

■ Gros plan sur l'annuaire de l'île des Médias

Présentation

La société L'île des Médias (<http://www.ile-des-medias.com/>) fournit un annuaire en marque blanche à de nombreux portails francophones (voir ci-après). La société, employant 58 personnes, est divisée en plusieurs pôles :

- Groupe de presse, avec une activité à la fois d'éditeur (overgame.com, loginews.com, annuaire blanc) et de co-éditeur (Marianne en ligne, Nouvel Economiste, Bodoï).
- Web agency avec une activité "classique" (hébergement, graphisme, création de sites, etc.) et des prestations plus spécifiquement éditoriales (Astérix - Editions Albert René, SNCF, Bifi, Paris Avenue, etc.).

Seule la fonctionnalité "annuaire blanc" sera traitée ici. Cette activité a été créée en avril 1999. Cette création s'est faite alors qu'il existait déjà un certain nombre d'annuaires sur le marché, avec les avantages (analyse des bons et des mauvais côtés des annuaires existants en essayant de ne pas commettre les mêmes erreurs) et les inconvénients (une partie du marché déjà occupée) de cette situation. Dans un premier temps, le but a été de créer une arborescence qui soit la plus "propre" possible. L'île des Médias a contacté des universités pour obtenir un thésaurus qui soit le plus efficace possible. Aujourd'hui encore, une personne à plein temps est employée pour travailler sur l'arborescence, la création de nouvelles catégories et le reclassement de sites déjà répertoriés dans de nouvelles rubriques. Le contenu de nombreuses catégories est, d'ailleurs, souvent revu et corrigé, certaines parties de l'arborescence qui méritaient une révision ont été reprises et complétées dernièrement (musique, loisirs, jeux vidéos, etc.).

Au plus fort de l'activité de création du service (juin 1999), 16 personnes travaillaient à temps plein sur l'annuaire. Aujourd'hui, ce sont 10 personnes qui sont occupées à mettre à jour l'outil de recherche. Chaque netreur reste en moyenne 12 à 18 mois en poste. Ils ne sont pas spécialisés sur une thématique donnée et "tournent" sur les grands domaines d'informations traités par l'annuaire. L'annuaire compte aujourd'hui 61 000 sites répartis dans - exactement - 5 290 catégories, avec un rythme de croisière de 3000 nouveaux sites par mois. Le but de l'île des Médias est de ne jamais aller au delà de la barre de 6 niveaux d'arborescence et de 60 à 80 sites proposés par rubriques. Au-delà, une nouvelle catégorie doit être créée.

Critères d'acceptation et de refus

Les critères d'acceptation et de refus sont classiques pour un annuaire (voir les articles dans les lettres précédentes sur Yahoo.fr, Lycos.fr, Nomade.fr et le Guide de Voila). Cependant, deux types de site sont automatiquement refusés :

- Les sites qui présentent un CV individuel
- Les petites annonces immobilières de particuliers

Les pages perso ne sont prises en compte que si elles présentent une réelle dominante sur un sujet précis, fournissant un contenu de bonne qualité. Sinon, elles sont refusées.

Les sites X ne sont acceptés que s'ils sont "softs", et en très petite quantité. Aucun site proposant des fichiers MP3 "pirate" n'est également accepté. Une fois pris en compte de façon positive, le site est rentré dans la base de données grâce à un outil "maison" (base de données écrite en VBScript sous Access), après 4 cycles de relecture : 3 humains et un grâce à un logiciel de correction orthographique.

