la fin de la semaine 46). Les webmasters peuvent soumettre directement leur site pour être référencé dans cet annuaire (lien "Ajouter votre site"). La procédure consiste à envoyer un mail à l'adresse shopping@lycos.fr, avec comme sujet "Ajout Guide Shopping", en précisant les informations suivantes :

- Adresse de votre site
- Catégorie dans laquelle vous souhaitez apparaître
- Type de commande (en ligne, par courrier, fax, téléphone...)
- Modes et types de paiement acceptés par votre site (carte, chèque / en ligne, par courrier, fax...)
- Type de sécurisation (SSL ou autres)
- Lieux de livraisons (Région, France, UE, Europe, Monde...)
- Langues proposées pour la consultation du site (En plus du Français)

Votre demande sera ensuite étudiée par l'équipe Lycos. Les sites sélectionnés seront ajoutés sous 10 jours en moyenne. Pour faire enlever votre site, ou demander la modification de sa description, contactez Lycos de la même façon (shopping@lycos.fr), en précisant dans le sujet "Modification Guide Shopping". Voir : <http://www.fr.lycos.de/webguides/shopping/liste/ajouter/index.html>

"Recommandé par Lycos"

Lorsque l'on parcourt l'annuaire en descendant l'arborescence (en cliquant sur les liens, de catégories en catégories), certains sites sont accompagnés d'un petit labrador noir qui remue la queue, avec la mention "Recommandé par Lycos". Il n'est pas possible de payer pour obtenir ce satisfecit ;-) qui est fourni par les netsurfers, en leur âme et conscience (si, si, un netsurfeur a une âme et une conscience... :-)).

Nous sommes cependant en mesure de vous donner un truc infallible pour obtenir le pictogramme "Recommandé par Lycos" : il suffit... d'avoir un bon site, source d'information incontournable, portail thématique, site de référence dans le domaine traité. Et le tour est joué ;-) Et je ne dis pas ça parce qu'Abondance l'a obtenu !!!

Cependant, le fait d'avoir "gagné" ce pictogramme ne signifie pas que vous l'ayiez ad vitam aeternam. Le netsurfeur peut prendre la décision de vous l'enlever, pour l'attribuer à un nouveau site qui lui semble plus pertinent.

Cela nous donne l'opportunité d'explicitier les résultats de l'outil de recherche Lycos France lorsqu'on saisit un mot clé sur sa page d'accueil (nota : une immense majorité des utilisateurs de l'outil ne tapent qu'un seul mot clé).

Lorsqu'on saisit un ou plusieurs mot(s) clé(s), les informations suivantes sont donc affichées, dans l'ordre :

- Tout d'abord les catégories (3 au maximum) qui contiennent le mot demandé dans leur énoncé ou les mots clés qui y sont rattachés. Pour avoir accès aux catégories suivantes (si il y en a plus de 5), il faut cliquer sur le lien "Plus de pages web sur...", ce qui n'est pas très logique, il faut bien le dire, puisqu'il s'agit de rubriques et non de pages...

- Puis, trois sites (maximum) issus de l'annuaire sont présentés. La dénomination "Recommandé par Lycos", dans ce cas, n'est plus en relation avec celle qui avait été donnée au sein de la rubrique par choix éditorial du netsurfeur (voir précédemment). Il s'agit, ici, d'évaluer une bonne adéquation entre la question posée et la fréquence d'apparition de ce terme dans le titre, le résumé et l'URL du site. Si par exemple, le mot demandé se trouve dans les trois champs, le site recevra obligatoirement la dénomination "Recommandé par Lycos". Un site donné peut donc obtenir ce pictogramme à l'intérieur d'une catégorie de l'annuaire, mais pas lors de la saisie d'un mot clé, dans la page de résultats. Et vice-versa.

- Ensuite viennent 5 résultats fournis par Direct Hit, système d'indice de clic (voir <http://www.abondance.com/trucs-et-astuces/outils4.html>).

