

## ■ Infoclic : bienvenue chez l'Ask Jeeves français

Infoclic a été conçu et imaginé par deux amis, Frédéric Plais et Renaud Lacroix, en 1999. Ce dernier s'occupait jusqu'alors de veille Internet chez Peugeot. Les deux fondateurs sont issus de Sup de Co Paris. La genèse de l'outil est basée sur une réflexion autour de la recherche d'information et des besoins des internautes dans ce domaine. Dans un premier temps, les deux compères avaient imaginé travailler autour des forums de discussion. Puis, ils découvrent le site Ask Jeeves (<http://www.ask.com/>) et se disent qu'il s'agit de façon évidente d'une voie d'avenir. Aussitôt dit, aussitôt fait (ou presque ;-)). En décembre 99, la société est créée et le site est mis en ligne en janvier 2001. Aujourd'hui, 9 personnes travaillent à la mise en place et à la maintenance de l'outil. En mai 2001, le site accueillait 15 000 visiteurs uniques pour un total d'environ 200 000 pages vues. Ce même mois, 35 000 questions ont été posées sur le site, avec une vague d'interrogations sans précédents sur... devinez quoi ? Eh oui, vous avez gagné : Loft Story (voir plus loin dans cette lettre) !

Une étude, menée il y a quelques mois et portant sur les outils de recherche Internet, indiquait que l'interrogation en langage naturel (le fait de poser ses questions sous la forme de phrases intelligibles comme "Quel temps fait-il à Paris aujourd'hui ?") avait grimpé, en quelques mois, de 0% à 11% chez les internautes. Pourquoi une augmentation si brutale ? Grâce à l'avènement d'un outil intéressant qui s'appuie sur ce type de technologies : Ask Jeeves, la "muse" d'Infoclic.

Pendant longtemps, on s'est désespéré de voir arriver dans l'Hexagone un équivalent à cet outil. Infoclic a donc comblé ce manque. Basé sur un concept très proche de celui d'Ask Jeeves, il permet de saisir une question en langage naturel. Puis, sur la page de résultats, plusieurs questions nouvelles, basées sur celle posée auparavant, sont proposées. Pour chacune de ces questions affinées, une seule réponse (un seul site) est affichée en cliquant sur un bouton adéquat.

Le but d'Infoclic est de répertorier une liste - la plus grande possible - de questions, puis d'y apporter une réponse et une seule de façon automatique. Objectif : lorsque l'internaute pose sa question, l'outil tente de la rapprocher d'une interrogation existante qui a déjà trouvé une réponse. C'est cette liste de questions "connexes" qui est proposée.

La base de données de questions et de réponses uniques commence à être conséquente pour un site si récent (il en répertorie 430 000 à l'heure actuelle avec un objectif d'un million à moyen terme). Bien entendu, elle n'est pas encore parfaite et vous risquez d'être parfois déçu car vous n'aurez pas trouvé chaussure à votre pied. Mais dites-vous que les responsables de la base de données d'Infoclic auront identifié votre question, posée en ligne, et auront vérifié qu'elle n'a pas réellement donné lieu à une réponse pertinente. Il y a de fortes chances pour qu'ils s'attendent alors à la tâche pour le prochain internaute ayant la même interrogation...

Il est également possible de consulter les réponses de Voila, Google et Altavista ou de demander conseil à un expert par le biais d'un partenariat avec le site Woonoz. Il s'agit là de simples partenariats, mais des discussions sont en cours avec certains annuaires francophones pour intégrer directement leurs résultats dans les pages de réponses d'Infoclic. Objectif : fournir une info avec le moins de clics possibles de la part de l'internaute.

### Structure de l'outil

-----

L'outil Infoclic est basé sur deux "briques" essentielles : lorsque l'internaute pose sa question en langage naturel, celle-ci est prise en charge par l'outil "Efficient Info". Il s'agit d'une application développée en utilisant les technologies Com+ de Microsoft. Celui-ci a la charge d'extraire de la question les mots importants et significatifs permettant de résoudre la question posée. Une fois ces termes extraits de la question, ils sont fournis à la deuxième "brique", le logiciel "Intuition" de la société Sinequa (<http://www.sinequa.com/>). Intuition va alors interroger une base de données de format propriétaire des 430 000 questions/réponses pour identifier celles qui sont le plus proches de la demande de l'internaute.

L'outil "Efficient Info" a pour objectif de mieux comprendre la question de l'internaute. Il utilise pour cela :

- Une base de synonymes.
- Une base d'expressions (par exemple des titres de films, de livres, de CD qui seront alors fournis intégralement à Intuition).
- Un système de création de formulaire pré-établis (cartes géographiques, calculs d'itinéraires, formulaires SNCF, codes postaux, etc. Le système comporte plus de 15 possibilités différentes).
- Un système d'épuration de mots morts, vides de sens ("je voudrais acheter", "quel est le prix de", "je recherche une information sur", etc.) ou d'"expressions types" (pas obligatoirement vides de sens) pour lesquelles il est indiqué au moteur qu'elles sont synonymes d'autres expressions déjà référencées.

L'outil permet également depuis peu un "plus" en termes d'interactivité (il répond "bonjour", "bonsoir", etc.) et renvoie directement au site demandé si la question posée se présente sous la forme d'une url ([www.abondance.com](http://www.abondance.com)), comme cela arrive parfois.

Une fois la question épurée et traduite sous une forme adéquate, elle est fournie au logiciel Intuition de Sinequa qui est chargé de trouver sa correspondance dans la base de données Infoclic des couples questions/réponses qui comprend 15 champs de recherche distincts.

### Création de la base de données

-----

La base de données de questions/réponse est bâtie selon plusieurs schémas et stratégies différents :

- La prise en compte d'un site très riche (par exemple une encyclopédie ou un site référence comme celui de Monsieur Cinéma) qui va permettre de créer bon nombre de couples "questions/réponses" en peu de temps (la liste des pages qui décrivent tous les plus grands sportifs mondiaux sur l'encyclopédie Hachette, par exemple, va fournir des dizaines de réponses en quelques minutes). Dans ce cas, le neteur ne répond pas obligatoirement à une question déjà posée sur le site mais enrichit la base de données avec des questions susceptibles d'être demandées dans l'avenir. Le but est, également, de choisir des sites-référence qui seront encore à jour et disponibles dans un an ou plus, pour éviter un trop gros travail de mise à niveau ultérieure des questions/réponses. Un logiciel de détection des erreurs 404 a été développé également.

- Un travail plus approfondi sur la base d'une recherche personnelle de la part d'un neteur. Par exemple, lorsque je suis passé dans les locaux d'Infoclic, l'un des neteurs avait pour tâche de répertorier tous les monuments et lieux intéressants touristiquement en Grèce afin de bâtir le plus de questions/réponses sur ce thème, toujours très demandé. Les outils utilisés sont, alors, très divers : le site LonelyPlanet (<http://www.lonelyplanet.com>) pour les indications touristiques sur le pays, puis Google pour approfondir chaque site touristique identifié et obtenir la meilleure réponse possible.

- La lecture des questions déjà posées par les internautes : celles-ci sont automatiquement révisées quotidiennement par une personne qui juge la réponse qui a été apportée au visiteur : Pertinent (tout va bien) / Non pertinent (à revoir) / A compléter (une source d'information globale est proposée à l'internaute, mais la réponse nécessite cependant une recherche de sa part au sein du site indiqué) / Poubelle (question ne méritant pas de réponse, allons, un peu d'imagination ;-)...). Chaque couple question/réponse est catégorisé dans un thésaurus qui pourra éventuellement permettre de vendre des bandeaux publicitaires ayant trait à un domaine précis. Sur la page de résultats, les couples questions/réponses sont affichés par ordre décroissant de pertinence (une "note" est appliquée automatiquement par l'outil Intuition de Sinequa à chaque réponse, mais cette note peut être modifiée manuellement pour changer un ordre de réponse si la question est à nouveau posée dans le futur).

