

■ L'intégration de Fast dans Lycos

Depuis quelques semaines, Lycos.fr intègre (en cela, il suit le site américain qui l'avait déjà fait depuis quelques mois) les résultats de FAST dans son moteur de recherche. C'était une bonne occasion de faire le point sur la façon dont est géré ce partenariat, notamment avec Felicitas Piegda (merci à elle), qui s'est occupée, entre autres projets, de cette intégration. Voici les informations qu'elle nous a fournies concernant Fast et sa mise en place dans l'outil de recherche Lycos...

Les serveurs de Lycos Europe (tous sites confondus pour ce continent) se situent en Allemagne, à Gütersloh, siège de Lycos Europe. L'outil de recherche Lycos fait appel aux résultats de Fast pour sa partie "moteur de recherche en texte intégral" (rappelons que l'annuaire de Lycos France est géré en interne, à Paris, sous la houlette de Florence Lassalle). Fast utilise en fait plusieurs index différents selon les sites qu'il dessert :

- Un index européen, basé à Londres, disponible depuis janvier dernier. Il est utilisé à l'heure actuelle par les sites européens de Lycos (dont Lycos France), mais peuvent également être utilisés par d'autres clients de Fast (comme Spray ou Web.de). Il n'est donc pas "exclusif Lycos".
- Un index aux Etats-Unis pour le site Alltheweb, qui est un peu la "vitrine technologique" de Fast. Alltheweb est également considéré comme un "laboratoire" par Fast pour intégrer de nouvelles fonctionnalités dans son moteur de recherche.
- Deux index aux Etats-Unis, qui sont des "back-up" l'un de l'autre (ils sont donc quasi identiques à un temps T) pour ses clients américains, et notamment Lycos US. Ces index sont basés à Boston (Massachusetts) et en Californie. Ils ne sont pas, non plus, exclusivement dédiés à Lycos.

L'index américain (et son clone) utilisé par Lycos US peut donc être légèrement différent de celui utilisé par Lycos France.

Lycos effectue également, pour gagner en temps de réponse, des systèmes de cache sur ses serveurs en Allemagne, pour les requêtes les plus demandées.

L'index de Fast compte environ 560 millions de pages dont 18 millions en français. Ce dernier chiffre devrait croître assez rapidement dans les mois qui viennent. Il faut noter que Fast est un outil de recherche assez jeune. Son premier objectif, en plus de la pertinence des réponses, bien sûr, a été de faire grandir cet index le plus rapidement possible pour arriver à une taille "critique" supérieure à 500 millions de pages afin de se retrouver au niveau des Google, Altavista et autres Inktomi. Cet objectif, couplé à la recherche de la meilleure pertinence pour ses résultats (donc l'objectif d'un algorithme de classement perfectionné), a fait que le moteur a pu être moins "regardant" sur d'autres projets, comme le spam ou les pages dynamiques, qui vont être pris en compte plus sérieusement dans les mois qui viennent, maintenant que l'index est à peu près stabilisé.

Affichage des résultats sur Lycos France

Cela nous donne l'opportunité de refaire un point sur les résultats de l'outil de recherche Lycos France lorsqu'on saisit un mot clé sur sa page d'accueil.

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. Ces catégories sont assorties d'un célèbre "chien d'or" qui bat la queue... Pour avoir accès aux catégories suivantes (si il y en a plus de 3), 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. Leur apparition se fait en fonction de la fréquence d'apparition du ou des termes de recherche dans le titre, l'URL et le commentaire. La dénomination "Recommandé par Lycos" avec le "chien d'or" a été abandonnée sur l'affichage de ces sites, car cela créait une confusion avec les pictogrammes identiques proposés sur les sites à l'intérieur des catégories (choix subjectifs des surfeurs de l'annuaire).
- Auparavant, 5 liens émanant de DirectHit étaient affichés. Ces infos ont disparu, Lycos estimant que les résultats provenant de Fast étaient assez pertinents et faisaient double emploi avec ceux de DirectHit.
- La page se termine donc avec 7 résultats issus du moteur de recherche de Fast.

Les liens en rouge, proposés en début de page (appelés "kittys"), sont soit achetés par des partenaires (exemple : "Retrouvez tous les livres sur "... avec BOL"), soit destinés à la promotion des chaînes et services du portail ("Rencontrez l'âme soeur sur Love@Lycos"). 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, s'ils existent.
- Ensuite, le complément modulo 10 (5 résultats si 5 sites "annuaires" sont présentés précédemment) en pages web issues de l'index de Fast.

Les pages de résultats issues de Fast sont affichées "telles quelles" pour l'instant. La requête de l'internaute est envoyée en temps réel par Lycos à Fast, qui renvoie les résultats. Ceux-ci sont insérés par Lycos dans les pages de résultats. Cette procédure pourrait changer dans les mois qui viennent, notamment si Fast gagne d'autres clients. Le but étant, alors, que tous les clients du moteur n'affichent pas les mêmes résultats pour un même mot clé. Tous ces clients auront alors la possibilité de modifier légèrement certains critères (filtre familial, langue par défaut, etc.) pour obtenir un affichage différent de celui proposé par Fast. Un filtre familial est, d'ailleurs, déjà disponible (Lycos fournit à Fast un certain nombre de mots "interdits" afin que les pages qui les contiennent ne soient pas proposées dans les résultats), mais il sera étendu par la suite dans ses fonctionnalités.

L'intégration d'informations issues des "chaînes thématiques" du portail seront, d'autre part, bientôt incluses dans les résultats pour les 1000 premiers termes les plus demandés.

Dans les semaines qui viennent, un index ainsi qu'un module de recherche multimédia sera proposé sur Lycos France. Il devrait être très puissant et permettra de proposer des requêtes sur 17 millions d'objets multimédia (images, sons, vidéos, etc.).