- La page se termine avec 3 résultats issus du moteur de recherche (en ce moment, c'est encore le "vieux" index de Lycos, mais le passage à Fast/Alltheweb est imminent, nous vous en reparlerons très bientôt). Lorsque Fast sera en place, l'objectif sera d'obtenir une mise à jour de l'index tous les trois jours. Alleluia !!

Les liens en rouge, proposés en début de page (appelés "kittys", exemple : "Retrouvez tous les livres sur "... avec BOL"), sont achetés par des partenaires. Ils ne sont pas, bien entendu, issus de l'annuaire ou du moteur. Cela va sans dire mais bien mieux en le disant...

Ce descriptif est valable pour la première page de résultats. A partir de la seconde, on trouve :

- D'abord 5 résultats de l'annuaire, si ils existent.

- Ensuite, le complément modulo 10 (5 résultats si 5 sites "annuaires" sont présentés précédemment) en pages web issues de l'index.

Les sites de l'annuaire étant toujours proposés en premier, les pages issues du moteurs sont donc très peu visibles (et avec ça on nous dira qu'on peut faire du positionnement sur Lycos... ;-))

Conclusion

L'annuaire de Lycos est un outil encore jeune, puisqu'il n'a que quelques mois d'existence. Mais il appartient sans conteste au groupe restreint des annuaires majeurs dans lesquels il faut être référencé, avec Yahoo! France, Nomade, le Guide de Voila et l'Open Directory francophone.

Petit point de détail, mais très sympa (enfin, je trouve...), il est le seul à présenter les netsurfeurs, photos à l'appui, avec leurs vrais prénoms (au début, c'étaient des pseudos qui étaient indiqués), dans chaque branche de l'arborescence. Cela donne un visage humain à l'outil qui nous le rend plus sympathique.

■ Au coeur d'Altavista France

Nota : merci à l'équipe d'Altavista France pour avoir pour avoir donné, annoté, validé ces informations et corrigé les erreurs éventuelles qui s'étaient glissées dans l'article, ainsi que pour avoir forcé quelques portes californiennes pour les obtenir ;-). Le but de cet article est de mieux comprendre le fonctionnement du moteur Altavista France, près de 9 mois après son lancement, le 15 février dernier et avant de nouvelles fonctionnalités qui vont arriver d'ici quelques mois (nous aurons l'occasion de vous en reparler...).

"Web français" ou "en français"

Tout d'abord, de nombreuses personnes se sont posé la question de la différence entre le menu déroulant "en français" et le bouton radio "Web français", sur la page d'accueil d'Altavista France. La question a été posée aux personnes rencontrées chez Altavista France. Leur réponse fait l'objet d'un article sur le site Abondance, à l'adresse : <http://www.abondance.com/trucs-et-astuces/recherche19.html>

Merci de le lire avec attention avant de parcourir l'article ci-dessous, car il y sera fait allusion aux - subtiles - différences entre "Web français" et "Tout le Web".

Ceci dit, commençons notre exploration du moteur...

Organisation interne

Les serveurs d'Altavista France se trouvent à Londres même. La banlieue de Londres (Maidenhead) héberge les services de développement, d'ingénierie, d'administration, de back-office. Quand on sait que les serveurs de Lycos France sont en Allemagne et ceux de Yahoo! France en Suède, on se dit que l'Europe est un bien grand continent ;-). En fait, la Grande-Bretagne héberge tous les serveurs européens d'Altavista, ainsi qu'une grande partie des services de développement informatique pour tout le continent.

L'organisation des index du moteur est la suivante :

- Le site Altavista.com utilise un index situé en Californie, à Palo Alto, siège d'Altavista.
- Le site Altavista.fr utilise un index basé à Londres, qui lui est spécifique, lorsqu'on prend en compte le bouton radio "Web français". Il reprend l'index de Palo Alto lorsqu'on choisit le bouton radio "Tout le Web".