Référencement d'un site web sur l'annuaire

Aucun des clients utilisant l'annuaire ne propose de formulaire de soumission d'un site dans l'outil de recherche, pas plus que sur le site de l'île des Médias. En effet, cette société a pris le parti de ne pas prendre en compte de soumission de cette façon, les sites étant plutôt entrés dans la base au travers du travail de veille des documentalistes (qui utilisent beaucoup la presse papier spécialisée, les sites web d'actualité, des livres de type "annuaire des X meilleurs sites web" ainsi que des annuaires professionnels papier). Cependant (rappelons que cette information n'est parue que dans la lettre "Réacteur" d'Abondance, ces mois derniers - merci à l'île des Médias pour leur confiance), l'annuaire propose la soumission par courrier électronique des sites à l'adresse referencement@annuaire-blanc.com.

Merci de préciser dans votre mail :

- L'adresse du site
- Son titre
- Un descriptif permettant de cerner le sujet.
- Des mots clés parfaitement descriptifs et spécifiques de l'objet du site (évités les mots clés top généralistes qui ne servent à rien).

La seule soumission volontaire possible sur cet annuaire s'effectue donc par e-mail, en suivant la procédure décrite ci-dessus. Merci de ne pas en abuser et de ne leur soumettre que des sites de qualité et surtout de ne faire aucune tentative de spam. En cas de problème, l'adresse serait irrémédiablement annulée. Personne n'a à gagner à ce que cela se passe ainsi. Merci de votre compréhension. Pour l'instant, le constat de l'île des médias sur cette procédure est que cela se passe plutôt bien, même si :

- Certaines sociétés spécialisées dans le référencement ne soumettent jamais aucun site.
- Certaines sociétés envoient, au contraire, beaucoup de soumissions.

Attention à l'overdose et à ne pas soumettre trop de propositions en une seule journée !

Tous les résumés mis en ligne sont rédigés par les documentalistes, donc donnez-leur un descriptif qui aide le (la) documentaliste dans son travail plutôt qu'un discours commercial. Les résumés sont, cependant, toujours réécrits par les netreurs, la formulation finale devant rester dans la ligne éditoriale de l'annuaire.

Cependant, si vous avez signalé certains mots clés essentiels pour l'activité du site (et pas trop génériques), les netreurs essaient toujours d'en tenir compte en rédigeant leurs fiches. Le délai pour que votre site apparaisse est d'environ 1 mois. Un mail de confirmation indiquant que le site est passé en traitement, qui se fait par vagues régulières pour les soumissions par mail, est envoyé. Ceci ne signifie pas que le site sera retenu. L'île des médias n'envoie pas de confirmation aux sociétés soumettant des sites en permanence.

Une demande de modification est à adresser à la même adresse e-mail. Ces demandes ne sont prises en compte que pour des modifications réellement argumentées et répondant à un besoin spécifique, portant sur un point précis et sur un manque spécifique de la fiche descriptive actuellement en ligne (aucune chance s'il s'agit de rajouter un terme générique pour des raisons de positionnement). Aucune fonctionnalité de prise en compte, référencement et positionnement payant n'est prévue dans les mois qui viennent sur cet annuaire, ce sujet ayant déjà fait l'objet d'un débat au sein des équipes de l'outil de recherche.

Traitement des sites inscrits - Mise à jour de la base de données

Le temps de traitement du site en interne est d'environ une semaine. Cependant, la base de données est basculée vers le site des clients une fois par mois. Pour cette opération (appelée "synchronisation de la base"), plusieurs opérations sont effectuées :

- Détection des erreurs 404 ;
- Intégration des nouveaux sites ;
- Elimination des sites obsolètes ("ménages") : 4 000 sites vont ainsi "sauter" lors de la prochaine "synchronisation" : l'annuaire passera alors de 65 000 à 61 000 sites.