Syntaxe d'interrogation

Pour l'instant, la syntaxe de Lycos est donc identique à celle de Fast. Voir :

<https://www.abondance.com/outils/alltheweb.html>

Il est à noter que la moyenne des requêtes des internautes sur Lycos sont sous la forme d'une suite de 2 à 3 mots. L'énorme majorité des mots clés demandés sont uniques. Seuls 0,1% sont répétés d'un internaute à l'autre. Comme le dit Felicitas Piegda, chez Lycos, "There are apparently many ways to ask for Britney Spears nude"... Le résultat est cependant assez bizarre. A mon avis, il faudrait revoir le programme de statistiques de Fast

;-)).

Critères de pertinence

Parmi les critères de pertinence pris en compte par Fast, les deux plus importants, de façon assez classique, sont le titre et le texte visible. Pour le texte visible, le nombre d'occurrences, la position dans le texte (début - fin) et la proximité des mots sont importants. En revanche, le fait qu'un terme soit mis en exergue (gras, Font Size, H1, texte d'un lien) n'est pas pris en compte. La rareté des mots (comme sur Altavista) non plus.

Les balises Meta, quelles qu'elles soient, ne sont pas prises en compte. Les commentaires non plus.

Les futurs noms de domaines accentués ne devraient pas poser de problème à l'indexation.

Les options ALT des balises images () sont prises en compte, mais avec un poids faible. Elles peuvent également servir pour l'index multimédia, afin de fournir un descriptif du contenu de l'image.

Le troisième critère, par ordre d'importance décroissant, est l'indice de popularité. Il est calculé à un seul niveau (uniquement quantitatif, il semble que l'indice des pages pointant vers un document ne soit pas pris en compte). Il n'existe pas, pour l'instant, de système d'indice de clic du type DirectHit qui soit proposé.

A priori, Fast n'aurait pas instauré de "tournante" sur le poids accordé à chaque critère de pertinence pour "faire tourner" l'index et les résultats qui en sont issus pour un même mot clé (l'information n'a pas été confirmée officiellement). Les résultats sont le plus souvent assez stables, donc, à l'affichage. Ceci dit, Fast va intensifier dans les mois qui viennent sa lutte contre le spam et il se pourrait que cette situation ne soit pas gravée dans le marbre...

Les pages dynamiques (ASP, CFM, PHP, JSP) ne sont que très peu prises en compte, même si elles ne contiennent pas de "?" dans leurs URLs. C'est l'un des "points faibles" de Fast, qui en est conscient. Le moteur doit donc travailler prochainement à une meilleure intégration de ce type de page.

En revanche, les frames sont plutôt bien intégrées (pages mère + pages filles) et le spider suit les liens Javascript. Les fichiers PDF seront plutôt trouvés dans l'index multimédia de Fast. Le format Flash n'est pas pris en compte.

Référencement sur Lycos / Fast

Lycos indique que si on veut être sûr d'avoir une présence optimisée chez Fast, donc sur Lycos, il faut soumettre son site sur Alltheweb, Lycos US et Lycos FR. Cela "nourrira" donc les trois index différents (plus la sauvegarde de l'index américain bien sûr). Cela explique également que, pour les mêmes mots clés sur Alltheweb, Lycos US et Lycos FR, vous n'obteniez pas la même réponse. Il est prévu, dans les semaines qui viennent, un délai d'une semaine au maximum pour la prise en compte d'une page soumise. A priori, il n'y a pas de différence pour Fast entre une page soumise manuellement et une page indexée par le spider en suivant un lien sur le web. Les deux documents seront traités de la même façon à l'indexation et dans le classement (ce qui n'est pas le cas de nombreux moteurs qui privilégient de plus en plus les pages crawlées de façon "naturelle" par le spider par rapport aux pages soumises manuellement, contenant 95% de spam selon Altavista et Inktomi).

Lorsqu'un site est soumis sur Lycos France, l'url est envoyée directement, en temps réel et sans traitement spécifique, à Fast. Cette situation pourrait changer à l'avenir, notamment pour des raisons de contrôle et/ou d'exploitation commerciale des données.

Fast n'indique pas de nombre maximal de pages soumises pour un même nom de domaine dans un délai de 24 heures. Cela pourrait éventuellement changer avec sa future politique de lutte contre le spam.

Fast annonce également un délai théorique de 14 jours entre deux rafraîchissements de son index.

Business Model

Le business model du contrat entre Lycos et Fast est comparable à celui d'Inktomi avec Hotbot (entre autres) : un paiement au nombre de requêtes. Le prix est gardé secret, bande de petits curieux ;-)...

Référencement payant

Des réflexions sont en cours (par l'observation de ce qui est proposé aujourd'hui sur le marché) en ce qui concerne la soumission payante de sites sur l'annuaire et le référencement de sites sur Fast. De la même façon, une réflexion existe pour ce qui est de l'intégration de résultats payés (de type de ceux proposés par GoTo) dans les résultats du moteur de recherche. Pour l'instant, rien n'est décidé, loin de là. Aucune décision n'est prise, Lycos serait plutôt en phase d'observation sur ce point. Restez connectés pour en savoir plus dans les mois qui viennent !

Adresses :

<http://www.fast.no/>
<http://www.alltheweb.com/>
<http://www-english.lycos.com/>
<http://www.lycos.fr/>

■ Altavista France : la nouvelle version est en ligne

ATTENTION : Je suis en attente de compléments d'information pour cet article, de la part d'Altavista France, notamment en ce qui concerne le projet "Britney". Ne désirant pas retarder plus l'envoi de cette lettre, je vous enverrai le complément de l'article dans un mail séparé, dès que j'aurai reçu les infos de la part du moteur. Désolé de ce contretemps...