Un spider spécifique à chaque pays est utilisé pour compléter et mettre à jour les index régionaux.

En fait, pour bien comprendre la façon dont fonctionne Altavista (mais également les prochains développements, qui seront en ligne en début d'année prochaine), il faut bien avoir à l'idée qu'il existe plusieurs familles de sites chez eux :

- Le site américain (www.altavista.com), qui possède son histoire propre, ses accords passés, ses propres fonctionnalités, depuis le jour de sa naissance en décembre 95.
- Les sites hors des Etats-Unis, beaucoup plus récents, donc plus "libres" au niveau des accords passés, plus autonomes, qui se différencient eux-mêmes en deux familles :

* Ceux qui sont sortis avec le nouveau "look" (Irlande, Espagne, Danemark, etc.), tous très semblables :

<http://ie.altavista.com/>
<http://es.altavista.com/>
<http://dk.altavista.com/>

* Ceux qui sont sortis avant que le nouveau "look" n'apparaisse (France, UK), mais qui vont rapidement y venir :

<http://fr.altavista.com/>
<http://uk.altavista.com/>

D'ici quelques mois, tous les sites non-américains, et notamment les européens,

s'uniformiseront au niveau de leur interface utilisateur. Le site américain, de son côté, devrait garder son apparence actuelle (différente, donc, des sites non-américains). Vous suivez ? Ca va ???

Bien, on continue...

L'index français spécifique compte actuellement 13 millions de pages en langue d'origine française.

Critères de pertinence

La plupart des informations sur l'algorithme de recherche sont gardées secrètes à Palo Alto avec à peu près autant d'acharnement - voire de paranoïa - que le secret de fabrication du Coca Cola à Atlanta. Difficile donc, même pour les équipes d'Altavista Europe, d'obtenir énormément d'informations. Mais, a priori, voici quelques données qui semblent intéressantes :

- La présence du mot demandé dans le titre des pages (balise <TITLE>) est toujours le critère de pertinence numéro 1. De gros efforts sont faits par les équipes en charge du développement des algorithmes de classement pour valoriser le texte lui-même plus que tout le reste. A l'heure actuelle, le titre reste encore prioritaire, mais dans les mois qui viennent apparaîtront de plus en plus des critères qualitatifs au niveau du texte des pages web : variété et richesse lexicale, niveaux de langues, correction syntaxique etc.

- En 2, la présence du mot demandé dans le texte visible de la page, avec les critères plus spécifiques suivants :

* L'occurrence des mots est prise en compte, mais il ya de fortes chances que les deux premières occurrences rencontrées dans la page soient effectivement les seules à être prises en considération dans le calcul de pertinence. Le fait que les balises Meta soient en début de page et que leur importance décroît fortement pour Altavista fait que le risque de voir une page contenant des balises Meta être mal classée sur ce moteur est de plus en plus grand. Voir l'article sur ce point précis à l'adresse http://abonnes.abondance.com/articles/meta_2.html).

* La proximité des mots dans la page (si plusieurs termes sont demandés).

* La rareté des mots (si on fait une requête sur "Michel Duchaussoy", les pages qui contiennent "Duchaussoy" seront mieux classées que celles qui contiennent "Michel", mot beaucoup plus courant dans les pages francophones.

* L'emplacement : un mot en début de page a plus de poids qu'un mot en fin de phrase. La première phrase est la plus importante.

* La mise en exergue : les mots en gras, de grande taille (balise Font Size), ont une importance quasiment infime. Aucune importance même si l'on en croit les ingénieurs londoniens. En tout cas, cela ne semble pas fortement jouer dans le cadre d'un classement concurrentiel.

- Les balises Meta sont en chute libre au niveau du poids accordé, largement au profit de l'indice de popularité, devenu le troisième critère par ordre d'importance décroissante, après le titre et le texte visible.