- Actualité : Des évènements comme Roland-garros, la Saint Valentin, le décès d'une personnalité, l'émission de télévision LoftStory etc. génèrent bon nombre de questions. Les réponses adéquates doivent donc être imaginées au préalable par l'équipe d'Infoclic et intégrées au plus vite dans la base. Une personne de l'équipe consulte l'actualité chaque matin pour répondre à ce type de problématique.

- Les membres : en règle générale, les questions "isolées" (prenant 7 minutes ou plus pour identifier une réponse) ne sont pas traitées directement par l'équipe de neturfers d'Infoclic, mais intégrées plutôt dans les interrogations proposées aux "membres" (<http://www.infoclic.fr/membres.asp>). Ceux-ci ont la possibilité de visualiser une liste de questions sur le site, peuvent y répondre et gagnent, lorsque leur réponse est pertinente, des points qu'ils peuvent convertir en place de cinéma ou en cadeau (tee-shirt, voire produits de beauté pendant la fête des mères par exemple) chaque fois qu'ils collectent 10 bonnes réponses. Le recordman a répondu à 180 questions dernièrement. Certainement un amoureux de cinéma.. ;-)

La pertinence des réponses des membres doit répondre à chacun des critères suivants :

- La page doit répondre entièrement et précisément à la question.
- Il faut donner l'adresse précise d'une page qui répond spécifiquement à la question.
- La page doit correspondre à la charte qualité d'Infoclic (<http://www.infoclic.fr/charte.asp>).
- La page doit être en français.

#### Pertinence

-----

Infoclic a développé un outil qui permet de mesurer la pertinence des réponses fournies aux internautes. Voici les chiffres relevés depuis le début de l'année :

Janvier 2001 : 59,75% de réponses pertinentes dont 29,7% restant à compléter  
Février 2001 : 62,4% de réponses pertinentes dont 15,9% restant à compléter  
Mars 2001 : 66,72% de réponses pertinentes dont 16,3% restant à compléter  
Avril 2001 : 69,1% de réponses pertinentes dont 13,1% restant à compléter  
Mai 2001 : 74,1% de réponses pertinentes dont 13,4% restant à compléter

L'objectif est aujourd'hui fixé à 80% de réponses pertinentes.

Point intéressant à noter : au vu des questions posées sur l'outil par les internautes, ceux-ci utilisent vraiment un langage naturel pour poser leurs questions et structurent leurs phrases sans opérateurs booléens ni mots clés isolés : ce sont vraiment, et de façon majeures, des phrases intelligibles qui sont proposées à l'outil de recherche.

#### Business Model

-----

Le business model envisagé par Infoclic n'est pas, pour l'instant, de gagner de l'argent avec le site Infoclic.fr lui-même (comme peut le faire Ask Jeeves qui propose bon nombre de solutions marketing/promotion de site autour de son outil), mais plutôt de revendre sa solution en marque blanche auprès de sociétés, notamment pour leur service après vente de premier niveau (call centers). En effet, ces entreprises ont souvent à faire face à une problématique de questions le plus souvent toujours identiques pour lesquelles un outil comme Infoclic est parfaitement adapté (voir l'exemple mis en place sur la page d'accueil du site Abondance). C'est d'ailleurs également l'un des business models majeurs d'Ask Jeeves aux Etats-Unis. Notons que, malgré de nombreux bruits d'adaptation de ce site à la France dans les mois précédents, celui-ci ne s'est toujours pas installé dans l'Hexagone. Pour l'instant, la situation semble gelée de ce côté-là.

#### Référencement

-----

Vous pouvez proposer votre site à l'équipe de neturfers d'Infoclic au travers du lien "Référencer votre site" de la page d'accueil du site. La page proposée (<http://www.infoclic.fr/partenariats.asp#referencement>) vous permet d'indiquer l'url de votre site ainsi qu'un commentaire explicitant en quoi il peut répondre aux interrogations des internautes potentiels sous la forme d'une question suggérée. Les sites référencés reçoivent un mail pour les prévenir qu'ils ont été intégrés dans la base.

#### Conclusion

-----

Bref, un outil qui semble bien né mais dont le succès futur sera certainement en corrélation avec l'agrandissement et le suivi de sa bases de données de questions/réponses. En effet, une réponse pertinente aujourd'hui ne le sera peut-être pas obligatoirement demain. Et plus le nombre de couples "questions/réponses" augmente, plus la base de données devient complexe à maintenir à jour. A tester absolument, en tout cas, et on leur souhaite bonne chance pour la suite...

#### Contacts

-----

Infoclic  
204 rue de Crimée  
75 019 Paris  
Téléphone : 01.55.26.42.80  
Mail : [presse@infoclic.fr](mailto:presse@infoclic.fr)  
Web : <http://www.infoclic.fr/>

#### ■ Comment mieux chercher sur Yahoo! France ?

On ne présente plus Yahoo! ni sa "filiale" française, bien entendu. Le mieux est de fournir, à son sujet, quelques chiffres :

- 140 000 sites référencés dans plus de 35 000 catégories.
- 3 224 000 visiteurs uniques par mois.
- Nombre de pages vues sur Yahoo! France au mois de mars 2001 : 487 179 412.
- Moyenne de temps passé sur le site par visiteur unique : 23,8 minutes.
- Taux de couverture : 40,9%.
- Nombre d'internautes en France de 15 ans et plus s'étant connectés sur Yahoo! France au cours des 12 derniers mois : 8 469 000.
- Notoriété assistée auprès des internautes de 15 ans et plus : 72%.

Des chiffres qui parlent d'eux-mêmes et qui consacrent - on s'en doutait un peu ;-)- Yahoo! France comme un outil de recherche majeur dans l'Hexagone. Il nous a donc semblé important de consacrer un article à la façon de mieux chercher l'info sur cet annuaire. Voici quelques "billets", fournies par l'annuaire (merci à eux), qui devraient vous aider :

Gestion des accents et de la casse des mots

-----  
- Yahoo! France ne gère pas les lettres accentuées. Plus exactement, la recherche est identique, quelque soit l'orthographe du mot demandé (accentué - non accentué) s'il n'y a qu'une erreur d'accentuation. S'il y a deux erreurs d'accentuation dans le mot demandé, les autres occurrences ne seront pas trouvées. Je ne doute pas que vous ayez parfaitement compris la phrase précédente, mais je vais quand même donner un exemple ;-):

jérôme trouve jérôme, jérôme et jerome  
jérôme trouve jérôme, jérôme et jerome  
jérôme trouve jérôme, jérôme et jerome

MAIS :  
jêrôme ne trouve ni jérôme, ni jérôme ni jerome  
car le mot contient deux erreurs d'accentuation (ê à la place de é et ò à la place de ô).. Ceci dit, deux erreurs d'accentuation dans un même mot, c'est quand même assez rare...

Yahoo! France connaît quelques problèmes, parfois, avec le tréma (ö) car le logiciel de recherche utilisé sur saisie de mots clés est identique pour toutes les versions "régionales" de Yahoo! et les allemands ont demandé un traitement spécifique pour ce caractère diacritique qui peut, parfois, induire en erreur l'algorithme lorsqu'un terme contenant un tréma est demandé. Dans ce cas, saisissez le mot en non accentué, vous n'aurez pas de problèmes.

- Yahoo! ne tient pas compte de la casse : vous pouvez saisir vos mots en minuscules ou majuscules, le résultat sera tout le temps identique.

#### Opérateurs booléens

-----

- Actuellement, l'espace est géré sur Yahoo! comme un ET strict (ET forcé) : s'il y a des résultats dans la base de données de Yahoo!, ils sont affichés, sinon, il passe la main à Google. Un OU n'est pas pris en compte si aucun résultat contenant tous les mots indiqués n'est trouvé. Ceci dit, il se peut que cette situation ne soit que transitoire (période de test) et qu'on revienne, comme il y a quelques semaines, à un algorithme du type :

\* D'abord un ET est appliqué.

\* Si il y a des résultats, ce sont eux qui sont proposés en premier.