Depuis le 8 mars dernier à 20 heures, l'AltaVista France nouvelle version est arrivé sur vos écrans. Il propose notamment :

- Un index francophone de 13,5 millions de pages. La structure des index est maintenant homogène pour l'Europe avec un index spécifique pour chaque pays.
- Un temps de réponse nettement amélioré par rapport aux versions précédentes.
- Un annuaire (Looksmart France) qui annonce 150 000 sites francophones.
- Une nouvelle interface, plus simple, plus épurée, plus intuitive.
- Un Espace Shopping et des pôles de recherche mis en place avec de sites partenaires.

Objectifs du nouveau site

L'objectif du nouveau site est d'associer la performance du moteur à de nouveaux outils de recherche. L'ensemble a été entièrement repensé pour permettre aux internautes d'optimiser leurs recherches d'information sur Internet. Les efforts d'amélioration ont porté notamment sur la simplicité d'utilisation, la rapidité d'accès aux informations, l'exhaustivité et la pertinence des résultats. La nouvelle version a demandé un gros travail de remise à plat et de modification radicale de l'outil, qui a été entièrement redéveloppé en base de données SQL7 Server. La version actuelle (nom de code "Monaco") est la V1.6. L'outil devrait très prochainement basculer dans sa version 2.0. Le basculement de tous les sites européens à la nouvelle technologie devrait être effectif avant l'été.

De plus, ce qui était jusqu'à maintenant un portail généraliste s'est recentré, comme il était annoncé depuis pas mal de temps, sur son coeur de métier, c'est-à-dire la recherche d'information. D'ailleurs, 90 à 95% des internautes qui utilisaient Altavista le faisaient dans cette optique, délaissant assez fortement les fonctionnalités "partenaires" qui "coûtaient" en ressources humaines et machines sans apporter, dans la pratique, de réel "plus" aux visiteurs. D'où le recentrage qui devenait, dès lors, inévitable...

Le moteur met également l'accent sur l'accès aux documents multimédia en offrant plus de 30 millions de documents (images, MP3/Audio, Vidéo).

AltaVista s'associe d'ailleurs en ce domaine à Canalweb, opérateur européen de télévision sur Internet (plus de 100 chaînes, 65 heures de direct chaque semaine et 10 000 heures d'émissions archivées). AltaVista propose par ailleurs de nouveaux outils de recherche. Ces derniers permettent aux internautes d'optimiser leurs recherches sur Internet et de personnaliser leurs résultats :

- Mes recherches : un outil pour enregistrer et transmettre ses recherches. Après avoir effectué une requête sur altavista.fr, l'internaute a dorénavant la possibilité de l'enregistrer et de la stocker parmi ses 25 recherches favorites. Il peut ainsi consulter à n'importe quel moment ses requêtes et transmettre les résultats par e-mail à un ami (ceci dit, il peut également les transmettre à une personne qui n'est pas son ami...;-)). Un projet d'alerte par e-mail ou message SMS est en train d'être mis en place. L'internaute sera ainsi averti lorsqu'un résultat a changé dans ses requêtes favorites : nouvelle url dans les 10 premiers résultats, changement de la date de dernière modification d'une page déjà présente, notifiant une mise à jour de celle-ci, etc.

Au sujet de cette fonctionnalité, on peut cependant regretter une certaine confusion entre le système de sauvegarde de la page entière de résultat ou au niveau du lien-résultat uniquement. La différence entre les trois zones de la page récapitulative ("Suivi des recherches", "Mes recherches favorites" et "Résultats de Mes recherches") n'est, dans ce sens, pas très évidente. Si l'idée est très intéressante, le tout semble encore un peu complexe à mettre en oeuvre. Mais l'outil est jeune et s'améliorera, notamment dans son ergonomie, certainement dans les mois qui viennent...

- Personnaliser : un outil pour paramétrer les pages de résultats : nombre de résultats affichés par page (de 10 à 50), descriptif complet ou succinct, langue des résultats de recherche, traduction dans la langue de son choix (Espagnol, Français, Allemand, Italien et Portugais) de tous les résultats obtenus.

- Dopez votre navigateur : l'internaute peut effectuer une recherche directe, simplement en surlignant un ou plusieurs mots d'une page web. Pour cela, il lui suffit de faire glisser le lien "Recherche rapide" dans la barre d'outils de son navigateur. Par la suite, il pourra à volonté sélectionner la partie du texte qui l'intéresse, et cliquer sur ce nouveau raccourci ajouté dans la barre du navigateur. La recherche se fera donc toute seule à partir d'une simple sélection du curseur !

- Les pôles de recherche : la nouvelle version d'AltaVista propose un accès à l'information par centre d'intérêt : emploi, automobile, petites annonces, shopping, TV-vidéos... Elle offre un accès à l'information grâce à une recherche par mot-clés ou par catégorie. ces zones sont réservées à des actions de partenariat qui permettent d'appliquer la technologie du moteur de recherche d'Altavista sur des sites partenaires. Il s'agit donc de réels partenariats qui vont bien plus loin qu'un simple co-branding.

Pages de résultats

Dans les pages de résultats, Altavista France a affiché pendant quelques temps les résultats issus de l'annuaire Looksmart France. Puis ces résultats ont disparu, a priori pour des raisons contractuelles entre Altavista et Looksmart. Toutefois, l'idée n'est pas abandonnée, bien au contraire. Il se pourrait que ces informations, qui apportent un réel "plus" à l'internaute, reviennent le jour prochainement.

Altavista France a également mis en place un système appelé "Intercept", qui permet de proposer une zone d'information "réglable" selon les mots clés saisis. Un exemple pour vous permettre de bien appréhender les potentialités du système : sur la page d'accueil d'Altavista France, tapez les mots clés "jazz" ou "DVD". Dans la zone affichée au dessus des liens-résultats, il apparaît une image et un lien de partenariat avec Amazon.fr, qui n'a rien à voir avec le bandeau publicitaire affiché en haute de page : il s'agit, dans ce cas, de mettre en valeur les partenaires d'Altavista France au travers d'une zone réactive aux mots clés tapés par le visiteur de l'outil. Cette zone est entièrement configurable et peut contenir de nombreuses informations, sans être uniquement un simple lien vers un site partenaire.