- La présence du mot clé dans l'URL et notamment dans le nom de domaine est importante. Cela peut donner un "coup de fouet" au classement d'une page.

- Le poids octroyé à chaque critère de pertinence est modifié au minimum plusieurs fois par semaine, à la fois pour faire "tourner l'index", ne pas avoir les mêmes pages toujours aux mêmes places, et pour éviter le spam.

- Les pages dynamiques sont prises en compte (sauf les pages PHP qui semblent poser de vrais problèmes d'indexation) s'il n'y a pas de point d'interrogation dans les URLs.

- Le Flash n'est pas pris en considération.

- Les commentaires ne sont pas lus.

- Les CGI en page d'accueil (détection de plug-in par exemple) ne semblent pas poser problème.

- Frames : elles sont prises en compte sur un premier niveau (page mère + pages filles). Si les pages filles contiennent elles-même des frames, il ya fort peu de chance qu'elles soient prises en compte.

- Les liens Javascript ne sont, en général, pas suivis dans les codes HTML.

- Autre critère, moins connu, et même si son importance est relativement faible dans l'algorithme de classement : le nombre de liens externes (vers des sites distants) de la page, qui donne une image, selon Altavista, de la "richesse d'ouverture" du site. Les liens en provenance et à destination d'autres sites mesurent tous deux l'indice d'interconnectivité du site. Les premiers sont plus valorisés dans la note finale que les seconds, mais ceux-là comptent néanmoins quelque peu. Ces deux critères sont issus de la célèbre "théorie du noeud papillon" chère à Altavista.

- La grande nouveauté est en fait l'implication prochaine d'algorithmes complexes d'ordre linguistique, grammaticaux et syntaxiques, pour "booster" les pages desquelles on peut faire ressortir un "sens". Difficile de donner plus d'informations pour l'instant (les systèmes sont encore en cours de développement), mais la voie suivie est passionnante : si

l'internaute tape un mot clé, le but sera de faire ressortir non seulement les pages qui contiennent, ici ou là, le terme demandé, mais surtout celles dont l'objet global peut être représenté par le mot saisi. Le "sens général" de la page peut être identifié par ce terme. Encore quelques soirées de travail en perspective pour les référenceurs ;-)

Le moteur n'a certes pas la prétention de véritablement comprendre le sens ultime d'une page, mais il peut analyser un certain degré de qualité du contenu et partant de là, effectuer un classement plus intéressant. L'objectif est de toujours d'améliorer la qualité des sites en tête de liste et, pourquoi pas, d'inciter les développeurs à passer plus de temps sur le contenu que sur les balises !!

Lutte contre le spam

Altavista indique, comme Inktomi d'ailleurs qui cite les mêmes chiffres, que 95% des pages qui lui sont soumises au travers de son formulaire de soumission directe, peuvent être considérées comme du spam, avec en priorité les pages contenant des scripts qui génèrent automatiquement d'autres pages (tous les moteurs font état des mêmes problèmes) ainsi que les pages contenant des mots répétés ad vitam aeternam. Y a quand même des webmasters qui sont lourds, des fois ;-)... Altavista gère une liste noire, dans laquelle sont compilés tous les noms de domaine "à problème".

Il existe de moulinettes informatiques, chez Altavista, qui tentent de détecter les tentatives de spam. Lorsqu'un doute émerge, la page est sortie de l'index et traitée par un être humain qui regarde son code HTML pour tenter de détecter d'éventuelles fraudes.

Altavista travaille énormément également sur la détection des techniques de cloaking (aisées à détecter pour un moteur en jouant sur les adresses IP de ses spiders). Attention donc à l'avenir si vous jouez avec ce type de pratiques...

