

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

**Octobre 2000**

### ■ Au coeur du moteur de Voila

Nota : cet article a été écrit sur la base des informations fournies par l'équipe d'Echo qui gère le moteur de recherche Voila dans ses locaux de Sophia Antipolis. Toutes les informations indiquées ci-dessous ont été vérifiées, annotées, corrigées et validées par l'équipe d'Echo. Qu'ils soient ici remerciés de leur aide et du temps qu'ils ont passé à m'expliquer le fonctionnement de leur "bébé".

### Présentation

-----

La dernière version du moteur de recherche Voila (intitulée "KE" pour "Knowledge Engine") a été entièrement réécrite par rapport à la première version, lancée avec le moteur Voila en juillet 1998. Cette version "KE" a demandé deux années de travail à une cellule de quatre développeurs situés chez Echo, à Sophia Antipolis, où sont également hébergés les serveurs de Voila, entre autres.

L'index de Voila comporte aujourd'hui 12,9 millions de pages francophones et en contiendra prochainement beaucoup plus (chiffre encore indéfini, certainement supérieur à 20 millions de documents). Une "vague" de prise en charge de nombreux nouveaux sites est programmée pour bientôt (jusqu'à maintenant, le nombre d'url était limité dans l'index, mais cette limite va "sauter" très prochainement).

La nouvelle version du moteur est donc maintenant en ligne et pleinement opérationnelle sur le site de Voila (<http://www.voila.fr/>). Elle sera implantée sur le site de Wanadoo (<http://www.wanadoo.fr/bin/frame2.cgi?s=recherche>) dans les prochaines semaines (restons prudents... ;-)). Les réponses moteur seront alors exactement les mêmes sur les deux outils. Une stratégie de montée en charge progressive des accès au moteur est la cause de ce décalage progressif entre les deux portails.

### Facteurs bloquants à la soumission

-----

Lorsque vous soumettez un site web au moteur Voila (<http://www.voila.fr/Submit/>), un certain nombre de facteurs peuvent entrer en ligne de compte et favoriser - ou bloquer - le référencement de vos pages. En voici une description :

\* Pages dynamiques : les pages réalisés en Cold Fusion (suffixe .cfm) et en PHP (suffixes .php, .php3...) ne sont pas prises en compte du tout. En ce qui concerne les pages ASP, seule la page d'accueil du site (default.asp) est indexée. Les pages en .shtml sont prises en compte.

\* URLs exotiques : les pages ayant une url contenant des caractères dits "exotiques" comme, notamment, le point d'interrogation (?), ne sont pas prises en compte.

\* Pages graphiques : pour être reconnues par l'analyseur linguistique de Voila (qui n'accepte que les pages en langue française), le code HTML de la page web doit contenir assez de mots et de phrases en "bon français" pour être acceptée. Cela signifie que si une page est entièrement graphique à l'écran (composée uniquement d'images), la balise Meta Description devra être excellentement travaillée car elle sera le seul point de repère de l'analyseur linguistique pour y repérer des phrases en français. En effet, dans le cadre de cette reconnaissance, l'analyseur linguistique ne recherche pas uniquement des mots isolés issus d'un dictionnaire, mais bien des phrases construites selon une syntaxe significative de la langue de Molière. Cela signifie qu'une simple liste de mots clés, comme celle de la balise Meta Keywords, ne suffira pas à accepter un document HTML, puisqu'elle ne contient pas de véritable phrase de type "sujet - verbe - complément". Il est donc essentiel que la balise Meta Description d'une page très graphique contienne au minimum 150, et idéalement 250 caractères et surtout des phrases écrites en français "intelligible".

\* Frames : toutes les pages, "mère" et "filles" sont prises en compte. Cependant, le lien de parenté entre "pages mères" et "pages filles" n'est pas gardé. Il est donc nécessaire d'insérer des codes javascript de reconstitution de ce lien dans les pages filles. Les balises <Noframes> sont lues et les liens éventuels à l'intérieur de celles-ci sont suivis. En suivant les indications d'Abondance à ce sujet, votre site devrait être bien référencé même s'il propose des frames. Voir : <http://www.abondance.com/docs/frames.html>

\* Flash : les fichiers HTML sont pris en compte, pas le contenu des animations Flash. Le

contenu de la balise <Noembed> est lu.