D'autre part, Altavista France propose également un autre format de bandeau publicitaire, appelé "Sky Scrapper" ou "gratte ciel" (format 120x600), sur la droite des pages de résultat. Ce format génèrerait, a priori, un taux de clic plus important que le traditionnel 468x60 horizontal.

Référencement

Depuis quelques jours, si vous soumettez un site sur le moteur d'Altavista France et que vous avez un message qui vous indique que l'url soumise aboutit à une erreur 404, soulignant que votre site sera déréférencé et enlevé de l'index, ne vous affolez pas : c'est un bug d'affichage. La page soumise est bien prise en compte pour indexation, sans problème. Le message affiché est "juste" un bug qui est en train d'être corrigé.

Les délais d'indexation d'une page soumise oscillent entre 7 et 10 jours. Il est à noter qu'Altavista France a largement amélioré dernièrement, la situation au niveau des soumissions par son interface manuelle.

D'autre part, a priori, les noms de domaines accentués ne devraient pas poser de problème à Altavista lorsqu'ils seront disponibles officiellement.

Projets en cours

Dans les résultats proposés par Altavista France, il n'y a pas encore de dictionnaire d'expressions (comparable aux "related searches" de la version américaine), mais ça ne saurait tarder. Le système est en cours d'évaluation et pourrait voir le jour prochainement. Contrairement à la version américaine (voir article dans la lettre "Réacteur" du mois dernier), c'est donc un OU qui est inséré entre les différents mots saisis par les internautes sur le moteur (s'il y en a plusieurs, bien entendu...).

L'un des gros projets d'Altavista est d'ailleurs de mettre en place un système d'analyse sémantique qui saurait reconnaître les utilisateurs, leurs centres d'intérêt et leurs façons de chercher pour leur proposer des résultats encore plus fiables à l'avenir. Aucun délai n'est défini pour ce projet qui est encore dans les laboratoires de Palo Alto. Mais il semble s'agir d'une priorité assez importante pour Altavista.

Enfin, plus immédiatement, Altavista France a mis en place le projet "Britney". Il a tout d'abord été mis en place sur Altavista Grande-Bretagne : lorsque vous saisissez, sur ce moteur, des mots clés comme "ferrari" ou "bmw", par exemple, vous voyez apparaître, dans les pages de résultats, une nouvelle fonctionnalité originale et innovante : une image vient mettre en valeur certains liens proposés. L'image est, bien entendu, extraite de la page en question. Ce projet, connu sous le nom de "Britney" en interne chez Altavista, est également disponible depuis peu sur la version française du moteur. Tapez un mot clé comme "Renault" ou "Peugeot" et vous verrez apparaître une image, issues des pages web proposées, qui met en valeur certains liens affichés comme résultat. Pour l'instant, ces images ne semblent apparaître que pour les recherches sur l'index

francophone (bouton radio "France", le système ne fonctionne pas sur une recherche au niveau mondial). L'image n'est pas calculée en temps réel car certaines pages qui aboutissent sur une erreur 404 se voient quand même mises en valeur par une vignette. Seules les images au format JPEG semblent être prises en compte. Un lien "Toutes les images de cette page" permet également de visualiser... eh ben, comme son nom l'indique ;-)

Le système n'est pas neutre, car il consiste en une mise en valeur du lien et donc à une incitation plus forte à cliquer dessus pour l'utilisateur du moteur. Nul doute que les référenceurs malins vont se pencher sur la question pour réaliser des pages compatibles avec ce nouveau système :-)

Je suis en attente d'un certain nombre d'informations complémentaires sur cette fonction et sur la façon de réaliser une page qui soit "Britney Friendly", c'est-à-dire de pouvoir dire à Altavista quelle image de la page doit être affichée, ce qui a une grande importance en ce qui concerne le référencement. Dès que j'aurai plus d'infos de la part d'Altavista France (c'est en cours), je vous les ferai parvenir par mail. Promis :-)

Conclusion

Bref, Altavista France propose, avec cette nouvelle version, de nouvelles fonctionnalités inédites qui replongent Altavista dans son rôle historique de défricheur de solutions originales adaptées à l'internaute pour la recherche d'information sur le Web. Ca commençait à nous manquer ;-)

Il est également à noter que les versions "hors USA" se démarquent de plus en plus fortement du site américain, qui n'a plus grand chose à voir, que ce soit au niveau du look que des fonctionnalités, avec les autres moutures d'Altavista. Et la situation n'est pas partie pour se modifier dans les semaines qui viennent...

<http://www.altavista.fr/>

■ Webhelp.fr, le moteur de recherche humain

Présentation de la société

La société Webhelp est née en France et en Europe d'un accord de partenariat entre Webhelp Inc, la "société mère" américaine, et Europ@web qui a investi près de 7 millions de dollars dans la société. Webhelp SA est dirigée par ses deux fondateurs, Olivier Duha, ESC Poitiers/Nantes - MBA INSEAD et Frédéric Jousset, HEC, tous les deux ayant quitté leur poste de consultant chez Bain & Company pour se consacrer au développement de ce concept.

Webhelp propose deux types d'offres :

- Une à destination du grand public (BtoC): Webhelp propose son portail généraliste Webhelp.fr intégrant son moteur de recherche à assistance humaine en temps réel. L'internaute rentre alors en discussion avec un des 150 WebWizards (littéralement "génie du web") qui tente d'identifier pour lui les sites susceptibles de correspondre à ses critères de recherche.
- Une à destination des entreprises (BtoB): Webhelp propose son infrastructure technique et ses logiciels ainsi que les ressources humaines spécialisés de ses Webcenters. Ses services de chat, d'Email Management et de selfhelp ont pour objectif de couvrir tous les cycles de la relation client Internet. Webhelp cible les entreprises de l'Internet (sites d'e-commerce, fournisseurs d'accès, portails,..) pour accompagner leurs visiteurs dans leurs processus de navigation, d'achat ou d'assistance technique.