Référencement sur Altavista France

Lors de la soumission d'une URL sur Altavista France (http://fr.altavista.com/help/remove_a_page.jsp), certains noms de domaine sont refusés en indiquant le message "AltaVista.fr accepte uniquement les domaines en .fr, .com, .net and .org". Par exemple, les sites Guadeloupéens (en .gp), comme <http://www.delgres.gp/>, sont refusés. Eh, Ti Mal, ou pa ka conait Gwadeloup ? (pour ceux qui parlent Créole uniquement ;-))

Les sites Belges (.be), Suisses (.ch) ou autres sont également refusés. Pour ceux-là, on peut penser que la raison est que des versions nationales du site Altavista dans ces pays seront bientôt disponibles. Les sites belges seront alors indexés sur Altavista Belgique, et idem pour la version suisse. L'index d'Altavista France ne prenant en compte que les pages d'origine française pour garnir son index accessible par le bouton radio "Web français", c'est finalement assez logique. Mais je devrais avoir plus d'informations sur les "filtres noms de domaine" lors d'une soumission sur Altavista France pour la lettre "R&R" du mois prochain. Les fins limiers sont sur des pistes... Encore un peu de patience. :-)

D'autre part, Altavista France a clairement connu, dernièrement, de gros problèmes d'indexation de pages, notamment sur les soumissions réalisées depuis sa fonction "Add URL". Si votre site n'a pas été indexé sur Altavista France dans les mois qui viennent de s'écouler, vous pouvez envoyer un mail à l'adresse webmaster@altavista.fr. Cette adresse arrive directement chez Altavista France et chaque cas est étudié manuellement par leurs équipes techniques qui trouvent une solution dans l'immense majorité des cas. N'hésitez donc pas à envoyer un message (sans spammer) en indiquant l'URL de votre site dans l'e-mail et, en deux lignes, à expliquer votre cas. Tout devrait se résoudre par la suite en 3 semaines, un mois (oui, je sais, c'est un peu long, mais c'est du traitement manuel)...

Sachez donc que l'équipe technique d'Altavista est consciente des problèmes rencontrés dernièrement, et il m'a semblé sentir souffler comme un vent de sincérité lorsqu'ils m'ont expliqué qu'ils faisaient tout pour réparer ces problèmes... ;-)

Le but, à partir du début de l'année prochaine, sera de régler tous ces problèmes, inhérents à des migrations d'index et à des travaux sur ces mêmes index, dûs au lancement des différentes versions européennes du moteur. Il est prévu de passer rapidement à un délai de 15 jours entre deux mises à jour de l'index. Le spider français intégrant 2 millions de pages par jour et l'index faisant 13 millions de pages, on devrait y arriver ;-)

L'objectif final, d'ici quelques mois, serait même d'avoir un index francophone rafraîchi toutes les semaines.

Conclusion

Altavista France est un moteur de recherche qui termine sa phase de maturité, après un lancement quelque peu cahotique en février dernier. L'année 2001 devrait voir s'ouvrir une nouvelle ère pour lui : réorganisation des index, nombreuses nouvelles fonctionnalités de recherche, nouveaux algorithmes de classement, possibilité de personnaliser ses recherches et ses pages de résultats (comme sur le site américain), etc. Bref, il y aura de quoi dire très bientôt dans cette lettre sur les prochaines nouveautés d'Altavista France !

■ Nouveau service de recherche pour MSN Search France

Nota : merci à l'équipe de MSN Search France pour avoir donné, annoté, validé ces informations et corrigé les erreurs éventuelles qui s'étaient glissées dans l'article.

MSN.fr a lancé, il y a quelques semaines de cela, un nouveau service de recherche, MSN Search (<http://search.msn.fr/>), qui se veut "le plus astucieux du Web". Il est basé sur un annuaire (proposant 62 000 sites, il est mis en oeuvre par la société Lamine), un moteur de recherche (fourni par Inktomi), la possibilité de poser une question par e-mail (fonction "Nous cherchons, vous trouvez !", également gérée par Lamine, nécessitant pour l'internaute d'avoir une adresse électronique chez Hotmail). Un forum de discussion entre internautes pour s'échanger les "bons plans" de recherche complète l'outil. Le but est de répondre de façon pertinente à 80% des recherches exprimées (la plupart du temps sous la forme d'un ou de deux mots clés) par les internautes.