\* Imagemaps : pas de problème, les "images cliquables" sont prises en compte et les liens qu'elles contiennent sont suivis par le robot.

\* Redirection : si une page de redirection contient une balise <Meta Refresh>, cette page est indexée, mais pas la page de destination (le robot ne suit pas le lien intégré dans la balise Meta en question). Si votre site s'ouvre sur une page proposant ce type de redirection, seul ce document contenant la balise <Meta Refresh> sera indexé et pas les autres pages du site.

\* Liens javascripts : les liens insérés dans des routines Javascripts ne sont pas suivis par le robot de Voila. Attention donc à la façon dont sont proposés vos liens, notamment sur la page d'accueil. Essayez toujours de proposer une version textuelle "simple" des liens proposés afin d'éviter tout problème, surtout si vous utilisez des "roll-overs" dans vos pages.

\* URL : les pages web de votre site ne seront prises en compte que si leur URL a une taille inférieure à 255 caractères.

\* CGI : les sites web "s'ouvrant" sur un CGI (détection d'un plug-in, de la plate-forme ou du navigateur de l'internaute, etc.) ne posent a priori pas de problème d'indexation.

### **Crawl des pages Web**

-----

Un site web est indexé par Voila avec une limite actuelle du point de vue du nombre de pages par nom de serveur (par exemple, www.site.com et sport.site.com sont deux noms de serveur différents pour Voila). Si un site web dépasse cette limite (cas des hébergeurs gratuits comme Multimania, Chez.com ou Geocities), le traitement est ensuite réalisé à la main par les ingénieurs de Voila, ultérieurement et au cas par cas. Ceci dit, pour les sites "petits" et "moyens", qui proposent quelques dizaines, centaines, voire milliers de pages, cela ne devrait pas poser de problèmes. Pour les très gros sites, leur cas est traité par l'équipe technique de Voila et ne pose pas de réel problème.

Il n'existe pas de limite de profondeur (nombre de sous-répertoires) pour crawler les pages d'un site. La tendance pour les semaines à venir est de prendre en compte le plus de pages possible pour un site web donné. Le fichier robots.txt est pris en compte pour interdire au robot d'indexer certaines zones du site (le système est en train d'être totalement finalisé en ce moment).

La balise <Meta Robots> est lue mais pas encore prise en compte, ce qui serait cependant une question de semaines. Pour interdire à Voila l'accès à certaines zones de votre site, la seule solution à l'heure actuelle reste donc le fichier robots.txt.

Voila effectue un dédoublonnage de ses bases. Donc, un site qui serait présent à la fois sur, par exemple ;-), www.abondance.com et www.abondance.net dans l'index de Voila par le jeu des liens trouvés sur le Web et pointant vers ces deux URLs, sera présent uniquement sur une adresse après dédoublonnage.

Rappelons également que tous les sites du Guide de Voila (la partie annuelle) sont automatiquement et régulièrement visités par les robots de Voila.

Enfin, l'objectif de Voila est de faire du J+1 entre la soumission et l'indexation de l'adresse du site soumis. Ce délai extrêmement court devait être en place et tourner sans problèmes d'ici la fin de l'année, peut-être plus tôt (pas de date précise encore).

### **Champs pris en compte**

-----

Lors de l'indexation et du classement d'une page web, les champs suivants sont pris en compte :

- Balise <Title> - Balises <Meta description> et <Meta Keywords> - Options ALT des balises IMG - Texte visible (balise <BODY>). Un texte mis en exergue (notamment au moyen de l'option "SIZE" des balises "FONT") a un poids plus important, ainsi qu'un texte à l'intérieur d'un lien (exemple : <A href="adresse">texte à l'intérieur d'un lien</a>).

Seules les trois premières occurrences d'un mot sont prises en compte dans le code HTML de la page. Ainsi, si un même mot se trouve dans les balises <Title>, <Meta Description> et <Meta Keywords>, son occurrence dans le texte visible ne sera pas prise en compte. Cela peut poser problème pour le bon classement d'une page, du fait que les balises <Meta> ont, chez Voila, un poids assez faible dans les algorithmes de pertinence, contrairement au texte visible. Le problème est identique mais moins critique que pour Altavista qui ne prend en compte, lui, que les deux premières occurrences d'un mot dans le code HTML d'une page.

Les commentaires ne sont pas indexés.