La procédure de mise à jour et d'hébergement est légèrement différente selon les clients de l'Ile des Médias :

- Excite.fr reçoit un fichier contenant toutes les données. C'est lui qui installe ensuite ce fichier sur ses propres machines. L'annuaire est hébergé chez Excite.fr, alors qu'il tourne sur les serveurs de l'Ile des Médias pour toutes les autres configurations.
- Nouvel Economiste, Mesactions.fr, Altavista France, Planetis (nouveau portail de NEC : en ligne prochainement), Spray France, autres (en cours) : l'annuaire est co-brandé et tourne sur les serveurs de l'Ile des Médias, hébergés chez Infosources.
- Lokace : leur cas est un peu particulier : il s'agit d'une copie de l'annuaire "classique", mais dont les résumés sont réécrits spécifiquement pour mieux coller à la cible (jeune) du portail Lokace. Notons que les cas d'Altavista France et de Spray France sont un peu particuliers : Altavista France a dernièrement remplacé l'annuaire de l'Ile des Médias par Looksmart France, un accord pan-européen avec cet outil ayant été signé pour tous les sites d'Altavista. Et Spray France est voué à disparaître, suite au rachat de la société par Lycos et à l'extinction de la marque en France.

Informations sur les sites

C'est certainement sur Lokace (en tenant compte du fait que, sur ce portail, les résumés sont réécrits spécifiquement) que l'on trouve le meilleur descriptif des fiches présentant les sites sur l'annuaire de l'Ile des Médias.

Les champs suivants sont renseignés (mais pas obligatoirement tous utilisés et/ou affichés par les clients diffuseurs de l'annuaire) :

- La taille du site (petit / moyen / grand)
- La fréquence de mise à jour
- L'adresse de messagerie électronique générique du site
- La présence d'un système de communauté (chat, forum...)
- Le fait que le site propose des fonctionnalités de e-commerce et de shopping
- Le type de site (société, site perso, association...)
- Le pays d'origine (France, Belgique, Canada francophone...)

Sur Lokace, un lien permet également de classer les résultats d'une rubrique par popularité. Cette donnée est calculée sur la base du nombre de clics enregistrés sur une période d'un mois glissant.

Les neturfers n'ont pas la possibilité d'indiquer des mots clés additionnels, de façon transparente, sur la fiche descriptive d'un site. Ils peuvent, cependant, le faire sur l'intitulé d'une rubrique (situation similaire à celle de Yahoo! France).

Modèle économique

L'annuaire de l'Ile des Médias est donc partagé entre plusieurs clients. Le prix demandé par l'annuaire est basé sur un coût au CPM ou sur la base d'un pourcentage sur les rentrées publicitaires des pages du site co-brandé (dans ce cas, le plus souvent, c'est le diffuseur de l'annuaire qui gère lui-même la régie de publicité). En règle générale, ce coût est de 26 à 40 F le CPM (coût aux milles pages affichées), selon l'audience du site diffuseur. Pour terminer, on peut indiquer que, tous sites confondus (hormis celui d'Excite France qui est géré en direct par le client), l'annuaire génère un trafic de plusieurs millions de pages vues par mois, ce qui en fait un acteur majeur du paysage de la recherche d'information en France.

■ Gros plan sur la société La Mine

Présentation

La société La Mine (<http://www.lamine.com/>) fournit l'annuaire du site portail MSN, appelé "Guide du Web" (<http://search.msn.fr/>). Il fournissait également, il y a encore quelques mois de cela, celui d'Ecila avant que ce site ne s'oriente vers l'Open Directory. La Mine a en outre participé à plusieurs services comportant des annuaires spécialisés comme Femme Online, Direct-Way, Shopsmart, Consomania et France.internet.com. Cet article complète celui consacré à MSN Search, dans le numéro de novembre dernier de cette lettre.

La Mine emploie, à l'heure actuelle, 43 personnes et a été créée au début de l'année 1997, notamment par Vincent Fournout qui travaillait en 1996 sur certains forums du service on line MSN. Un dialogue constant entre Vincent Fournout et Microsoft a abouti à la création du Guide web de Microsoft dont la première mouture date de 1996, dans le cadre d'un partenariat fort ayant pour objectif de mettre en place un contenu à forte valeur ajoutée destiné notamment au site Internet de MSN.