Seule la fonctionnalité de recherche humaine au travers des WebWizards nous intéressera dans cet article.

Webhelp.fr représente donc la structure française de Webhelp.com, et a commencé à s'internationaliser en Europe avec la création, depuis quelques semaines, du site allemand (<http://www.webhelp.de/>). Il est à noter que, contrairement à sa mouture française, la "maison-mère" américaine est devenue un site entièrement payant, hormis en ce qui concerne la mise à disposition de la base de données de questions déjà posées (et des réponses correspondantes), qui reste encore en libre accès.

Modèle économique

Car le modèle économique de Webhelp.fr est clairement axé, pour sa partie "moteur de recherche humain", sur la gratuité. Il ne semble pas dans les projets de Webhelp.fr de suivre, pour l'instant, le site américain. L'outil devrait donc demeurer gratuit dans les mois qui viennent. L'offre de recherche d'information est en fait un produit d'appel B2C, une vitrine technologique, qui amène Webhelp à vendre ses solutions B2B de E-CRM.

Procédure de réponse aux questions des internautes

Lorsqu'un internaute arrive sur le site de Webhelp, il pose sa question dans le cadre prévu à cet effet. L'objectif de l'outil est qu'un nouveau venu sur le site soit pris en charge dans les 45 secondes qui suivent le moment où il a posé sa question. A ce délai, il faut ajouter les délais de transmission sur la plate-forme de WebHelp (45" également), ce qui amène à un délai maximum de prise en compte "admissible" de 1'30" en moyenne. Prochainement, ce délai de transmission sera ramené à 15" avec la mise en oeuvre d'une nouvelle plate-forme "0 latency" ramenant le délai total à 1' environ. Rentre alors en scène un "WebWizard" (un netsurfeur) qui va tenter de répondre aux interrogations de l'internaute au travers d'un "chat" permettant de cerner sa recherche. Sur la droite de l'écran va s'afficher le chat tandis que la partie gauche a deux utilités : soit afficher les pages de sites partenaires (publicité proposée en fonction des termes recherchés par le visiteur - dans ce cas, le taux de clic est en moyenne de 52% sur ce "billboard"), soit afficher (Webhelp emploie le terme de "pousser") les sites proposés par le WebWizard.

Celui a à sa disposition plusieurs sources d'information pour mener à bien ses investigations :

- En premier lieu, des listes de sites considérés comme pertinents et reconnus au préalable comme pouvant répondre à de nombreuses problématiques le plus souvent demandées : sites de téléchargement de MP3, banques de données d'images, sites de référence, encyclopédies, etc.
- La base de données des questions déjà posées par le passé et des réponses qui y ont été apportées (Selfhelp). Un système de validation par l'internaute demandeur permet également d'avoir une idée de la satisfaction de ce dernier. Les WebWizards sont d'ailleurs notés en partie sur cet indice de satisfaction. Le WebWizard peut taper une question et l'outil va lui proposer celles qui sont les plus proches précédemment posées et leurs réponses respectives. En cela, cet outil peut être comparé à un système proche d'Ask Jeeves ou d'Infoclic, mais utilisé en interne, sur la base propre de Webhelp.fr.
- Des outils de recherche "classiques", et notamment Google et Voila. Autres outils éventuellement utilisés : Altavista (notamment pour les recherches d'images), Yahoo!, Lycos et Northernlight.
- En dernier lieu et plutôt en bouée de sauvetage, une recherche sur Copernic est lancée.

Aucun site payant n'est utilisé comme source d'information car il ne serait pas possible, juridiquement parlant et dans ce cas, de "pousser" la page estimée comme pertinente sur l'écran de navigation de l'internaute. Seules des sources d'information gratuites sont donc prises en compte.

Un chat représente une durée moyenne de 10 à 12 minutes.

Il est demandé aux WebWizards de traiter un maximum de deux questions au cours d'un même chat + 2 approfondissements pour chaque question par interlocuteur. En effet, si les discussions s'éternisent et si l'internaute pose trop de questions, la file d'attente risque de s'allonger

dangereusement. Les WebWizards sont évalués sur de nombreux points. Tous les dialogues sont stockés (questions / réponses). Ces contrôles peuvent se faire en différé (examens des dialogues passés du WebWizard et évaluation de leur pertinence), surveillance en temps réel à distance (permet de contrôler le timing et de vérifier que la durée d'un dialogue ne s'allonge pas trop) et sur place, à côté du WebWizard, avec une aide humaine (comment surfe-t-il ? respecte-t-il les procédures ? connaît-il vraiment bien les outils de recherche et leur syntaxe ?).

Certains thèmes ne sont pas traités : le sexe, la violence, l'incitation à la haine raciale, etc. Les dilemmes sont courants : si quelqu'un pose la question "Où puis-je trouver "Mein Kampf" ?", il peut s'agir soit d'un étudiant en histoire qui désire se documenter sur la deuxième guerre mondiale, soit d'un pur néo-nazi. Pas si simple de faire le tri en quelques minutes. Le credo adopté par le WebWizard est le suivant : "si tu hésites sur le bien fondé de la demande, tu ne donnes pas l'info". Cependant, seuls 1 à 2% des questions ont trait à des thèmes "extrêmes", ce qui n'est pas représentatif du phénomène observé sur les annuaires et moteurs "classiques". Deux raisons à cela : d'une part, l'internaute peut savoir que Webhelp.fr ne traite pas ces questions (fidélité) et d'autre part, il pose quand même, même par clavier interposé, sa question à un être humain directement, ce qui est très différent du processus d'un annuaire ou d'un moteur.