L'indice de popularité est également utilisé (voir plus loin la description des algorithmes de classement). Une base de données des liens le plus souvent cliqués pour tel ou tel mot clé est en train d'être constituée (système d'indice de clic de type DirectHit), mais n'est pas utilisée pour l'instant dans les algorithmes de pertinence.

Lors d'une interrogation sur le moteur de recherche, la troncature n'est pas appliquée (la saisie de "abon" cherche uniquement "abon" et pas "abondance"); A terme, une option de stemming (la demande d'un mot au singulier trouve le pluriel et un pluriel trouve son singulier) pourrait être mise en oeuvre par la suite.

Sur ce sujet, un prochain article de la lettre "Recherche et Référencement" vous donnera quelques "tuyaux" pour mieux effectuer vos recherches sur le moteur Voila.

## **Algorithmes de classement**

---

Le moteur Voila utilise plusieurs algorithmes de classement différents pour présenter ses résultats. Ces algorithmes sont utilisés dans certains cas, différents, selon ce que saisit l'internaute ou les fonctionnalités qu'il choisit (notamment la prise en compte de l'indice de popularité); Ils peuvent être très rapidement et simplement "basculés" de l'un vers l'autre par l'équipe technique de Voila, selon les circonstances, ce qui explique que, parfois, les résultats peuvent changer du tout au tout sur ce moteur. On peut en fait classer ces algorithmes sous trois familles principales :

Famille 1 : algorithmes faisant jouer la typographie et / ou la densité  
Famille 2 : la famille 1 à laquelle sont rajoutées des notions de popularité  
Famille 3 : la famille 1 à laquelle sont rajoutées des notions de proximité

Ces trois familles se basent sur 6 algorithmes de pertinence

- Algorithme 1 : Typographie uniquement : la présence du mot demandé est détectée dans différentes zones de la page (par exemple : balise <Title>, texte visible, etc.). Le nombre d'occurrences du mot demandé n'a que peu d'importance. Une pondération est mise en place selon les zones dans lesquelles se trouve le mot. Zones les plus importantes : balise <Title>, texte visible. Zones secondaires : balises <Meta>, options ALT.

- Algorithme 2 : Typographie + densité : l'algorithme 1 est repris et on lui ajoute une notion de densité (nombre d'occurrences du mot demandé par rapport au nombre total de mots dans la page). En clair, le nombre d'occurrences du mot est ici important, en fonction de la taille de la page. Plus le document est long, plus le mot en question pourra être répété. La densité consiste en un chiffre codé sur 8 bits et fournit un résultat oscillant donc entre 0 et 255.

- Algorithme 3 : Correspond à l'algorithme 1 ou 2 avec prise en compte de l'indice de popularité "nouvelle formule" chaque semaine, un "index de popularité" est créé, contenant, pour toutes les pages, la liste des pages qui pointent vers elle. Cet index sert alors à calculer par la suite l'indice de popularité d'une page. Lors de calcul, l'indice de la page en question est pris en compte, mais le fait qu'une page contenant le dit lien se trouve dans le lot de pages contenant le mot est important. Hum, hum, reprenons... ;-) Un internaute tape le mot clé "abondance" sur Voila. Le moteur trouve, par exemple, 5 000 pages contenant ce mot. Pour la page P1, le calcul d'indice de popularité indique que 200 pages de l'index complet (les 12,9 millions de pages) pointent sur P1. Mais, parmi ces 200 pages, 50 d'entre elles font partie des 5 000 pages contenant le mot clé "Abondance". Ce point sera alors pris en compte dans le calcul de pertinence, la page P1 ayant alors un poids plus fort que si, parmi les 200 pages contenant un lien, aucune ne faisait partie des 5 000 pages trouvées. Une aspirine ?

- Algorithme 4 : Notoriété : l'indice de popularité "simple" est ici calculé, sans pondération telle que prise en compte dans l'algorithme 3.

- Algorithme 5 : Correspond à l'algorithme 1 ou 2 avec prise en compte de la proximité. Si un seul mot est demandé, cet algorithme n'est pas pris en compte. Si plusieurs mots sont demandés, la proximité entre ceux-ci à l'intérieur d'une page est importante. Une page contenant les deux mots l'un à côté de l'autre sera mieux classée qu'un document contenant les deux mots, mais l'un au début et l'autre à la fin du code HTML. Plus les termes demandés sont "proches" les uns des autres dans le document et meilleur est son classement.