Organisation

La société La Mine est organisée en trois grands pôles de production : B2C (voyages, sport, etc.), Shopping et B2B (finances, entreprises, etc.). Une quatrième division, transversale, s'occupe notamment de la veille : collecte des informations, traitement, diffusion vers MSN ou d'autres clients des contenus éditoriaux.

Le coeur des préoccupations de La Mine est le contenu, bâti spécifiquement pour le Web. Au départ, des sites comme Metanews, France.internet.com, puis Consomania ont été créés et sont maintenus par une rédaction comprenant à l'heure actuelle 35 rédacteurs/journalistes. La Mine produit également du contenu éditorial pour d'autres clients comme Wanadoo, Usinor, OK Assurance, Canal Satellite, etc.

L'annuaire de MSN contient aujourd'hui 60 000 sites. Il comprend non seulement l'annuaire "classique" proprement dit, mais également de nombreuses informations pour les chaînes thématiques du site MSN ainsi que pour ses déclinaisons internationales. Chaque rubrique de l'annuaire comprend au maximum 30 à 40 sites, une nouvelle catégorie étant créée ensuite, si trop de sites apparaissent dans une catégorie spécifique.

Les critères d'acceptation et de refus sont très classiques (voir les articles dans les lettres précédentes sur Yahoo.fr, Lycos.fr, Nomade.fr et le Guide de Voila), mais un accent est surtout mis sur la crédibilité de la source (qui fournit l'info ?) et l'originalité du contenu proposé.

Une soumission est traitée en moyenne en une semaine. Aucun mail n'est envoyé, que ce soit en cas d'acceptation ou de refus d'un site. Il faut donc aller sur le site de MSN vérifier si le site a bien été pris en compte. Le site reçoit en moyenne 1000 à 2000 soumissions par jour ce qui est énorme. ce chiffre serait, de plus, en phase de croissance constante.

Des "ménages" sont périodiquement effectués, certains sites considérés comme obsolètes ou moins intéressants sont ainsi enlevés de la base. Les sites répondants à des erreurs 404 ou 403 sont également enlevés. Vérifiez donc périodiquement que votre source d'information est encore présente.



Vincent Fournout, fondateur de la société La Mine

dans l'annuaire et resoumettez-le éventuellement. La Mine gère également une "black-liste" pour les sites qui auraient tenté de spammer l'annuaire.

L'objectif de La Mine et, donc, de MSN, est de proposer aux internautes un annuaire intéressant au niveau qualitatif, pas obligatoirement au niveau quantitatif (avec 60 000 références, l'annuaire de MSN propose la moitié du nombre de sites indexés par des outils comme Yahoo.fr ou Nomade.fr) : leur credo est d'avoir "de plus en plus de bons sites, pas obligatoirement de plus en plus de sites".

Le profil des intervenants sur l'annuaire est à 35% composé de documentalistes, à 50% de journalistes et à 15% de passionnés, spécialistes d'un domaine précis et notamment des sciences humaines.

Enfin, La Mine s'occupe également de la partie "Recherche d'experts" baptisée "Nous cherchons, vous trouvez" sur MSN, dont le but est de fournir des pistes de recherche aux internautes (voir notre article sur MSN dans la lettre "Réacteur" de novembre dernier).

En gros, ce service traite les demandes des internautes selon les pourcentages suivants :

- Un quart de demandes professionnelles
- Un quart de demandes grand public précises
- Un quart de demandes très générales
- Un quart de demandes de rencontres privées plus ou moins tendancieuses (eh oui...).

Le délai de réponse est officiellement de 48 heures, mais il est le plus souvent anticipé par les équipes de La Mine. Enfin, La Mine a lancé, il y a quelques temps, le site "Paye per search" (<http://www.paypersearch.com>), qui n'a jamais vraiment été réellement promu. Cependant, ce site génère quand même des demandes de recherche d'information chaque semaine. Il se pourrait qu'il retrouve, dans les mois qui viennent, une "nouvelle jeunesse"...