La fonction "Nous cherchons, vous trouvez", notamment, est un système de questions/réponses avec des experts de la recherche d'information. Lorsqu'on pose une question, un premier mail de confirmation de réception de la demande est envoyé. Puis, 24 à 36 heures plus tard, un deuxième mail est envoyé, contenant des propositions de ressources pertinentes par rapport à la question posée.

MSN France s'attend à répondre aux interrogations des internautes selon les voies suivantes : 50% par la saisie de mots clés, 30% par l'exploration de l'annuaire en descendant l'arborescence (en suivant les liens), 10% par l'intermédiaire du système de questions-réponses avec les experts, et 10% par le biais de la communauté.

Affichage des résultats

Lorsqu'on tape, donc, un ou plusieurs mots clés sur le formulaire de recherche de la page d'accueil, les pages de résultat sont formatées de la façon suivante :

- Tout d'abord, le Realnames est présenté, s'il existe. Point intéressant, il est représenté, pour la première fois sur un outil de recherche, comme un résultat à part entière, et non comme un choix issu d'un partenariat. Le choix de MSN donne un poids plus fort à cette fonctionnalité de Realnames et renforce l'idée qu'il vous faut réserver le vôtre.
- Puis, 4 liens au maximum sont indiqués, présentant des sites web issus de l'annuaire ou provenant de choix spécifiques de l'équipe de l'outil de recherche de MSN. Dans cette zone apparaîtront notamment de nombreux sites de la galaxie Microsoft, en relation avec les requêtes demandées, bien évidemment.
- Puis sont indiquées un maximum de 4 catégories en provenance du guide web indiquant le contexte des catégories proposées.
- Les sites en provenance de l'annuaire classés hiérarchiquement en fonction de la qualité des sites par rapport à la requête demandée sont ensuite proposés. Les critères de pertinence sont classiques : titre, résumé, URL. Des mots clés additionnels peuvent être rajoutés de façon transparente sur une fiche descriptive de site ou un énoncé de rubrique (catégorie).

Ce descriptif de la page de résultats est valable pour la première page fournie lorsqu'on tape un ou plusieurs mot(s) clé(s). Nota : un partenaire de MSN peut acheter un positionnement dans les résultats de l'outil de recherche, mais uniquement si celui-ci est logique (le site marchand de la Fnac pour la requête "achat livre", par exemple).

A partir de la deuxième page de résultats, les indications suivantes seront affichées :

- En premier, les autres sites de l'annuaire, s'ils existent (10 au maximum).
- Puis, les pages issues de l'index d'Inktomi (15 au minimum).

A priori, c'est de l'index européen d'Inktomi (basé chez BT, à Londres) que proviennent les résultats.

L'outil de recherche a essayé de faire un gros travail de fond pour reconnaître des mots mal orthographiés ("pokmon" sera synonyme de "pokémon") ou de sens proche (dans la

même veine, "pikachu" sera synonyme de "pokémon").

Le but de MSN France est de proposer un outil simple et intuitif, sans fioritures (pas de recherche multimédia, de photos, de cours de bourse, de vidéos, etc.). Simple et efficace semble être son credo.

Enfin, le site propose une recherche avancée, mais elle ne porte que sur les résultats du moteur Inktomi. Elle est très peu utilisée.