- Algorithme 6 : "mix" entre les algorithmes 5 et 3. Le 5 est d'abord appliqué (proximité), puis le 3 (popularité).

A l'heure actuelle, lorsqu'un seul mot est demandé par l'internaute sur Voila, c'est

l'algorithme 3 qui est prise en compte.

Si plusieurs mots sont demandés, c'est l'algorithme 6 qui est pris en compte.

Pour tous les algorithmes décrits ci-dessus, la rareté du terme demandé est prise en compte. Par exemple, si un internaute recherche "Michel Duchaussoy", l'expression comprend le terme "Michel", très souvent présent sur les pages du Web francophone et "Duchaussoy", beaucoup plus rare. Imaginons que l'on trouve 1000 pages contenant "Michel" et 10 contenant "Duchaussoy". Cette dernière aura un poids, pour le critère "rareté", 100 fois supérieur à celle contenant le mot clé "Michel".

### **Lutte contre le spam**

-----

Voilà lutte et va lutter de façon encore plus importante contre le spam dans les mois qui viennent. Le test de la langue, en entrée, constitue déjà un filtre important, tout comme la prise en compte de l'indice de popularité.. La répétition abusive de smots clés est pris en compte par l'algorithme 2. Les plus gros problèmes sont posés à l'heure actuelle par des sites "roses" qui peuvent parfois soumettre près de 40 000 pages en une seule journée, alors que le nombre moyen de soumission reçues par Voila dans une journée par Voila oscille entre 10 et 15 000 urls :-(. Ceci dit, Dans ce chiffre, il y a déjà beaucoup de spam quotidien...

#### **■ Nouvelle interface pour Nomade.fr**

Nomade.fr a mis en ligne le mois dernier une nouvelle interface pour son site-portal. Une bonne occasion pour faire le point sur la recherche d'information sur cet outil.

La recherche simple est bien entendu accessible à l'adresse :

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

La page d'accueil a maintenant "viré" très "portail" (mais de façon moins "fouilli" que sur l'ancienne version), de nombreuses informations de tous domaines étant disponibles, autour de quatre grandes préoccupations de l'internaute : "Acheter", "Communiquer", "Trouver de l'information et des services", "Trouver du contenu". Seule cette dernière option, la recherche d'information, nous intéressera ici.

La page d'accueil propose un formulaire de recherche avec les seuls choix "Francophone" (coché par défaut, recherche sur Nomade) ou "Mondial" (recherche sur Inktomi).

Sur l'annuaire, la syntaxe d'interrogation est toujours la même que précédemment.

Rappelons les spécificités de Nomade :

Minuscules/majuscules : Nomade ne tient pas compte de la casse des lettres.

Exemples :

ibm trouve ibm, IBM ou Ibm.

Ibm trouve Ibm, ibm ou IBM.

IBM trouve IBM, ibm ou Ibm.

Lettres accentuées : Nomade ne tient pas compte de l'accentuation des lettres.

Requête saisie sans lettres accentuées : toutes les occurrences sont recherchées. Exemple : electricite trouve electricite, électricité ou electricité.

Requête saisie avec des lettres accentuées : toutes les occurrences sont recherchées.

Exemple : électricité trouve électricité, electricite ou electricité.

Ordre des mots : dakar paris donne un résultat identique à paris dakar

OU : "OU"

Exemple :

paris OU dakar

trouve les fiches descriptives qui contiennent soit paris, soit dakar, soit les deux.

ET : opérateur "ET" ou signe + (Opérateur par défaut)

Exemple :

paris ET dakar

ou

paris dakar

ou

+paris +dakar

recherchent les fiches descriptives qui contiennent à la fois paris ET dakar.

SAUF : opérateur "SANS" ou signe -

Exemple :

dakar SANS paris

ou

dakar -paris  
recherchent les fiches descriptives qui contiennent le mot dakar mais qui ne contiennent pas le mot paris.

Expressions : Guillemets.

Exemple :

"paris dakar"

recherche cette expression dans les fiches descriptives, les deux mots l'un à côté de l'autre et dans cet ordre.

Troncature : par défaut

Exemple :

cherch

recherchera :

cherche, chercheur, chercheurs, chercheuse, etc.