Lors de votre première recherche, Webhelp.fr vous demandera d'enregistrer un nom d'utilisateur. Une fois enregistré, le système saura vous identifier. Webhelp propose également un service dit "Express" qui est un service prioritaire. Les abonnés payants à Webhelp Express bénéficient d'un service plus rapide qui vous permettra d'être traité comme un client privilégié et donc prioritaire. L'outil propose trois durées d'abonnement (1, 3 et 6 mois).

Dans ce cas, vous pouvez poser autant de questions que vous voulez pendant toute la période de votre abonnement. Les prix sont de 59 F, 159 F ou 259 F selon la durée prise en compte.

Plus d'infos :

<http://french.webhelp.com/french/MemberPurchase.execute.j>

Organisation interne

L'équipe des WebWizards est constituée de 150 personnes travaillant dans un "Web center" en Roumanie. Webhelp assure également des services B2B depuis l'Ile Maurice.

Les personnes répondant aux questions sont donc des roumains, sachant très bien écrire le français. Ce sont des professeurs de français, des techniciens ou des étudiants. 80% des personnes travaillent à temps plein pour Webhelp, 20% à temps partiel. Pour l'anecdote, sachez que chaque mois a lieu l'élection du meilleur WebWizard.

Le niveau d'étude du WebWizard "classique" est de Bac+4. Il suit une formation de 3 semaines avant de commencer, puis fait l'objet d'une formation continue tout au long de son travail.

Le travail s'effectue, en règle générale, par tranches de 6 à 8 heures comprenant une pause toutes les deux heures.

Webhelp base son système informatique sur l'architecture distribuée de Webhelp.com : les bases et les serveurs sont en France et aux Etats-Unis. Le système de chat évolue, de son côté, totalement sur une toute nouvelle technologie.

Profil des utilisateurs du site

Selon une étude basée sur un sondage effectué en décembre dernier (1024 utilisateurs du service ont été interrogés), les utilisateurs du service Webhelp.fr ont le profil suivant :

Sexe :

Homme = 67 %

Femme = 33 %

Age :

- de 15 ans = 1 %

15 - 24 ans = 32 %

25 - 34 ans = 31 %

35 - 49 ans = 20 %

50 - 64 ans = 13 %

+de 65 ans = 2 %

Activité :

Prof. Libérales / cadres = 31 %

Étudiant = 28 %

Employés = 20%

Professions intermédiaires = 8 %

Chefs d'entreprise = 7 %

Inactifs = 4%

Agriculteurs = 1 %

Ouvriers = 1 %

Répartition géographique :

Paris = 26,5 %

Province = 48 %

Etranger = 25,5

Utilisation d'Internet :

- de 6 mois = 9 %

Entre 6 mois et 1 an = 16 %

+ d'un an = 75 %

Lieu de connexion :

Domicile = 56 %

Professionnel = 32 %

Scolaire = 8 %

Fréquence de connexion :

Au moins 1 fois par jour = 82 %

Au moins 1 fois par semaine = 17 %

Au moins 1 fois par mois = 1 %

Achat sur le Net au cours des 6 derniers mois :

Oui = 56 %

Non = 45 %

Thèmes les plus souvent demandés

Webhelp.fr traite 2 000 questions par jour. Chaque WebWizard prend en charge plus de 1 000 interrogations par mois.

Les thèmes traités sont très divers. Sur 2 800 questions, dans les semaines qui viennent de s'écouler, voici les domaines qui ont été demandés par les internautes :

Logiciels et nouvelles technologies : 14%
Sciences et médecine : 8%
Divertissements : 8%
Informations : 8%
Socio-culture : 7%
Sites Internet : 6%
Emploi : 6%
Entreprises : 5%
Médias : 5%
Musique : 5%
Littérature Arts : 5%
Economie Finance Bourse : 4%
Cours, apprentissage, formation : 4%
Législation : 3%
Personnalités, people : 3%
Achats ventes : 3%
Voyages : 2%
Adresses, annuaires : 2%
Dictionnaire, traduction : 2%
Cartes et plans : 1%
Données chiffrées : 1%

Les clients de Webhelp sont cependant satisfaits, selon une étude menée par l'outil, à 79% et sont fidèles à 80%.

Soumission d'un site

Il est possible de proposer un site à Webhelp à l'adresse suivante :

Service-client@direct.webhelp.fr

Le site est alors évalué par le contrôle qualité et, s'il est jugé comme pertinent, il sera ajouté à la base de données de liens utilisés par les WebWizards de façon quotidienne selon les thèmes traités. Un e-mail "type" ("votre demande a bien été bien prise en compte") est envoyé au webmaster dès réception. Cependant, aucun e-mail n'est envoyé pour notifier l'acceptation ou le refus d'un site qui n'est, de toutes façons, considéré que comme une suggestion. Il est très important pour Webhelp de répondre à tous les mails qui lui sont adressés, puisque cela fait partie des prestations qu'il vend à ses clients.

Principales difficultés rencontrées par les Webwizards

Les principales difficultés auxquels sont confrontés les WebWizards au quotidien sont les suivantes :

- Questions très complexes, parfois quasi insolubles.
- Questions mal formulées, parfois incompréhensibles.
- Questions illégales.
- Tentation de l'internaute de faire tourner le chat à la conversation ("tu es brune ou blonde ? ;-)), ce qui n'est pas le but de l'outil.

Les questions les plus difficiles concernent l'installation et l'utilisation de logiciels, drivers, cartes pour lesquels les WebWizards du B2C ne peuvent se substituer aux hotlines des éditeurs et constructeurs. C'est d'ailleurs ce type de service que Webhelp offrent aux abonnés de certains de ses clients dans le cadre de son offre B2B.