\* Puis les résultats du OU sont affichés. ceux-ci sont affichés en premier s'il n'y pas de résultats avec le ET.

Pour l'instant, il n'en est rien. Affaire à suivre en tout cas...

Le ET peut sinon être demandé sous la forme du signe +. Toutes les syntaxes, avec ou sans espace entre les mots, fonctionnent :

+moteur +recherche

+moteur+recherche

Si l'on tape :

moteur +recherche

ou :

moteur+recherche

Logiquement, le mot moteur devient accessoire, mais comme, par défaut, c'est un ET qui est pris en compte actuellement lorsqu'on met un espace, cela revient au même que :

+moteur +recherche

Il pourrait en être différemment lorsque l'algorithme aura changé et passera à un ET+OU pour l'espace comme indiqué ci-dessus. Voici un article qui commence dans la simplicité la plus totale... ;-)

- La troncature est effectuée par défaut pour les mots à partir de 4 caractères :

capl trouve capitaine et capitale

cap ne trouve que cap. Si on veut "tout ce qui commence par cap", il faut utiliser l'astérisque : cap\*

- Les expressions peuvent être recherchées en utilisant les guillemets ou le tiret. Les requêtes suivantes sont identiques :

paris-dakar

"paris dakar"

- L'ordre des mots est neutre : la recherche sur :

paris dakar

donnera le même résultat que si vous demandez :

dakar paris

#### Recherche avancée

-----

La recherche avancée de Yahoo! France (<http://fr.search.yahoo.com/search/fr/options?p=>) semble ne pas fonctionner : le résultat est identique quelle que soit l'option de recherche : "Laisser le moteur décider", "Rechercher avec l'expression telle quelle", "Rechercher avec tous les mots" ou "Rechercher avec au moins l'un des mots...". Bizarre... Encore mieux, le résultat est toujours indentique avec les options "les catégories uniquement" ou "Tout Yahoo!" et n'enlève que les résultats de type "catégories" pour les options "les sites uniquement" et "les dépêches d'actualité" (qui affiche les sites quand même)... Bref, pas la peine d'utiliser la recherche avancée pour effectuer ce type de filtre...

En revanche, la recherche sur les dates d'entrée dans la base fonctionne, ainsi que le nombre de résultats par page.

Bon, ceci dit, moins de 0,2% des utilisateurs du site utilisent la recherche avancée sur le site. On ne peut donc pas dire que ce chantier soit prioritaire chez Yahoo! France ;-)

#### Recherche d'information sur Yahoo! France

-----

Cela nous donne l'occasion de donner quelques chiffres sur l'utilisation de l'outil par les internautes. La recherche d'information sur l'annuaire représente environ 20% du trafic global à l'heure actuelle. En règle générale, la part du "search" décroît en pourcentage sur Yahoo! France au profit des services thématiques (mais Yahoo! propose de plus en plus de services thématiques, parfois proches et très complémentaires du "search" comme l'encyclopédie).

Ainsi, Yahoo! propose depuis peu, en haut de la page de résultats, un bandeau qui permet d'effectuer une requête sur d'autres zones du portail :

Recherchez "mots clés demandés" dans les services Yahoo! suivants :

Encyclopédie - Actualités - Sport - Musique - Clubs - Shopping - Enchères

Ce qui contribue également à "faire fuir" les internautes de l'annuaire pour aller vers les zones thématiques de l'outil.

Cette barre est toujours la même, quelle que soit la recherche, mais un projet de contextualisation de cette information est à l'étude.

Dans la partie "recherche d'information sur l'annuaire", on peut dire que sur 10 internautes :

- 4 tapent des mots clés et aboutissent à des résultats issus de la base de données de Yahoo!
- 2,5 tapent des mots clés et se voient proposer des résultats issus du moteur Google.
- 3,5 naviguent dans l'arborescence pour arriver à la bonne catégorie.

Une immense majorité d'internautes ne tapent qu'un mot clé. Ceux qui saisissent deux mots demandent, la plupart du temps, une expression consacrée (Paris Dakar, Loft Story, etc.). Les expressions sur trois termes et plus sont quasiment inexistantes dans les logs.

Une immense majorité également, ne va pas au delà de la première page de résultats. Très très peu de gens affichent la deuxième page (et les autres, bien sûr).

Enfin, selon les tests qu'a pu faire Yahoo!, une minorité d'internautes fait du "scroll" sur la page de résultats pour la consulter en entier à l'aide de l'ascenseur. La plupart des visiteurs se satisfont de la page présentée "telle quelle", à la longueur de l'écran utilisé, et cliquent sur les premiers liens proposés, qu'il s'agisse de catégorie ou de site (c'est d'ailleurs pour cela que Yahoo! a adopté une nouvelle présentation pour les catégories avec le lien hypertexte présent uniquement sur le dernier terme de l'intitulé de la rubrique).

#### Affichage des résultats dans une catégorie

-----

Lorsqu'on se trouve dans une catégorie donnée, l'affichage des sites se fait encore par ordre alphabétique des sites, mais :

- Les sites avec lunettes sont affichés en premier. Rappelons que cette distinction s'attribue, le plus souvent, uniquement aux sites non commerciaux, pour éviter des problèmes du type "Pourquoi mon concurrent a une paire de lunettes et pas moi ?".
- Le site officiel, pour une marque par exemple, est également mis en exergue, en premier.
- Un projet du type des "most popular sites" (basé, sur yahoo.com, sur les résultats d'indice de popularité fournis par Google) est à l'étude sur le site français.
- L'offre de positionnement "Sponsored sites" (<http://sponsoredsites.yahoo.com/>) est également à l'étude. Plus d'infos en fin d'année peut être...
- Enfin, il n'existe, pour l'instant, pas de limites théoriques au nombre de sites dans une catégorie donnée. Ceci dit, il est, logiquement, plus complexe d'entrer dans une rubrique qui contient 50 sites plutôt que dans une autre qui en propose 2...

#### Affichage des résultats sur saisie de mots clés

-----

Lorsqu'on saisit un mot clé, les critères suivants sont pris en compte pour classer les catégories et sites proposés :

- Pour les catégories :
  - \* Présence du mot demandé dans l'intitulé.
  - \* Présence du mot demandé dans les mots clés ajoutés de façon transparente par les navigateurs à la catégorie (surtout des synonymes).v \*
  - Longueur de l'intitulé (profondeur de l'arborescence).
  - \* D'autres critères sont à l'étude (voir ci-dessous)
- Pour les sites :
  - \* Présence du mot demandé dans le titre.
  - \* Présence du mot demandé dans le résumé.
  - \* Présence du mot demandé dans l'url.
  - \* Critères pris en compte pour les catégories (voir ci-dessus).
  - \* Nombre d'occurrences du mot demandé dans la fiche descriptive.
  - \* Si plusieurs mots demandés, proximité des mots entre eux (notamment dans le résumé).
  - \* D'autres critères sont à l'étude (voir ci-dessous)
  - \* Il est à noter que la présence des lunettes pour un site donné n'a pas d'incidence sur l'algorithme de classement sur saisie d'un mot clé.

ATTENTION : il y a une différence majeure sur l'affichage des résultats suite à la saisie d'un ou plusieurs mots clés, depuis quelques jours : un site n'est plus affiché qu'une seule fois dans la page de résultat, même s'il est présent dans plusieurs catégories de l'annuaire. C'est donc différent d'il y a quelques temps, lorsque 4 occurrences du site étaient proposées dans la page de résultats si le site était référencé dans 4 catégories différentes. Une seule apparition dans cette page de résultat ne signifie donc pas que votre site a été enlevé des autres rubriques. Important...

D'autre part, Yahoo! France (et en fait tous les Yahoo! d'Europe) prépare pour les mois qui viennent pas mal d'innovations, à la fois dans l'affichage des pages à l'intérieur d'une catégorie, mais également sur les résultats du "search" (algorithme de pertinence et affichage des résultats). Toutes ces modifications seront proposées entre cet été et la fin de l'année, normalement. Un article spécial sera proposé dans la lettre "Réacteur" lorsque le nouvel algorithme de classement de Yahoo! sera mis en ligne. Promis ;- ) On devrait avoir du nouveau sur ce point d'ici à la rentrée de septembre.

