



# Internet

## La trousse à outils de l'internaute errant

### Moteurs de recherche, troisième partie

Après avoir abordé par deux fois et de manière quelque peu générale le sujet de combien important des assistants au service de la recherche de l'internaute errant en quête d'informations, fi des introductions qui risquent de lasser les lecteurs de cette rubrique. Entrons donc dans le vif du sujet par la présentation des techniques plus tangibles qui permettent de structurer et de cibler la fouille dans la jungle du Web.

Qu'il me soit permis de rappeler que l'adresse e-mail ci-dessous est également au service du courrier des lecteurs. Si vous avez des questions, des remarques ou des critiques à formuler, n'hésitez pas. Un média interactif a besoin d'échanges interactifs!

Thomas Vauthier  
th.vauthier@bluewin.ch

### Recherche par têtes de chapitre

La recherche par rubriques, telles qu'elles figurent sur les pages d'accueil des grands moteurs de recherche, est la méthode la plus simple et la plus primitive. L'internaute se contente de suivre la liste des sélections proposées (p.ex. *Sport, Santé, Culture*, etc.), derrière lesquelles se cache un certain nombre d'adresses directement accessibles par un simple clic de souris. Il s'agit plutôt d'une ballade guidée, mais cette manière de



procéder peut s'avérer fort utile aux débuts de l'exploration d'Internet, car elle permet de survoler aisément l'offre présélectionnée d'un centre d'intérêt donné.

Les grandes machines comme *Yahoo!* ou *Excite* ont affiné les possibilités, en laissant d'abord entrer l'utilisateur dans un domaine défini en suivant des liens vers des sous-chapitres, puis de définir par mot-clé la recherche ciblée dans la banque de données y relative. Une telle recherche à un niveau plus précis ne permet pas seulement d'économiser passablement de temps; les taux de succès, donc de retrouver une adresse intéressante, sont également bien plus élevés.

### Recherche par mot(s)-clé(s)

La recherche par des termes plus précis permet d'explorer de manière élégante l'offre gigantesque et d'aboutir directement aux renseignements souhaités. La technique est simple: il suffit d'entrer le mot-clé dans le champ prévu à cet effet – et le robot fera le reste. Le bouton miracle situé à côté de la fenêtre de saisie s'appelle dans la plupart des cas *go*, *search* ou *submit* query, alors que les moteurs francophones utilisent en général les commandes *Recherche(r)* ou *Soumettre*. Méthode plus sophistiquée, le recours aux opérateurs booléens (ci-après) permet d'enrayer très nettement le domaine que le robot va fouiller, puisque la commande associe alors deux ou plusieurs mots-clés, ce qui va d'inclure ou exclure d'emblée un certain nombre d'adresses ou de documents.

### Opérateurs booléens

Tous les grands robots de recherche offrent la possibilité d'utiliser les commandes AND, OR, NOT et NEAR. Vu la disparité actuelle en la ma-

tière, il est toutefois préférable de se renseigner sur la syntaxe de recherche précise nécessaire pour chaque robot individuel. En général, les informations y relatives se cachent sur une page intitulée «recherche avancée» ou similaire. En effet, alors que pour les uns, il convient de respecter un espace entre les mots-clés et les commandes, pour d'autres, il n'en faut surtout pas; d'autres encore simplifient en admettant des symboles comme + ou -. AND ou + est certainement la commande booléenne la plus utilisée, parce que l'association directe de deux ou plus de termes pour restreindre le champ de recherche est la plus usuelle et la plus logique. Une recherche par le seul terme «hôpital» est bien trop vague et nous livrera des milliers d'adresses, de quoi se perdre; en y associant «+universitaire» et mieux encore «+Genève», nous allons tomber assez près du but. En fait, il s'agit d'appliquer de façon condensée la technique que nous utilisons en parcourant de notre index un listing dans un annuaire téléphonique.

### Metawords

Plus pointue encore, la recherche par des *metawords* ou opérateurs de syntaxe avancée permet de limiter la sélection désirée à des titres de documents, des adresses URL, etc. Il est aussi possible d'exclure toutes les informations sauf celles sous forme de vidéo/audio ou graphique. En entrant *t:spaghetti*, la machine va afficher seulement les documents contenant le terme «spaghetti» dans leur titre. En mettant plusieurs mots entre des guillemets, on peut rechercher par des phrases entières.



### Langue

Internet est le monde type de l'hégémonie anglophone quasi absolue. Peu étonnant donc que les moteurs de recherche d'emblée francophones sont rares, tout comme ceux proposant un filtre par langue. *AltaVista* et *HotBot*, pourtant de source américaine, permettent de présélectionner la langue de recherche (elle est définie dans une fenêtre à part). C'est un réglage qui peut s'avérer fort utile et qui permet d'être plus efficace et plus rapide.

### Majuscules

Jadis (vous souvenez-vous de ces imprimantes paléontologiques?) pour ainsi dire obligatoire, la saisie des mots-clés par des lettres majuscules a heureusement cédé à un libéralisme nouveau. Une seule précaution demeure: les fautes de frappe et autres erreurs de saisie demeurent fatales, car la machine ne pense pas!

### Opérateurs de langue naturelle

L'un des objectifs essentiels du développement actuel de logiciels pour les applications Internet est de créer la possibilité de la saisie des données dans une forme aussi proche de la langue naturelle que possible. En s'adressant à un robot «intelligent», l'internaute du futur pourra s'adresser à un interlocuteur virtuel. Bien qu'étant un premier pas dans cette direction, les syntaxes de recherche avancée ou des systèmes de reconnaissance vocale sont hélas encore loin de cette vision d'avenir. Toutefois, l'exploration des aides à la recherche avancée, dont disposent pratiquement tous les robots dans une rubrique ad hoc, valent toujours la chandelle. Remarque: à l'instar de la plupart des applications en informatique, les configurations standard sont nettement insuffisantes!



A suivre ...