Pour retrouver la syntaxe de Nomade sur le site Abondance :

[http://www.abondance.com/outils/no\\_syntaxe.html](http://www.abondance.com/outils/no_syntaxe.html)

et l'officielle :

<http://www.nomade.fr/aide/recherche/rechercher.asp>

La recherche avancée (disponible en cliquant sur le lien "Plus d'options" sur la page d'accueil, ou directement au travers de la page de résultat sur une recherche "simple") propose plusieurs possibilités, sous la forme d'onglets spécifiques :

- Recherche sur le Web francophone (<http://rechercher.nomade.fr/recherche.asp?avc=o>)  
Permet de faire une recherche au sein de l'annuaire de recherche "classique" de Nomade.fr.

Pour chaque site, Nomade.fr affiche le titre du site, son URL (adresse Internet), le nom de son éditeur, sa nature (personnel, entreprise, association, éducation, public) ainsi que sa ville et son pays. Un descriptif du contenu du site en quelques lignes est proposé, le public auquel il s'adresse (Tous publics, ados/adultes, adultes, professionnels, enfants) et les catégories de Nomade.fr dans lesquelles le site est répertorié. Le nombre entre parenthèses correspond au nombre de sites référencés dans la catégorie.

Pour effectuer une recherche, saisissez un ou plusieurs mots dans le champs de saisie proposé. Elle s'effectue alors sur les descriptions des sites : le titre, le résumé, les informations concernant l'éditeur, ainsi que sur quelques mots clés complémentaires. Malheureusement, Nomade n'effectue toujours pas de recherche par défaut dans l'URL des sites, contrairement à tous les annuaires majeurs actuels. Un bug à corriger au plus vite, à notre avis. Cette recherche est cependant possible dans les options complémentaires (voir ci-dessous).

La recherche par options (+ d'options) vous permet de faire une recherche sur chacun de ces critères d'indexation spécifique à Nomade :

- l'adresse du site ou URL ex : www.nomade.fr
- l'éditeur du site, ex : france télécom
- ville de l'éditeur, ex : Paris
- nature de l'éditeur (entreprise, association, public, personnel, éducation)
- public visé : professionnel, enfants, ados-adultes, adultes, tous publics

Se reporter à la lettre "Recherche et Référencement" de septembre dernier pour bien comprendre la signification de ces champs.

L'interface de recherche propose plusieurs onglets différents. Le fait de cliquer sur un de ces onglets induit une recherche sur le mot saisi pour la recherche précédente, dans la zone d'investigation choisie :

- Recherche sur le Web mondial (<http://ink.nomade.fr/ink.asp?src=&opt=1&MT=&GL=INTL>)

Une recherche sur le web mondial vous permettra de faire une recherche sur les pages web indexées par le moteur de recherche Inktomi.

- Recherche sur le Web régional (<http://rechercher.nomade.fr/recherche.asp?src=&opt=2&s=>)

Pratique pour faire une recherche locale, elle vous permet de faire une recherche par ville ou région sur l'origine géographique de l'éditeur dans Nomade.fr. Très intéressante, cette fonctionnalité est l'une des solutions de recherche les plus pointues proposées par Nomade et l'une de celles qui fait sa spécificité. A utiliser sans modération pour des requêtes régionales.

- Recherche sur l'actualité (<http://rechercher.nomade.fr/recherche.asp?src=&opt=3&s=>)

Pour retrouver de l'information sur l'ensemble des dépêches d'actualité publiées sur Nomade.fr ou sur des thèmes spécifiques : politique, sport, multimédia, etc.

- Recherche sur les sélections (<http://rechercher.nomade.fr/recherche.asp?src=&opt=4&s=>)

Une recherche sur les Sélections vous renvoie sur des sites sélectionnés par Nomade.fr tant pour la qualité de leur contenu, leur originalité que pour leur qualité graphique. Les sites sélectionnés apparaissent avec une "palme cliquable". Pfff, Abondance n'en a même pas une, nul Nomade... ;-)))))) (je plaisante, Nomade est abonné à cette lettre, je les aime donc beaucoup ;-)).

- Recherche sur les forums (<http://rechercher.nomade.fr/recherche.asp?src=&opt=5&s=>)

Recherche sur les forums Usenet français (fr.\*). Fonctionnalité nouvelle et très intéressante. Chaque message est daté et le forum dont il est issu est indiqué.

Prochains projets de Nomade : la personnalisation du portail, personnalisation active selon les choix de l'internaute. Ce chantier est aujourd'hui prioritaire sur le portail. Une chaîne "Femmes" et de nombreuses fonctionnalités de shopping en ligne sont également au programme pour très bientôt.