#### Mise à jour de la base de données

-----

De temps en temps, des "ménages" sont faits sur Yahoo! France pour épurer la base de données, et notamment les sites qui n'existent plus (liens morts), mais c'est assez rare car la plupart du temps, cela se fait de façon plutôt sur réception de mails à l'adresse [webmaster@fr.yahoo-inc.com](mailto:webmaster@fr.yahoo-inc.com) qui signalent ce type d'anomalie. Du coup, le problème est, la plupart du temps, résolu dans la journée. N'hésitez donc pas à signaler les liens morts (surtout s'il s'agit de sites concurrents, arf arf).

La procédure automatique, voire manuelle (élimination de sites trop anciens, pas mis à jour ou obsolètes bien qu'existants) de "ménage" est plus souvent mise en place sur des catégories très "populaires", très demandées, en fonction de la charge de travail des navigateurs.

#### Conseils de recherche

-----

Voici quelques conseils pour être plus efficace dans vos recherches sur Yahoo! France :

- Effectuez, dans un premier temps, plutôt des recherches sur les catégories plutôt que sur les sites lorsque vous tapez des mots clés. Il vaut mieux, tout d'abord, demander des termes qui ne soient pas trop précis trop rapidement, identifier la bonne catégorie, puis aller sur celle-ci pour obtenir une liste plus exhaustive de sites répondants à votre problématique. La liste de résultats "sites" fournie par Yahoo! est toujours moins exhaustive que le contenu d'une rubrique.
- Lorsque vous avez identifié une catégorie qui peut sembler intéressante, n'hésitez pas à élargir votre recherche en cliquant sur le lien supérieur dans l'arborescence (à "monter d'un niveau").
- Utilisez la possibilité du menu déroulant "Recherche uniquement dans cette catégorie", présent sur toutes les rubriques, à droite du formulaire de recherche. Cette fonctionnalité est très peu utilisée, certainement à tort, sur l'annuaire car elle permet de restreindre une recherche à une zone, plus ou moins grande, de l'arborescence et à éliminer des sites qui ne répondraient pas à la demande. On peut noter que cette fonction est présente sous

la forme de boutons radio sur Yahoo.com, ce qui est peut-être plus logique (pourquoi proposer un menu déroulant pour deux choix ?). Cette modification pourrait éventuellement arriver sur Yahoo! France bientôt.

- Choisissez bien vos mots clés. Ne demandez pas des termes trop précis. N'oubliez pas que Yahoo! est un annuaire et qu'il répertorie des sources d'informations, pas l'info elle-même. Ayez toujours à l'esprit que sur un annuaire de ce type, on recherche des sites, pas des pages.

Projets

Les projets en cours chez Yahoo! France sont les suivants :

- Une subdivision d'un grand nombre de catégories dans les deux domaines "pour les particuliers" et "pour les professionnels" est en train de se terminer. Gros boulot depuis quelques mois pour arriver à ce résultat (ce qui explique peut-être un certain ralentissement dans le référencement de nouveaux sites de la part des netsurfeurs).

- D'autre part, la présentation géographique ("Exploration géographique") va devenir plus homogène : la même organisation de l'arborescence sera proposée pour Paris et Brives-La-Gaillarde. La seule différence entre les deux villes est que les catégories qui ne contiendront pas encore de sites (par exemple, pour Brives-La-Gaillarde) ne seront pas affichées. Mais elles seront "disponibles" pour les netsurfeurs de l'outil qui pourront les renseigner à l'envi. Toutes les villes disposeront donc d'une "arbo théorique" identique, dont une partie seulement sera proposée sur l'annuaire.

- Enfin, des offres de soumission payante devraient voir le jour d'ici la fin de l'année...

Peut-on acheter les gens de Yahoo! ? ;-)

On me signale de plus en plus de cas "limites" sur des gros annonceurs ayant demandé des modifications dans l'annuaire, sur leur site ou sur celui de concurrents ? Parano totale ou fond de vérité ? Voici la réponse "officielle" de Yahoo! : "nous recevons quotidiennement des demandes de modifications d'indexation (changement d'url, changement d'activité et donc nouvelle catégorie ou refonte du commentaire) c'est normal et même plutôt bien considéré : les sites évoluent, l'annuaire évolue en conséquence ! Par contre nous ne recevons pas (et n'accepterions pas si cela arrivait) de demande de modification d'un site vers son concurrent... Rumeurs, rumeurs ;-)"  
Voilà, cela semble clair...

#### ■ Copernic 2001 : l'incontournable métamoteur francophone

Cet article a été écrit par Béatrice Foenix-Riou (<http://www.fla-consultants.fr/>, désolé, il y avait une erreur dans la lettre du mois dernier), et est extrait de son excellent livre "Recherche et Veille sur le web visible et invisible", aux éditions Tec&Doc : [https://www.abondance.com/docs/biblio\\_recherche.html](https://www.abondance.com/docs/biblio_recherche.html)

Présentation de Copernic

Copernic est sans doute l'un des agents de recherche les plus connus dans le monde. Et pour cause, la société québécoise qui le développe est l'une des premières à avoir choisi de lancer une version gratuite de son produit (en complément de deux versions payantes) et à faire l'effort de développer une interface en plusieurs langues : en plus de l'anglais et du français, Copernic est depuis peu disponible en allemand, en espagnol, en italien et en portugais ! Les résultats ne se sont pas fait attendre, et le nombre des utilisateurs de Copernic - affichés fièrement sur l'écran d'accueil du site - est à la mesure de son succès : près de 14 millions à la mi-juin 2001 !

À ces conditions attirantes, Copernic ajoute deux qualités primordiales : cet agent est particulièrement convivial et réellement performant. La version gratuite de Copernic 2001 (Copernic Basic) est en fait un métamoteur offline, qui permet d'interroger plus de 90 sources d'information (moteurs, annuaires et banques de données) réparties dans huit domaines, selon leur couverture : le Web mondial (AltaVista, Euroseek, Fast, HotBot, ODP, etc.), le Web en français (AltaVista en français, Ecila, Francite, Nomade, Voilà, Yahoo! France, etc.), les groupes de discussion, les adresses de courrier, les achats de livres, les achats de logiciels et les achats de matériel informatique ; un huitième domaine, lié à une langue ou un pays, est proposé lors de l'installation.

Les possibilités de recherche comme l'interface sont les mêmes que pour les versions payantes, mais celles-ci donnent accès à un nombre de sources bien plus important. Elles permettent en effet d'envoyer une requête à plus de 1 000 sources, classées dans 93 domaines.

En complément des catégories précédentes, diverses "chaînes" sont offertes comme affaires & finances, articles sur les technologies, fils d'agences de presse, journaux, santé, logiciels, dépêches en français, sécurité informatique, sciences, nouvelles technologies, recherche de FAQ, etc. Il est possible de mettre à jour les versions Copernic Plus et Pro depuis le site, en rajoutant d'autres sources ou domaines.

Au-delà de l'éventail des sources, les logiciels Copernic Plus 2001 (39,95 US\$) et Copernic Pro 2001 (79,95 US\$) ont un autre avantage : ils n'affichent pas les classiques bannières publicitaires que l'on retrouve sur Copernic Basic. Le logiciel Copernic Pro a quant à lui une fonctionnalité supplémentaire : il permet de programmer les mises à jour des recherches, avec une fréquence définie par l'utilisateur ; un rapport de recherche est ensuite envoyé par e-mail à un ou plusieurs destinataires, et signale les nouveautés.

Le point fort de Copernic (versions gratuite et payantes) est, incontestablement, la convivialité de l'interface. Alors qu'avec d'autres agents, une lecture un peu détaillée du mode d'emploi est indispensable, l'interface de Copernic est suffisamment claire pour permettre aux internautes pressés de se lancer immédiatement dans la recherche.