■ Looksmart France : naissance d'un nouvel annuaire francophone

Présentation

Depuis quelques jours, l'annuaire Looksmart (ou plutôt BTLooksmart, fruit d'une joint-venture entre BT et Looksmart pour l'Europe) France, dernier né des annuaires francophones (et qui sera officiellement lancé à la fin du mois de janvier) remplace l'Ile des Médias comme guide des sites sur Altavista France. Looksmart France propose une base de données de 125 000 sites web francophones et était déjà présent depuis quelques semaines sur les sites suivants :

- Altavista France (<http://www.altavista.fr/>)
- Club Internet (<http://www.club-internet.fr/>)
- Worldonline (<http://www.worldonline.fr/>)
- Freesurf (<http://www.freesurf.fr/>)

Première constatation : le nombre de grandes rubriques proposé sur les pages d'accueil n'est pas le même d'un portail à l'autre : 13 sur la plupart des clients de Looksmart France (Club-Internet, Worldonline, Freesurf), mais seulement 4 sur la page d'accueil d'Altavista.fr. En fait, on a bien accès sur cet outil de recherche à l'annuaire Looksmart au grand complet, mais seulement à partir de cette adresse : <http://fr.altavista.com/cgi-bin/query?pg=dir&cn=fr&cl=fr> La situation n'est certainement que temporaire. Une soumission de site n'est cependant possible que sur ce moteur (aucun formulaire n'est disponible sur les sites de Club-Internet, Worldonline et les autres).

Description de l'annuaire

Looksmart France propose, en temps normal, 13 grandes rubriques :

- Achats et services
- Actualité et médias
- Arts et culture
- Economie et commerce
- Education et emploi
- Famille et maison
- Informatique et Web
- Loisirs et Discussion
- Santé et mise en forme
- Sciences et références
- Société et politique
- Sports et plein air
- Tourisme et voyage

Le "déroulé" des rubriques et sous-rubriques est assez classique. La version d'Altavista France propose cependant, là encore, une différence par rapport à Club Internet et Worldonline : l'annuaire d'Altavista propose une liste de sites intermédiaires, dans des rubriques proposant des liens vers des sous-rubriques, alors que, sur les autres moteurs, seule une liste finale de sites est proposée, sans aucune indication à l'intérieur des catégories intermédiaires.

D'autre part, à l'intérieur d'une catégorie finale, les sites ne sont pas classés de façon alphabétique. Plus d'infos sur ce point le mois prochain.

Altavista, Club Internet et Worldonline proposent également un module de saisie de mot clés sur l'annuaire (Altavista propose, en fait, les résultats de l'annuaire avant ceux de son moteur de recherche, ce qui est une petite révolution chez cet outil qui avait, jusqu'à maintenant, toujours privilégié les résultats de type "moteur" sur ceux des annuaires). Chez Club-Internet et Worldonline, les classements des sites proposés sont identiques d'un outil de recherche à l'autre. Altavista France propose, de son côté, un autre classement, car il prend certainement en compte un certain nombre de critères de pertinence qui lui sont propres (issus de sa partie "moteur").

Sont présentés, de façon classique, dans les fiches descriptives de site : le site, un résumé ainsi que la catégorie à laquelle appartient la source d'information. Pour des mots clés génériques, les catégories les plus pertinentes sont également affichées. Bon, il reste encore quelques petites imperfections à régler (la mention "Next 10" sur un outil de recherche francophone... ;-) mais l'annuaire semble aujourd'hui bien au point pour son lancement officiel.