Référencement sur l'annuaire

Au niveau du référencement sur l'annuaire (géré par Lamine mais appartenant clairement à MSN), 50% des sites sont référencés sur la base d'une proposition (le formulaire se trouve à l'adresse <http://www.msn.fr/leguide/suggestions.asp>). Le reste est effectué par le biais d'un travail de veille perpétuelle de chaque netsurfeur. Chaque site soumis est traité (évalué/accepté/refusé) en une semaine environ. Si un site est refusé, aucun e-mail n'est envoyé au webmestre qui l'a soumis. Il faut resoumettre pour avoir une chance d'être à nouveau évalué. Une demande de modification de la fiche descriptive du site est, elle, prise en compte (si justifiée) dans les 24 à 48 heures. Les critères de sélection se trouve a l'adresse <http://www.msn.fr/help/aideguide.asp>

L'annuaire reçoit de 500 à 1 000 soumissions par jour. La moitié émanent de logiciels de référencement automatiques. Elles sont tout aussi automatiquement refusées.

20 personnes travaillent actuellement (équipes de Lamine comprises) sur le portail de recherche de MSN.

Conclusion

Après avoir avoir un peu somnolé les années précédentes, l'outil de recherche de MSN France semble être sorti de sa léthargie et propose un vrai outil qui semble pertinent, bien né, et qui combine surtout, pour la première fois, plusieurs voies distinctes : un annuaire et un moteur (mais, là, c'est très classique), un site de questions/réponses et un environnement communautaire pour s'échanger les "bons plans" de recherche d'infos. Bref, un ensemble cohérent qui semble tenir la route. Il pourrait donc devenir l'un des grands acteurs de la recherche d'information dans les mois qui viennent. Une affaire à suivre...

■ Hotbot France arrive !

La version française de Hotbot devrait être disponible en ligne d'ici peu en version beta. La version finale sera mise en fonction début 2001. Il s'agira donc d'une adaptation française du site hotbot.com, le moteur de recherche devant proposer beaucoup plus de filtres disponibles et une vraie recherche multimédia (recherche par auteur, par formats audio, etc.). L'index sera bien entendu fourni par Inktomi. Sans que cela soit officiel, on peut raisonnablement penser que le moteur "attaquera" l'index européen d'Inktomi, situé à Londres, chez BT.

L'objectif de la version européenne d'Hotbot est de répondre aux besoins européens et non pas d'être une pale copie ou adaptation du Hotbot US. Lycos Europe (rappelons qu'Hotbot fait partie du réseau Lycos) a travaillé plusieurs mois sur des études internes pour savoir ce que les utilisateurs européens dans les différents pays attendent d'un nouvel outil de recherche et pour comprendre comment ils appréhendent aujourd'hui la recherche sur l'Internet. Ils ont également testé plusieurs designs, fonctionnalités, etc.

Hotbot ne se veut pas le moteur de recherche des spécialistes, contrairement à son image américaine. Le but de Hotbot est de permettre à tout utilisateur de prendre le contrôle de sa recherche et d'apprécier au maximum l'expérience de recherche sur l'Internet en effectuant des recherches très précises et par le biais d'un design et d'une structure très claires, d'options approfondies, d'aides détaillées.

Hotbot sera destiné à tous les publics mais il se focalisera sur l'aspect recherche et ne sera pas un portail multi-fonction généraliste. Bienvenue au nouvel outil de recherche !

■ Délais actuels des moteurs de recherche entre une soumission manuelle et la prise en compte effective d'une page

Au tableau d'honneur : les moteurs qui indexent de nouveaux sites assez rapidement (moins de 15 jours) : Altavista US et France.

Les moteurs qui sont "moyens" (certains sites indexés de façon plus ou moins rapide, d'autres non pris en compte, bref des délais assez variables mais qui restent raisonnables) : Google, Ecila, Northern Light, Excite US.

Les moteurs à la peine ce mois-ci : Infoseek, Voila, Excite (pour les sites en français), Hotbot (donc Inktomi), Alltheweb, Lycos, Lycos France.

