

## Les outils de recherche

Le Web vient de dépasser le milliard de pages. Sauf à connaître l'URL, il est impossible de trouver l'information que l'on cherche sans les outils appropriés.

Ces outils sont indexés sur le portail *d'Anasys*, à la rubrique : *outils de recherche*.

### Généralités sur les outils de recherche

Par abus de langage on appelle ces outils des « moteurs de recherche ». En réalité il y a quatre types d'outils de recherche : les (vrais) moteurs de recherche, les annuaires, les méta-moteurs et les agents de recherche.

### Les (vrais) moteurs de recherche

Le meilleur exemple est Altavista<sup>®</sup>. Ce site dispose de robots qui sondent le texte intégral de la totalité de l'Internet et classent ensuite les pages selon un index de mots-clés.

### Les annuaires

Yahoo<sup>®</sup> en fait partie. Ce sont des répertoires dont les entrées ont été réparties par une main humaine, et non par un ordinateur, en fonction d'une classification logique (par thème). La recherche se fait par mots-clés ou bien en remontant l'arborescence correspondant à son thème de recherche.

On trouve dix fois plus de pages sur Altavista<sup>®</sup> que sur Yahoo<sup>®</sup>. Certains prétendent que Altavista<sup>®</sup> est la mémoire de l'humanité ! La richesse de son contenu est réellement surprenante.

Pour une bonne recherche, il est utile d'essayer les deux types d'outils.

On notera d'ailleurs que Yahoo<sup>®</sup> propose, en bas de page, de transférer votre requête à ses concurrents (il vous en fournit la liste). Il suffit de cliquer sur l'un d'eux pour que le transfert se réalise. Ce service mérite d'être souligné !

### Les méta-moteurs (ou méta-chercheurs)

Ils permettent de mener la même requête sur plusieurs moteurs à la fois. Ceci procure un gain de temps considérable.

Le plus connu est Metacrawler<sup>®</sup> (<http://www.metacrawler.com>).

Un méta-moteur spécifiquement dédié aux moteurs de recherche francophones peut être consulté à : <http://www.ariane6.com>.

### Les agents de recherche

Copernic<sup>®</sup> est le plus utilisé. A la différence des méta-moteurs de recherche, il nécessite le téléchargement d'un petit logiciel. Ce logiciel est gratuit dans sa version de base. Il a l'avantage de posséder une version française et d'être très efficace pour une première approche d'un thème.

L'adresse de téléchargement est <http://www.copernic.com/fr/>

Copernic<sup>®</sup> permet de faire sa requête sur les moteurs de recherche du Web francophone ou sur ceux du monde entier.

Il est efficace pour la plupart des recherches. Son utilisation nous semble très performante dans les recherches de documents rares.

Voici un exemple de recherche avec Copernic<sup>®</sup> sur le Web francophone (consensus ou référence sur le diabète). Nous visualisons la fenêtre de requête. Copernic<sup>®</sup> lance douze moteurs. Les résultats s'affichent avec un score de pertinence, sur le navigateur. Il ne reste qu'à aller les visualiser en cliquant sur leur adresse (voir figures 1, 2, 3, 4).

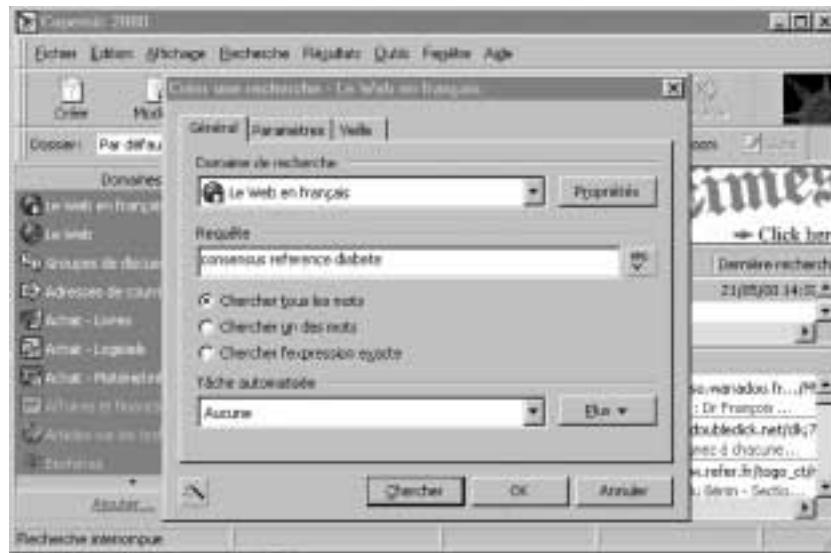


Figure 1 : Fenêtre de recherche de Copernic®.