### ■ 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) : toujours Altavista et All The Web donc Lycos US et bientôt Lycos FR.

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) : Infoseek, Google, Ecila, Northern Light, Excite US.

Les moteurs à la peine ce mois-ci : Altavista France, qui semble "en travaux", Voila, Excite (pour les sites en français), Hotbot (donc Inktomi), Lycos France, Lokace.

### ■ Bruits et chuchotements

Une nouvelle 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 France travaillerait actuellement à une toute nouvelle version du moteur prévue pour début 2001. Ces modifications feraient partie d'une stratégie mondiale de tous les sites d'Altavista en dehors des Etats-Unis. Vous en saurez plus très bientôt sur Abondance, mais le projet risque de modifier la façon dont nous utilisons le moteur à l'heure actuelle. A très bientôt, promis...

-> Courant octobre, sortie de la nouvelle page d'accueil du portail Voila et accès des la page d'accueil aux catégories du Guide Voila :

1/ Nouvelle page d'accueil avec 4 objectifs :

- simplifier et clarifier l'accès à tous les services en les classant par grands ensembles thématiques : Acheter, Communiquer, etc.
- donner un accès direct au Guide du Web en le séparant des chaînes
- renforcer la visibilité des contenus éditoriaux et d'actualité (AFP, Sports, dossiers événementiels, etc.)
- rendre l'interface plus vivante, plus conviviale.

2/ Le Guide du Web

- accès direct dès la page d'accueil de Voila à une page d'accueil du Guide
- accès direct dès la page d'accueil de Voila aux catégories principales du Guide
- sélection de sites hebdo mise en avant par l'équipe de documentalistes
- classement des sources d'infos (services Voila, actu...) dans les pages des catégories du Guide.

Si vous êtes gentils, vous aurez droit en avant-première d'ici quelques jours sur Abondance au nouveau look du portail...

-> Un bruit a couru sur l'Internet francophone, comme quoi Lycos Europe choisirait finalement Inktomi à la place de Fast pour ses index européens. Il n'en est rien, c'est bien Fast qui fournira ces index, et notamment celui de Lycos France. Cependant, Lycos Europe travaille bien avec Inktomi sur d'autres projets encore confidentiels pour l'instant. Des annonces sur ce sujet sont prévues pour début novembre.

-> Hotbot serait en train de lancer une filiale française. Pas plus d'infos pour l'instant. On creuse...

-> L'index Fast devrait être disponible sur lycos Europe et intégré au catalogue Lycos ainsi qu'à la technologie Direct Hit courant Novembre.

### ■ En Bref

-> Attention : Lorsque vous utilisez l'interface de vérification de la présence d'un site sur le

moteur de Voila

([http://www.voila.fr/Informations/Moteur\\_guide/recherche\\_inscription.html](http://www.voila.fr/Informations/Moteur_guide/recherche_inscription.html)), le système ne marche pas si votre nom de domaine comporte un tiret (immobilier-paris.com par exemple). Le bug vient du fait d'une confusion dans les interprétations du tiret et du signe "moins". Bug signalé et en cours de réparation chez Voila.

-> Altavista, sur sa page de soumission d'URL, propose depuis quelques semaines un lien pour soumettre le site à Looksmart, exactement comme sur Excite.

-> Excite : son algorithme prendrait en compte de façon assez forte, en ce moment, l'indice de popularité. Autre point très important : la présence du site sur L'open Directory, Yahoo! ou Looksmart - surtout - qui dope les pages du site dans le classement du moteur. Enfin, la présence de smots clés à l'intérieur des textes des liens (<a href="adresse">Texte du lien</a>).

-> Google favoriserait, dans ses classements, les pages faisant moins de 10 Ko (code HTML uniquement, ce chiffre ne tient pas compte du poids des images).

-> Altavista France propose en ligne un Guide de la recherche très pointu (<http://fr.altavista.com/doc/guide/>). Ce guide sera bientôt disponible sous d'autres formes (informations encore confidentielles, patientez quelques semaines, mais le résultat risque d'être assez sympa... ;-)).

-> L'outil de recherche de MSN a complètement changé son "look" (<http://search2.msn.fr/>). Vous en saurez plus dans le prochain numéro de la lettre "Recherche et référencement".