Les perles

Pour nous faire passer un bon moment, voici un collier de perles confectionné avec les questions les plus délirantes ou les plus "tordues" qui ont été posées aux webwizards pendant les derniers mois (si vous avez la réponse, ils sont preneurs... ;-)) :

- Combien d'escaliers a l'Elysée ?
- Comment puis je me marier avec une américaine?
- Bonjour j'ai besoin d'aide
- Je cherche des familles au pair pour des hommes...
- Les classes sociales sont elles toujours pertinentes pour decrire la société actuelle?
- Je cherche une photo de taille A4 d'un crumble aux pommes. Je n'ai pas réussi à trouver par moi-même. Pourriez- vous m'aider svp ?
- Je cherche un site où je pourrais donner mon chien à un particulier
- Est-ce que les femmes ayant les cheveux blonds (au naturel) ont aussi le poil pubien blond?
- Je cherche des informations sur les années 1890, que s'est-il passé d'important à cette époque, comment était la vie, etc... Merci!!!
- Je cherche un fille scandinavienne pour le mariage
- Je cherche un logiciel pour se teleporter
- Bonjour je voudrais une photo d'une plante en train de transpirer
- Je cherche une ferrari rose
- S'il vous-plait pouvez-vous m'envoyer des sites payant, je veux dire les sites qui peuvent me payer?
- Je suis à la recherche d'un site où acheter des colliers de fleurs (comme dans les îles)
- Comment fait-on pour recevoir les adresses des femmes qui cherche un mari ?
- En quoi José Garcia est-il l'acteur comique que les francais attendaient
- Bonsoir! je recherche des expressions sur l'ambition ou la patience.
- Quelle est la définition du mariage
- Salut! je voudrais savoir comment faire partir des suçons sur le cou au plus vite possible
- Pourquoi peut-on voir plus d'étoile dans la campagne que dans la ville?
- Si rien ne colle au teflon, comment colle-t-on le teflon sur la poele?
- Parlez moi d'amour
- Il me faut un sujet ou je met en relation deux phenomenes , un qui explique et l'autre explicatif pour cela il me faut des chiffres , par exemple quels sont les criteres pour avoir le droit a un pret bancaire de telle somme?
- Je cherche 20.000 francs

- Je voudrais des photos de pieds nus
- Quelle est- donc la recette du bonheur ?
- Un site qui lorsqu'on tape une question, nous crée des dissertations... sur tout sujet...
- Bonjour ! La vie est belle pour vous ?
- Client : Dites ouaf ouaf si vous êtes une vraie personne. Wizard : ouaf ouaf. Client : remuez la queue, maintenant.

Projets

Webhelp prépare de nouveaux projets pour les mois qui viennent : - Expansion vers d'autres pays européens. De nombreuses réflexions sont en cours sur ce point.

- Nouvelle technologie de chat, plus rapide et plus conviviale.
- De nouveaux services : recherche approfondie en différé, accès à la base de connaissance Webhelp, service prioritaire,
- Fourniture de l'outil Webhelp en marque blanche auprès de grands portails.
- Spécialisation par thèmes de l'outil de recherche.
- Un objectif de certification qualité en 2001.

WebHelp :
<http://www.webhelp.fr/>

La FAQ sur Webhelp :
http://www.webhelp.fr/faqs/faq_compagnie.php3?n=1

■ Le portrait d'un Websurfeur chez MSN

Un petit texte qui nous a aimablement été envoyé par MSN France. Merci à eux :
 Valérie Lecaille Bourgier, 30 ans, est Responsable du moteur de recherche MSN Search. En quelques lignes, elle décrit le métier de websurfeur chez MSN.

"Sur MSN Search, notre intervention en tant que "fouilleurs du Web" est présente à tous les niveaux du service de recherche, du démarrage à la sélection des résultats. Pour cela, nous travaillons en permanence à la sélection, l'indexation et la mise à jour des meilleurs sites Internet dans le souci de toujours mieux satisfaire les intérêts des internautes".

1/ Sélection des sites

"Notre rôle dans la sélection des sites Web consiste à parcourir le Web de façon quotidienne, à prendre en compte les sites qui nous sont proposés par nos utilisateurs, toujours à la recherche de nouveaux sites, de nouvelles tendances et à l'écoute de l'actualité du moment. Ceci dans le but d'anticiper les attentes de nos utilisateurs. Chaque site est examiné avec soin, commenté puis évalué en fonction des différents publics d'internautes (enfants, adolescents, adultes). Nos principaux critères de sélection prennent en compte la qualité et la richesse du contenu du site, la conformité à la loi française et la notoriété ou la fiabilité de l'éditeur".

2/ Indexation

"Nous participons également activement à l'indexation des sites de l'annuaire. Selon la pertinence des sites sélectionnés, nous choisissons de mettre en avant et d'apposer trois étoiles à ceux qui nous semblent les plus appropriés et au contexte et/ou à l'actualité. Ainsi, dans la catégorie Emploi, Formations et Organismes Publics, nous proposons en premier les Organismes nationaux officiels (ANPE, APEC, ASSEDIC), que nous jugeons "incontournables" par rapport à toute question dans le domaine de l'Emploi.

Notre valeur ajoutée en tant "qu'humains" se matérialise également par notre capacité à référencer l'ensemble des sites associés à un ou plusieurs mots clés. Par exemple, si vous tapez aujourd'hui "ESB", "Vache folle" ou "Prion" sur le moteur, vous obtiendrez la même liste de sites dont les premiers correspondront à l'actualité sur la sécurité alimentaire, ceci étant la préoccupation majeure. L'actualité du jour et la découverte de sites intéressants nous incitent donc à créer ou à affiner une catégorie de l'annuaire pour une meilleure navigation dans l'arborescence."