L'écran d'accueil propose ainsi une série d'icônes :

L'icône Chercher est destinée à la création d'une recherche ; un clic sur Chercher lance un assistant, qui permet de choisir un domaine de recherche (Web, Web en français, groupes de discussions...), dans une liste plus ou moins étoffée selon la version (Basic ou Plus/Pro).

Le bouton Propriétés de l'assistant de recherche permet ensuite de choisir les sources, parmi celles qui sont prédéfinies pour chaque domaine. On regrettera ici, c'est l'un des rares points faibles de Copernic, qu'il soit impossible d'envoyer une même requête à des sources de différents domaines. On peut interroger simultanément toutes les sources d'un domaine (dans une limite de 32), mais pas celles de chaînes différentes. On peut ensuite écrire les termes de sa requête, en précisant si elle doit prendre en compte tous les mots, un des mots ou l'expression exacte. En cas de doute sur l'orthographe, un dictionnaire est là pour aider l'utilisateur. On peut demander la validation des résultats, ce qui permet l'élimination automatique des liens morts. Pour les internautes pressés, une barre d'outils "Recherche rapide" permet de lancer une recherche utilisant implicitement l'opérateur ET, sans passer par l'assistant ; mais il est alors impossible de configurer sa stratégie de façon précise.

Copernic envoie ensuite la requête aux sources indiquées, puis affiche la liste des résultats obtenus, après les avoir dédoublonnés et triés ; les mots-clés apparaissent en surbrillance. En double-cliquant sur l'un des titres, on affiche la page en question. Sophistication intéressante, la version 2001 propose la nouvelle barre "Aperçu des documents", qui offre dans un petit encadré une copie miniature de la page, permettant de déterminer sa pertinence avant de l'ouvrir.

Parmi les autres fonctionnalités intéressantes de Copernic, on citera :

- L'édition d'un rapport de recherche très agréable à consulter, au format d'une page Web ; il est possible de choisir les documents qui doivent

figurer dans la liste, ainsi que leur classement (pertinence, ordre alphabétique, source...).

Pour chaque page Web identifiée par l'agent, sont indiqués son titre, les premières lignes, l'URL, les moteurs qui l'ont identifiée, la pertinence... Les termes de la requête sont surlignés. Comme ce rapport est au format html, on se connecte d'un simple clic à la page décrite. Enfin, ce rapport peut être envoyé par e-mail aux destinataires de son choix,

- L'option Mettre à jour une recherche (icône Actualiser) ; une icône spécifique signale alors les nouveaux documents,
- La fonction "Raffiner" est intéressante lorsque la recherche est relativement large. Elle permet de télécharger les résultats de son choix sur le disque dur, puis de lancer une requête sur le contenu de ces pages, avec cette fois des critères plus sophistiqués (ET, OU, SAUF, PRES). Cette fonction permet ainsi de réduire les contraintes dues à l'utilisation d'un métamoteur. Mais il faut penser à "nettoyer" régulièrement son disque dur, sous peine de saturation rapide !
- Des commandes et boutons sont intégrés pour résumer les documents, lorsque le logiciel Copernic Summarizer est déjà installé.

A savoir pour optimiser sa recherche

Sources interrogées

Domaine d'exploration : Version Basic : 90 sources réparties dans huit domaines : Web mondial et en français, groupes de discussion, adresses e-mail, achat de livres, de logiciels et de matériel informatique, ainsi qu'un domaine facultatif lié à une langue ou un pays, à choisir dans une liste lors de l'installation du logiciel.

Copernic Plus et Copernic Pro 2001 : plus de 1 000 sources au total, réparties dans 93 "chaînes" (fils de presse, logiciels, articles sur les technologies, affaires & finances, santé...). Choix des sources : Oui. Les icônes Chercher et Modifier permettent d'indiquer le domaine de recherche. On peut alors, en cliquant sur l'onglet Propriétés, choisir dans une liste la ou les sources à interroger. Il est impossible d'interroger simultanément des sources de différentes chaînes.

Les versions Plus/Pro permettent le rajout d'autres sources, depuis le site Web de Copernic.

Paramétrage du nombre de résultats par source : Oui. L'onglet Paramètres des icônes Chercher et Modifier propose plusieurs choix prédéfinis :

Recherche Rapide (10 résultats par source, 100 maximum par recherche), Normale (10/200) ou Détaillée (30/300).

L'option Recherche Personnalisée offre une formule sur-mesure, avec un maximum de 300 résultats par moteur et 3 000 résultats par recherche.

Comment bien interroger

Opérateurs booléens : Oui. Les icônes Chercher ou Modifier permettent de saisir plusieurs mots, en précisant que seuls doivent être sélectionnés les documents comprenant tous les mots (ET), un des mots (OU) ou l'expression exacte. La barre d'outils "Recherche rapide" permet de lancer directement une recherche sur plusieurs mots, reliés implicitement par l'opérateur ET.

Sinon, l'option Raffiner permet de télécharger les documents sur le disque, puis de lancer une requête sur leur contenu en utilisant les opérateurs ET, OU, SAUF, PRES (termes de recherche à au plus 10 mots l'un de l'autre).

Mots composés / phrase : Oui. Les mots doivent être tapés "entre guillemets". Copernic signale les moteurs qui ne comprennent pas les guillemets.

Troncature : Selon les moteurs interrogés, la troncature est faite de façon implicite ou non. Pour une recherche exhaustive, il est donc important de comparer les résultats avec les formes singulier / pluriel des mots. Inversement, il est possible de préciser que la recherche doit se faire sur "l'expression exacte" (icônes Chercher ou Modifier) ou sur les mots "en entier" (option Raffiner).

Ordre des mots : La prise en compte ou non de l'ordre des mots dépend des moteurs interrogés. Une recherche avec deux termes ne donnera donc pas exactement les mêmes résultats, selon l'ordre des termes.

Caractères admis : La prise en compte ou non des majuscules et des minuscules, comme des caractères accentués, varie selon les moteurs interrogés. Pour une recherche sur un mot accentué, il est donc prudent de comparer les résultats en écrivant le mot avec et sans les accents.

Sinon, l'option Raffiner permet de spécifier que la requête doit tenir compte des majuscules

Autres fonctions : Oui. La commande Dupliquer (menu Recherche) permet de copier dans une nouvelle recherche les paramètres d'une recherche existante.

Présentation et gestion des résultats

Critères de classement : Oui. Il est possible de trier les résultats par ordre alphabétique des titres ou de l'URL, par moteur, par pertinence ou selon la date de la recherche (menu Affichage, choix Trier les résultats).

Elimination des doublons : Oui.

Vérification de la validité des liens : Oui, en cliquant sur l'icône Valider.

Indication de la source ayant fourni la réponse : Oui.

Format de visualisation : Titre - avec lien -, première ligne, URL, score, moteur(s) ayant fourni la réponse, type de document (accessible, inaccessible, nouveau, téléchargé, visité...). Les mots de la requête sont surlignés.

Edition personnalisée : Oui. La commande Consulter permet d'éditer un rapport de recherche au format d'une page Web ; il est possible de choisir les documents qui doivent figurer dans la liste, ainsi que leur classement.

Téléchargement des résultats : Oui. La commande Télécharger permet de télécharger sur son disque les documents identifiés, afin de les consulter hors ligne ; on peut également les envoyer par e-mail à une autre personne. Les liens non valides sont éliminés automatiquement au cours de cette étape.

Recherche dans la base des documents téléchargés : Oui, la commande Raffiner permet de télécharger les documents de son choix (pour sélectionner des documents, il faut cliquer sur chacun en utilisant le bouton gauche de la souris, tout en tenant enfoncée la touche Ctrl) ; on peut ensuite formuler sa requête en utilisant les opérateurs ET, OU, SAUF, PRES, les parenthèses...

Mise à jour des recherches / Veille : Oui. L'icône Actualiser permet de relancer une recherche, en signalant spécifiquement les nouveaux documents.