Soumission sur Looksmart France

Pour 159 euros (soit 1043 FHT, et donc 1247,43 FTTC), Looksmart France propose l'offre "Soumission Express", assez similaire aux offres américaines et anglaises, déjà disponibles. Voici quelques informations complémentaires sur cette offre :

- Il n'existait pas, dans un premier temps, d'offre gratuite, seule l'offre payante a été tout d'abord disponible. Une offre gratuite a ensuite vu le jour (http://submit.looksmart.co.uk/submit_non.asp?isp=ut) pour les associations et organismes à but non lucratif.
- Contre cette somme, il est offert une garantie que le site sera évalué par les netsurfeurs sous les 5 jours ouvrés suivants.
- S'il est accepté, le site sera classé dans les catégories les plus appropriées de l'annuaire.
- Il n'est possible d'effectuer une soumission que dans une seule catégorie. Mais les netsurfeurs peuvent décider, par la suite, d'inscrire le site dans

plusieurs d'entre elles, éventuellement.

- Les sites acceptés le sont pour une durée illimitée et sont visibles dans le répertoire deux semaines après leur indexation dans la base de données. Que le site soit accepté ou refusé, un e-mail est envoyé à la personne ayant soumis le site.

- Le fait d'être inscrit au travers de l'offre proposée sur Altavista France inscrit le site sur l'annuaire fourni à tous les autres clients de Looksmart France (Club-Internet, Worldonline, etc.).

- Il ne s'agit en rien d'une garantie d'inscription, mais bien d'une prime à la rapidité de prise en compte du site de la part de l'outil de recherche. - De même, il ne s'agit pas d'une proposition de positionnement du site dans les résultats d'une recherche.

- La somme versée n'est pas remboursée si le site est refusé. Cependant, dans un délai de 30 jours, il est possible d'effectuer une nouvelle soumission par mail à l'adresse correction-fr@looksmart.net, après avoir corrigé certains points "noirs" du site ayant entraîné le refus (liens cassés ou zones en cours de travaux par exemple).

Les critères d'acceptation (http://submit.looksmart.co.uk/submit_guide.asp?isp=ut mais également développés à l'adresse http://submit.looksmart.co.uk/submit_faq.asp?isp=ut) sont très classiques pour un annuaire. Il est également indiqué que Looksmart France fera régulièrement des "ménages" sur son outil de recherche et qu'un site ne répondant pas à la charte de qualité de l'annuaire peut être enlevé de sa base de données.

La FAQ sur l'offre "Soumission Express" de Looksmart France : http://submit.looksmart.co.uk/submit_faq.asp?isp=ut

■ 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, Hotbot (donc Inktomi), Google, Excite US, Fast.

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) : Ecila, Northern Light, Excite (pour les sites en français), Altavista France, Lycos.

Les moteurs à la peine ce mois-ci : Infoseek, Voila, 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 ;-)

-> Fin janvier, la société La Mine aurait l'intention de lancer un livre blanc traitant du contenu sur l'Internet en général et le Web en particulier. Cet ouvrage ferait 200 pages et serait destiné en priorité aux gros clients et prospects de la société, une version serait également téléchargeable sur le site de la Mine par la suite.

-> Infoseek (Go.com) semble ne plus prendre en compte les sites qui sont soumis au travers de la procédure de soumission gratuite. Il faut donc payer pour être présent dans ce moteur. Mais, comme cet outil s'est recentré sur les domaines du loisir aux Etats-Unis, on peut douter des retombées positives éventuelles d'un référencement sur Go.com pour un site francophone, à moins qu'il ne soit spécifiquement ciblé vers les internautes américains.

■ En Bref

-> Une version 4 du logiciel eCatch (<http://www.ecatch.com/>), aspirateur de sites évolué, devrait sortir avant l'été. Il sera plus fiable et l'organisation des données collectées sera organisée sous forme de base de données offrant plus de fiabilité pour les corpus (les "agents") composés de nombreux sites et supérieurs à 10 000 pages. Cette version proposerait d'autres améliorations plus ponctuelles. L'outil est utilisé par environ 1500 utilisateurs payants et 10 000 utilisateurs gratuits.