3/ Mise à jour

"Dans un souci permanent d'amélioration des résultats de recherche, nous assurons une mise à jour quotidienne des liens que nous proposons, en veillant toujours à la pertinence de notre sélection. Cette réactualisation se fait aussi au travers des réponses que nous apportons aux internautes sur le service "Nous cherchons, vous trouvez". A ce titre, régulièrement, les internautes nous posent des questions plutôt surprenantes, comme "Comment trouver l'amour ?" ou "Dieu existe-t-il ?" Ce qui montre bien que les internautes attendent tout d'un moteur 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 France, Google, Fast (donc Lycos et Lycos 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) : Altavista, Inktomi (donc Hotbot), Ecila, Excite France.

Les moteurs à la peine ce mois-ci : Voila, Excite.

■ 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 ;-)

-> Looksmart.fr semblerait proposer des offres de soumission payante "en gros" à certaines sociétés de référencement. Ils proposent aux sociétés qui soumettent plus de 20 sites par mois une remise sur volume. La tarification serait la suivante:

- pour un volume supérieur à 20 sites soumis : 10% de remise.
- pour un volume supérieur à 40 sites soumis : 20% de remise.

Une procédure spécifique n'étant pas encore en place, Looksmart France propose au référenceurs d'utiliser la procédure existante disponible sur www.looksmart.fr (formulaire et paiement par CB). Ils procèdent à la fin de chaque mois à une régularisation par un remboursement correspondant au montant de remise sur le volume de soumissions.

-> Puisqu'on en est au référencement payant, un outil de recherche francophone majeur va mettre en place, dans les semaines qui viennent, plusieurs offres de ce type. Non, désolé, je ne peux pas vous dire qui ;-)

-> Le moteur Kartoo sera lancé officiellement d'ici quelques jours (éventuellement vendredi 20 avril ?). Dans tous les cas avant la fin du mois d'avril.

-> Free ne propose plus les résultats du moteur de recherche en langage naturel Albert qui travaillait sur la base de l'index de Fast (qui fait partie des actionnaires d'Albert). Renseignement pris, Free semble avoir été mécontent de la qualité des réponses de Fast, raison pour laquelle ils sont passé sur Google. Mais Albert fonctionne encore sur leur pages perso et reste en discussion pour d'autres projets sur le portail. La technologie n'est donc pas, a priori, rejetée.

-> Une nouvelle maquette du site Francopholistes (<http://www.francopholistes.com/>) aux couleurs de Club-Internet devrait être disponible en ligne au plus tard début mai.

-> Le nouveau système de soumission d'url d'Altavista.com (image affichée proposant des chiffres et des lettres à ressaisir à la main), serait interprétable sans trop de problèmes grâce à des logiciels d'OCR. Altavista sait que ces codes sont parfois difficilement lisibles par un être humain. Il travaille à les rendre plus accessibles. En revanche, depuis que cette fonctionnalité a été mise en place, les pages disponibles à des urls "exotiques" ("?" dans leur énoncé) seraient à nouveau prises en compte.

-> Le guide de Voila travaille de façon approfondie sur une nouvelle version de leur outil (appelé "Super Guide"), ce qui explique un léger (?) ralentissement des prises en compte de sites actuellement par les netsurfeurs. Rendez-vous très bientôt !

-> Le nouvel index de Voila.fr devrait contenir 60 millions de pages web francophones. La taille actuelle serait de 30 millions. Voila a finalement arrêté d'afficher les Realnames dans ses résultats, euh, on va dire "pour des raisons techniques" ;-)

-> Inktomi tient maintenant compte de la proximité des termes demandés dans une requête (une demande sur "Jacques Chirac" classera en premier les documents qui contiennent ces deux mots l'un à côté de l'autre).

-> Nomade.fr met en place une nouvelle architecture (arborescence) pour son annuaire. Vous avez peut-être déjà vu pas mal de modifications être mises en place ces jours derniers. Fin des travaux prévue pour fin avril.

-> Sur les 300 000 soumissions que reçoit le site Altavista.com chaque jour, la moitié correspondraient à des urls "cassées" (erreurs 404). Sur les 150 000 restantes, 142 000 correspondraient à des pages satellites ou pages alias. Seules 7 à 8 000 pages correspondraient donc à des documents "propres" pour Altavista. Selon le moteur, seuls 400 webmasters utilisent chaque jour cette fonctionnalité (soit environ 20 pages soumises en moyenne par référenceur et par jour). Avec le nouveau système de soumission, la limite de 5 pages maximum soumises par jour semble ne plus exister. Il n'y aurait plus de limite. Le délai d'indexation serait d'une semaine à dix jours environ.

■ En Bref

-> Lycos France a mis en place une nouvelle rubrique : "les sites préférés des netsurfeurs" :
http://www.fr.lycos.de/dir/Informatique_Multimedia/Internet/Recherche_sur_Internet/Les_sites_preferes_des_surfeurs/

-> Si vous avez regardé vos logs (la mémoire des connexions de votre serveur), vous avez peut-être trouvé trace d'un robot appelé "Antibot", très actif en ce moment. Il s'agit du "spider" de la société Antidot (groupe IXO, ex Infosources), entreprise qui travaille sur la maintenance du moteur Lokace, mais également sur de nouveaux produits et de nouvelles offres commerciales basées sur le "business model" d'Inktomi (vente de solution de recherche d'information clés en mains). L'outil de recherche est disponible sur le site Antisearch. Il s'agira d'outils de type "agents intelligents" dont l'un d'entre eux s'occupera de rapatrier des pages web pour bâtir un index qu'il sera possible de questionner en texte intégral par mots clés. Le tout sera lancé d'ici la fin du mois de mai.

Plus d'infos :

<http://www.antibot.net/>

<http://www.antisearch.net/>

<http://www.antidot.fr/>

-> Sur Voila, 20 000 requêtes correspondent à 50% des demandes des internautes.

-> Une nouvelle FAQ est disponible sur Lycos France, à l'adresse <http://www.fr.lycos.de/help/faq/>

-> L'annuaire de Lookmsart France est depuis peu disponible sur le site Génie (<http://www.genie.fr/>). De nouveaux partenariats devraient être annoncés très bientôt.