La commande Veille (uniquement dans Copernic Pro) permet de programmer les mises à jour, avec une fréquence définie par l'utilisateur. Copernic Pro se charge de l'ouverture et de la fermeture automatiques de la connexion Internet par modem, au début et à la fin de l'exécution d'un calendrier de veille.

Archivage des recherches : Oui. L'historique des recherches s'affiche sur l'écran principal, avec un certain nombre de renseignements pour chacune ; on peut les trier, les ranger dans des dossiers...

Spécifications techniques : Equipement minimum : 486-66 Mhz

15 Mo de mémoire vive et 10 Mo d'espace disque

Système d'exploitation : Windows 95, 98 ou NT4

Plus d'infos :

<http://www.copernic.com/fr/>

## ■ Le phénomène Loft Story vu au travers des outils de recherche

Loft Story, on aime ou on n'aime pas (bon, moi, personnellement, bof, mais enfin... Ceci dit, j'ai quand même regardé trois quarts d'heure, pour me rendre compte). Mais personne ne peut nier l'incroyable phénomène qu'il provoque à l'heure actuelle, dans la vie réelle et sur l'Internet. Même dans mon village, au fin fond de l'Alsace profonde, on n'entend parler que de ça.

En effet, habituée des records d'audience télévisés, "Loft Story" multiplie aussi les bons scores sur Internet, puisque près de 30% des internautes, soit plus de 1,63 million de visiteurs uniques, se sont connectés au moins sur l'un des sites liés à l'émission, selon les résultats mensuels de Médiamétrie eRatings sur le sujet. Dans un communiqué, Médiamétrie eRatings observe que "quel que soit l'indicateur de mesure choisi, il est

question de Loft Story sur Internet en France au mois de mai 2001".

Le site loftstory.com a enregistré la meilleure progression en pourcentage d'audience en France, avec 507.439 visiteurs en mai, contre 105.271 le mois précédent, soit une hausse de 382%. Loftstory.fr est entré à la neuvième place du palmarès des dix domaines les plus visités en France, 1.446.000 des internautes ayant surfé à domicile étant allés sur le site, ce qui représente 26% des internautes ayant été naviguer sur la Toile en mai. Les pages permettant un lien direct avec l'actualité du "Loft", regroupées dans la rubrique "live", ont été les plus consultées, devant celles présentant les meilleurs moments des derniers jours, rassemblées dans les rubriques "top vidéo" et "news". Quant au site "non officiel" bofstory.fr.st, il a réalisé la meilleure entrée du mois dans le Top des domaines les plus visités en France, en accédant au rang 54.

M6 profite aussi de l'effet Loft en termes de notoriété, puisqu'elle entre en dernière position du palmarès des cinq annonceurs les plus vus par les internautes depuis leur domicile, 30,4% des internautes ayant surfé en mai ayant été exposés à au moins une bannière ou un bouton publicitaire de la chaîne.

Cette enquête a été effectuée auprès d'un échantillon de 8.000 personnes, représentatif de la population des internautes de deux ans et plus résidant en France et disposant d'un accès internet à domicile.

Il était donc intéressant de demander aux outils de recherche comment les internautes posaient leurs questions sur le sujet. C'est également une façon de voir comment ces outils de recherche traitent un sujet d'actualité comme celui-ci.

Les mots clés le plus souvent demandés sur les outils de recherche francophones

Par exemple, Depuis plusieurs semaines, "Loft Story" est la requête la plus tapée sur **Nomade.fr** (voir <http://www.nomade.fr/motsclés/>). Plus fort que "sexe", il fallait le faire ! Une "performance" qui profite également au mot clé "M6", 238ème à la veille de la première diffusion de l'émission et qui se classe actuellement en 3ème position. De son côté "Loana" atteint le 18ème rang et détrône Tibéri (Jean, pas Xavière, quand même) dans les personnalités les plus recherchées sur le Net et se positionne sans concurrente sérieuse. Sa suivante immédiate, Lolita, "coince" 100 places plus loin. Le mot clé "loftstory", en un mot, est également bien classé. Malheur, en revanche, pour les exclus du loft qui n'apparaissent plus dans les 300 requêtes les plus tapées sur Nomade.fr quelques heures après leur éviction. On vit dans un monde sans pitié... En tout, les mots clés sur Loft Story ont représenté environ 2% du nombre total de requêtes sur le mois de mai.

Côté **Lycos France** (<http://www.lycos.fr/webguides/top/>), sur le mois de mai, "loft story" a été demandé 837 000 fois, loftstory 404 000, loana 121 551, david 32 364 et Laure 27 000 fois. Le mot clé bofstory a, de son côté, été demandé à 20 000 occurrences quand même !

Sur **Webhelp**, le volume des questions sur l'émission a représenté 2 à 5% du volume des chats, mais seulement pendant les premiers jours, puis les demandes ont baissé rapidement. Les sujets demandés le plus fréquemment concernent, bien sûr, les vidéos censurées, mais elles ont été classées "recherche interdite" par Webhelp. Autres questions posées très souvent :

- Quel est le soft nécessaire pour voir l'émission ?
- La musique du générique,
- Des photos agrandies et en couleur de Loana
- L'actualité : qui est avec qui, qui a fait quoi ou qui pense quoi ...

Sur **Looksmart France**, les mots clés plus demandés sont, sans surprise, "loft story", "lofstory" et "M6". Ils représentent environ 1 à 2% des requêtes.

Sur **Lokace**, on voit apparaître en 3ème position (derrière "sexe" et "photo") le mot "story" et en 4ème "loft". En descendant un peu, on trouve ensuite "M6" en 14ème position, et juste après "warez", c'est "Loana" qui arrive à la 45ème place. Ensuite, peu de mots relatifs à l'émission, sauf un peu plus bas, en 183ème place où l'on retrouve "David".

Du côté de **Yahoo! France**, voici quelques statistiques sur le nombre de fois où les mots clés relatifs à l'émission ont été demandés pendant le mois de mai 2001 :

loft story (866 217)  
loftstory (524 919)  
loft (65 288)  
loana (62 037)  
"loft story" (31 558)  
bofstory (29 308)  
www.loftstory.fr (19 317)  
bof story (18 500)  
lost story (18 219)  
story (17 671)

Chez **Voila**, les mots clés les plus souvent demandés par les internautes concernant cette émission sont les suivants : si l'on s'en réfère à la journée du 11 juin par exemple, "Loft story" est la requête la plus fréquemment demandée sur le moteur. Ses différentes déclinaisons sont également bien placées dans le top 50, avec "loftstory" en 6ème position, et "loft" en 22ème position. "M6" remonte en 19ème position, et détrône "TF1" qui est relégué à la 44ème position. "Loana", classée en 35ème position, devient la femme la plus populaire dans les requêtes des utilisateurs, après "Clara Morgane". En ce qui concerne les autres habitants du loft, seuls "Steevy" et "Kimy" apparaissent dans les 1 000 requêtes les plus fréquentes sur la période du 26/04 (1er diffusion) au 12/06.

Sur **Altavista France**, 384 807 requêtes en mai ont trait à loftstory et consors, soit 28,7% du nombre des requêtes du mois (les 1000 premières qui représentent 98% du nombre total de requêtes). Ce chiffre dépasse celui des requêtes d'ordre sexuel (25% des requêtes pour le mois de mai), traditionnellement le sujet le plus recherché en nombre de requêtes sur les moteurs. Voici quelques statistiques fournies par Altavista France :

loft story (118 571)  
loftstory (94 270)  
loft +story (57388)  
loftstory (41974)  
loana (21 497)  
m6 (12691)  
loft story (12598)  
loana (8571)

Sur **KartOO**, le phénomène est également très présent. Prenons par exemple pour le 2 mai dernier :

"loftstory" : 445 requêtes (0,24% du total), c'est moins que "sexe" (612 req), mot le plus tapé sur KartOO (non, sans blagues...)

"loft story" : 352 requêtes (0,20% du total) soit un peu moins que "sex"

"loft-story" : 55 requêtes (0,03% du total) à égalité avec "roses" par exemple

"piscine" : 90 requêtes (0,05% du total) soit le même score que "rencontres"

"loana" : 21 requêtes (0,01% du total) mais l'émission n'était lancée que depuis quelques jours...