■ Bruits et chuchotements

Une rubrique qui regroupe tous les bruits et rumeurs dans le (petit) monde des outils de recherche mondiaux et francophones. Rien n'est obligatoirement vérifié, mais toutes les infos sont données... de source sûre ;-)

-> Contrairement à ce qui a été évoqué dans cette lettre les mois précédents, Excite ne changerait pas de fournisseur pour son index de recherche prochainement. Mais il pourrait y avoir de nombreux changements sur le moteur très prochainement et notamment de gros progrès - très attendus - sur la mise à jour de son index. Il se pourrait qu'une fonction de recherche multimédia (son/audio/vidéo) fasse son apparition sur Excite France, en collaboration avec Inktomi.

-> L'annuaire Blanc de l'Ile des Medias (Excite.fr, Altavista.fr, Lokace, entre autres) comporterait environ 70 000 sites.

-> Après avoir changé son annuaire réalisé en interne (Hachette.net, qui contenait 12 300 sites) par Looksmart France, Club-Internet réfléchirait maintenant au remplacement d'Inktomi par un autre prestataire. Mais rien n'est fait pour l'instant.

-> Il se murmure que le site Question.fr (<http://www.Question.fr/>) serait en train de mettre en place un système d'outil de recherche humain, avec réponses d'experts, en marque blanche. Cet outil, appelé "ExpertLink", ayant nécessité 6 mois de développement, serait notamment proposé aux grands portails et outils de recherche francophones en complément de leurs fonctionnalités habituelles. Un peu ce que fait Lamine sur MSN Search avec l'option "Nous cherchons, vous trouvez". Plus d'infos sur cette offre de Question.fr sera disponible sur Abondance dans quelques jours.

-> Il se pourrait que sur certains moteurs, et notamment Google, le fait que le spider ait du mal à indexer un page (ralentissement ou autre) joue également sur son positionnement. Un critère de pertinence comme un autre...

-> La version actuelle de Looksmart France générerait mal les accents. Une nouvelle version, réglant ce problème, est attendue pour la mi-décembre.

-> Lycos, le labrador noir, existe bel et bien : il se promène dans les locaux de l'outil de recherche et bouffe les pulls de tout le monde (on en trouve, de l'info utile, dans la lettre payante d'Abondance ;-))

■ En Bref

-> Sur Yahoo.com, dans de très nombreuses rubriques, les sites proposés en "Most popular searches" sont classés selon l'indice de popularité fourni par Google.

-> Sur Yahoo.fr, un site à objectif commercial n'aura quasiment aucune chance d'obtenir la "paire de lunettes" qui permet d'être classé en premier dans certaines rubriques. Cette distinction est, dans une immense majorité des cas, réservée aux sites à but non commercial.

-> Looksmart France compte 23 éditeurs rentrant chacun 40 sites par jour. L'annuaire compte à l'heure actuelle 80 000 sites. Pour le moment il n'est pas possible de soumettre des sites à ses rédacteurs/surfeurs, ni sur Worldonline France, ni sur Club-Internet, qui sont les deux portails utilisant son annuaire. Looksmart fera une conférence de presse dans la deuxième semaine de janvier 2001 pour annoncer officiellement son offre.

-> Un petit scoop pour Abondance : vous avez certainement vu que, dans chaque rubrique de l'annuaire de Lycos.fr, un lien est disponible vers la catégorie correspondante de l'annuaire de Lycos US. Très bientôt (d'ici la fin de l'année), il sera possible d'obtenir les sites américains immédiatement traduits en français, au moyen d'un système de traduction automatique (ce serait l'outil Reverso qui serait pris en compte).

-> Yahoo.fr semble avoir donné un "coup de collier" au référencement de sites pendant le mois qui vient de s'écouler. De nombreux sites ont été intégrés dans l'annuaire en octobre/novembre. Nomade.fr est recordman du délai d'épreuve en compte : moins de 15 jours pour voir son site intégré dans la base. Bravo !

-> L'annuaire Indexa (<http://www.indexa.fr>) a passé la barre des 25 000 sites référencés.