Bref, le loft représentait déjà, à cette époque, presque 1% du trafic si l'on compte les associations de mots et les requêtes plus rares comme

"loftstoryx", "loftstoryz", "loftconnerie" (si si !),

Un mois plus tard, la situation était la suivante :

"loftstory" est au coude à coude avec le mot "sexe". Si l'on observe les proportions, on s'aperçoit que la demande pour les sites "coquins" a légèrement baissé, contrairement à celles ayant trait au loft. C'est un simple rapprochement de faits ;-).

Certains mots sont de plus en plus demandés, par exemple, les prénoms des occupants du Loft, Loana en tête bien sûr. Quelques associations de mots qui reviennent souvent (pour les référenciers) :

videos loftstory (ces deux mots réunis dominent nettement le classement des associations)

loftstory cameras  
loftstory images  
loftstory loana  
loana nu  
loftstory candidats  
loftstory photos  
loftstory news  
loftstory inedites  
loft photos  
humour loftstory  
loftstory delphine

...

on trouve aussi au hasard des logs :

loftstory poules  
loftstory loftstorysex  
"photos de loana"

entrevue loana kimy

Sur trois mots, on a souvent :

loana edouard piscine

L'association la plus longue est, sans surprise :

loftstory +cameras +photos +pornographie +erotique +gratuit

Sur **MSN France**, le palmarès est le suivant :

Loft story, 6ème

M6, 12ème

TF1, 29ème

Big Brother; 2133ème

Bigdil, 3125ème

Qui veut gagner des millions, 3770ème

Survivor, 6897ème

Loana, 1016ème

Kenza, 37421ème

Les requêtes "Loft Story" et "M6" enregistrent ainsi sur cet outil de recherche les progressions les plus spectaculaires de ces douze derniers mois.

En avril la requête "Loft Story" devient en moins de six jours la 6ème plus populaire sur MSN Search pour venir talonner "sexe" ! Quant à la requête "M6", elle atteint la 12e place du classement. Le lendemain du lancement de l'émission, la requête "Loft Story" était la plus demandée et totalisait à elle seule 22% des requêtes saisies sur MSN Search parmi les 100 plus populaires. Le site de M6 quant à lui n'était pas en reste, capitalisant 15% des requêtes. Pour clore ces résultats, notons enfin que lors de cette même journée de lancement, l'émission doublait le nombre de requêtes habituel du leader du classement. Du jamais vu !

Parmi les lofteurs, seules les filles semblaient intéresser les internautes au mois d'avril dernier. Les participants masculins sont en effet absents des requêtes. Et si Kenza suscite un attrait certain, c'est avant tout Loana qui récolte les fruits de l'incroyable popularité de l'émission. Incroyable ;-)

D'autre part, le débat qui a opposé les dirigeants de TF1 à ceux de M6 ne concerne pas que les réseaux hertziens. En effet, avant le début de l'émission TF1 devançait systématiquement M6 en termes de requêtes sur MSN France, or dans la semaine précédant la première de l'émission, la Petite chaîne a décidément monté pour multiplier le nombre de requêtes la concernant par 36 (vous avez bien lu !) et devancer TF1 au classement général. Par ailleurs, les émissions phare de TF1 (Bigdil et Qui veut gagner des millions ?) semblent également boudées par les internautes... même si la riposte est imminente : Survivor, le projet concurrent qui ne sera diffusé sur TF1 qu'au mois de septembre, intéresse déjà quelques internautes, ainsi que Big Brother, première émission du genre diffusée aux Pays-Bas et aux Etats-Unis. Des requêtes qui traduisent sinon un intérêt durable du moins une curiosité globale pour ce type d'émissions.

Sur **Francité**, le mot clé "loft" a été demandé 2 147 fois, "story" 2 010, "loana" 613 et "laure" 634. Malheureusement, nous ne savons pas sur quelle durée. Cela nous donne quand même un résultat relatif des mots clés entre eux...

Enfin, la folie Loftstory s'est également emparée d'**Infoclic**, l'outil de recherche en langage naturel (voir article dans cette lettre). Le lendemain de la fameuse scène dans la piscine, plus de 15% des questions posées sur le site concernaient le désormais célèbre loft et ses locataires ! Classées par thèmes, vous trouverez ci-dessous quelques unes des questions posées par les internautes :

- Les questions les plus posées :

\* Où trouver les scènes censurées de loft story ?

\* Quels sont les sites parodiant loftstory ?

- Les questions pratiques :

\* Comment empêcher ma femme de regarder Loftstory ?

\* Où trouver la même table basse que dans Loftstory ?

\* Je comprends rien à Loftstory : on vote pour ou contre un candidat ?

\* Comment faire comprendre aux enfants que voter par téléphone à Loftstory ça coûte cher !

- Les questions sur les candidats de Loftstory :

\* Comment se faire Loana, et comment la plaquer le lendemain ?

\* Comment avoir le même regard intelligent que David ?

\* Est ce que Steevy est Gay ?

\* Philippe va t-il créer une start-up dédiée au rangement et au ménage ?

\* Comment trouver d'aussi bonnes cachettes que Kenza derrière une porte ?

\* Delphine a t-elle réellement un enfant ?

\* Pourquoi Aziz s'est fait viré le 1er, c'était le plus cool pourtant ?

\* Comment devenir aussi bon imitateur que Jean Edouard ?

\* Je supporte pas Laure, comment la dégager du Loft ?

\* Kimy est-elle bisexuelle, si oui elle n'a aucune chance avec les filles du Loft ?

\* L'idylle entre Christophe et Julie symbolise t-elle le parfait amour ?

\* Pourquoi les filles ne trouvent-elles pas Fabrice beau ?

- Les questions hots :

\* Ils font quoi la nuit avec leur peluche dans loftstory ?

\* Y'a t-il un micro dans la cuvette des toilettes ?

\* Où trouver des photos de la piscine ou en infra rouge où on voit du sexe sur loftstory ?

- Enfin, les questions de société :

\* Régresse t-on à force de trop regarder Loftstory ?

\* Qu'est ce qui est le plus triste, love story ou loftstory ?

Eh oui, on vit une époque formidable...

Enfin, pour terminer cette étude-express, sachez que l'expression "Loft Story" apparaissait également, il y a encore peu de temps, dans le Top 30 du site... Lozère On line (<http://www.lozere-online.com/webmaster/top.php3>). Comme quoi, c'est vraiment un thème universel !

D'autre part, il nous a semblé intéressant de voir comment les grands annuaires francophones avaient traité le sujet au niveau de leur outil de recherche, comment ils avait "couvert" cette actualité.

Couverture du phénomène par les annuaires francophones majeurs

-----  
**Yahoo! France :**

-----  
\* 5 catégories créées :

- Actualités et médias > Télévision > Émissions et séries > Jeux télévisés > Loft Story
  - Actualités et médias > Télévision > Émissions et séries > Jeux télévisés > Loft Story > Anti-Loft Story
  - Actualités et médias > Télévision > Émissions et séries > Jeux télévisés > Loft Story > Humour
  - Actualités et médias > Télévision > Émissions et séries > Jeux télévisés > Loft Story > Participants
  - Divertissement > Sites à voir > L'intégrale des sélections de la semaine > Divertissement > Télévision > Loft Story
- \* Un dossier d'actualité ([http://fr.fc.yahoo.com/l/loft\\_story.html](http://fr.fc.yahoo.com/l/loft_story.html))  
\* 57 sites référencés

**Nomade.fr :**

-----  
\* 4 catégories créées :

- Actualité, presse > Radio, TV > Télévision > Films, émissions, séries > Emissions de télévision > Jeux télévisés > Loft Story
  - Actualité, presse > Radio, TV > Télévision > Films, émissions, séries > Emissions de télévision > Jeux télévisés > Loft Story > critiques, débats
  - Actualité, presse > Radio, TV > Télévision > Films, émissions, séries > Emissions de télévision > Jeux télévisés > Loft Story > Parodies
  - Actualité, presse > Radio, TV > Télévision > Films, émissions, séries > Emissions de télévision > Jeux télévisés > Loft Story > Soutien aux lofteurs
- \* Pas de zone d'actualité spécifique a priori.  
\* 27 sites référencés

**Guide de Voila :**

-----  
\* 1 seule catégorie créée :

- Actualité, média > Télévision > Emissions de télévision > Loft story
- \* Mais un site web quasiment entièrement créé pour la circonstance (Voila est l'un des sponsors de l'émission) :  
<http://loftstory.event.voila.fr/index.html>  
Celui ci a réalisé une très bonne audience avec 1,2 million de pages vues sur Voila depuis sa date de mise en ligne le 3 mai. Le pic d'audience ayant été atteint le 7 mai avec 720 00 pages dans la journée. Les chats avec les lofteurs ont également bien marché, comme par exemple le chat avec Aziz qui a eu 1 000 connections simultanées.  
\* 120 sites référencés !

**Looksmart France :**

-----  
\* 5 catégories créées :

- Actualités et médias > Télévision > Loft Story > Règlement / Actualité
  - Actualités et médias > Télévision > Loft Story > Débats sur Loft Story
  - Actualités et médias > Télévision > Loft Story > Candidats de Loft Story
  - Actualités et médias > Télévision > Loft Story > Parodies de Loft Story
  - Actualités et médias > Télévision > Loft Story > Pages de fans de Loft Story
- \* Pas de zones d'actualité (ce n'est pas réellement le métier de Looksmart France)  
\* 31 sites référencés

**Lycos France (annuaire) :**

-----  
\* 1 seule catégorie créée :

- Actualité, Médias > Télévision > Émissions > Jeux > Loft Story
- \* Pas de zone d'actualité spécifique a priori.  
\* 23 sites référencés

**Open Directory :**

-----  
\* 1 seule catégorie créé :

- Français: Actualité: Télévision: Emissions et séries: Loft story
- \* Pas de zone d'actualité spécifique a priori.  
\* 11 sites référencés

**MSN France :**

-----  
\* 1 catégorie créée :

- Infos - Météo > A la une > Loft Story
- \* Pas de zone d'actualité spécifique a priori.  
\* 21 sites référencés.

■ **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, Altavista France, Google, Voila, 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) : Personne.

Les moteurs à la peine ce mois-ci : Inktomi (donc Hotbot), Ecila, 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 ;-)

-> Altavista.com donnerait encore un meilleur classement aux pages anciennes (dont la date de dernière modification est plus ancienne qu'une autre). Ce critère de pertinence semblait avoir disparu, mais il semble qu'il soit revenu "à la mode" du côté de Palo Alto si l'on en croit cette url (<http://www.linux.com/article.php?sid=14>) et d'autres rumeurs concordantes.

-> Le site Canada.com utilisant maintenant Dogpile pour ses résultats à la place d'Inktomi, l'état se resserre autour de Hotbot et Hotbot France pour soumettre vos pages à Inktomi. Rappelons qu'Inktomi propose de nombreuses solutions payantes pour l'indexation de sites web. Il y a fort à parier - relation de cause à effet ? - que l'indexation gratuite par le biais du lien "Add URL" de Hotbot devienne de plus en plus anecdotique. Autre solution pour soumettre sur Inktomi : le site japonais Goo (c'est en japonais : premier champ = url du site, deuxième champ = e-mail) : <http://www.goo.ne.jp/help/info/url.html>

-> Le site Wordtracker (<http://www.wordtracker.com/>) comprendrait à l'heure actuelle 350 millions de mots clés et expressions.

-> Google recevrait quotidiennement 29 millions de requêtes. Ses critères de pertinences principaux seraient (en dehors de l'algorithme PageRank, basé sur l'indice de popularité à deux niveaux des pages indexées) : l'indice de densité des mots, la proximité des termes demandés (pour des requêtes de deux mots et plus), la mise en exergue (début ou fin de texte, mise en gras, taille de la police de caractère, texte d'un lien). L'option ALT des images est également prise en compte. Point important : le texte d'un lien pointant vers vos pages peut également avoir une importance. Pas bête, car dans le texte de ce lien, un résumé du contenu de la page peut effectivement être présent... L'outil ne prendrait pas en compte la taille du site web lui-même comme critère de pertinence. Toutes les pages sont traitées de la même façon, qu'elles soient issues d'un "petit" site comme d'un "gros". Les balises Meta ne sont pas prises en compte. Les pages dynamiques sont en train d'être prises en compte, petit à petit (on trouve, effectivement, certaines pages avec des urls "exotiques" dans l'index du moteur), mais avec prudence. Les index de Google seraient remis à jour tous les mois. Toutes les pages d'un site ne sont pas obligatoirement crawlées, mais il n'y a pas de limite officielle au nombre de documents indexés pour un site unique. Selon Craig Silverstein, l'un des responsables techniques de Google, les sociétés utilisant des technologies de cloaking sont exclues de l'index et mises sur liste noire. On peut difficilement être plus clair...

-> Le site Woonoz (<http://www.woonoz.com>), outil de recherche basé sur des experts humains, a déposé son bilan dernièrement et a été repris par equesto. La marque Woonoz est conservée (même si l'équipe est 100% différente).

-> Looksmart France s'apprête à lancer un nouveau produit de référencement. A priori, c'est une question de jours, voire d'heures...

#### ■ En Bref

-> Infoclic (voir ci-dessus) a signé un contrat avec QXL France (<http://www.qxl.fr>) pour mettre en place sur la rubrique "Aide" de leur site un outil d'aide automatisée. Il s'agit bien sûr de proposer à leurs visiteurs une assistance sur l'utilisation de leur site de vente aux enchères.

-> Quelques chiffres sur l'outil de recherche Voila :

- 3,8 millions visiteurs uniques/mois
- Taux de couverture de 43 %
- Entre 9 et 10 millions de pages vues en moyenne chaque jour.

-> Indexa, l'Annuaire Internet des Professionnels compte désormais plus de 30 000 sites référencés, dont plus de 2 000 sites de vente en ligne. L'inscription d'un site dans Indexa reste gratuite. Cependant, depuis plus d'un an, l'annuaire propose une inscription sous 48H pour 500 FHT. Par ailleurs, l'annuaire propose une méthode pour augmenter l'audience des sites référencés : La Fiche Flash.

- Pour le propriétaire du site, elle fonctionne comme un achat de mots clés lui garantissant une présence dans les résultats de recherche. - Pour l'utilisateur, elle apporte des informations supplémentaires sur l'entreprise et un accès direct aux sous-parties du site.

Quelques chiffres sur Indexa pour le mois de Mai 2001: 82 500 visites, 702 000 PAP, 29 900 sites référencés.

-> L'offre est passée plus ou moins inaperçue, elle est pourtant d'importance. Fast lance une offre couplée moteur de recherche interne - référencement. L'offre Fast PartnerSite propose un outil de recherche déporté pour site web (de type AtomZ.com - l'outil de recherche du site Abondance - ou Extense par exemple) avec une possibilité de voir son site inclus de façon garantie dans l'index de Fast plus une mise à jour quotidienne. L'offre sera lancée au troisième trimestre de cette année, mais on peut penser que c'est un peu ce qu'a fait Fast pour le site IBM.com ([http://www.fastsearch.com/index.php?d=press&c=display&pr\\_rel=60](http://www.fastsearch.com/index.php?d=press&c=display&pr_rel=60)). D'autre part, des rapports de positionnement seront également fournis par l'outil de recherche à leurs clients. L'offre devrait être payante, mais il n'est pas donné plus d'informations sur le site de Fast. Affaire à suivre...

Pour en savoir plus :

Fast ([http://www.fastsearch.com/index.php?d=press&c=display&pr\\_rel=59](http://www.fastsearch.com/index.php?d=press&c=display&pr_rel=59))