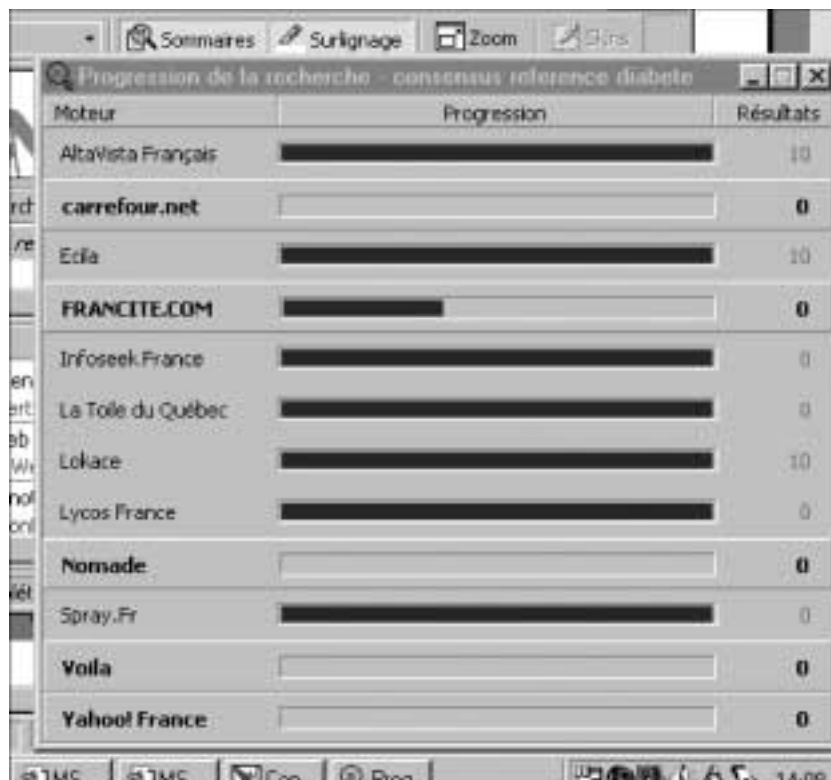


Figure 2 : Suivi de la recherche de Copernic® sur différents moteurs du Web francophone.

### Comment réaliser une requête sur Yahoo® ou Altavista® ?

La recherche par arborescence ne requiert qu'une démarche logique. Nous ne ferons que la citer.

Nous étudierons plutôt la recherche par mots-clés que l'on inscrira sur la fenêtre du moteur de recherche.



Figure 3 : Résultats de la recherche de Copernic® sur consensus, references, diabete.



Figure 4 : Résultats de la recherche de Copernic® sur consensus, diabete.

On commencera par la version française de Yahoo® : <http://www.yahoo.fr> (version américaine en .com au lieu de .fr).

Altavista® existe, depuis peu de temps, en version anglaise (<http://www.altavista.com>) et en version française (<http://www.altavista.fr>).

En général, **le choix du mot-clé** le plus signifiant est aisé. Il faut ensuite éviter quelques embûches en relation avec les spécificités linguistiques (pas d'accent en anglais) ou informatiques (les serveurs Unix® distinguent la casse des lettres).

Voici quelques règles à respecter :

- écrivez dans la langue des documents que vous recherchez.

Vous pouvez aussi procéder à une recherche en français suivie d'une autre en utilisant la traduction anglaise de votre mot-clé ;

- écrivez sans faute d'orthographe.

Si vous n'êtes pas sûrs d'un mot, beaucoup de moteurs accepteront que vous remplaciez un groupe de lettres par une \* (attention : je dis un groupe de lettres et non une lettre). Pour remplacer une lettre on utilise, en général, un point d'interrogation : « ? ». On a aussi intérêt à utiliser une troncature à droite avec \* pour récupérer tous les termes ayant la même racine. Par exemple evalua\* récupérera : evaluation, évaluable, évaluée ;

- écrivez sans accent.

Souvent, les développeurs ont volontairement supprimé les accents, même dans du texte français (pour mieux passer les tris des moteurs anglophones) ;

- écrivez au singulier.

Cela vous permettra de récupérer aussi les pluriels (sauf si vous avez mis le mot entre guillemets). Avec les pluriels, votre recherche sera limitée aux pluriels ;

– écrivez en minuscule.

Ainsi vous récupérerez les pages contenant le mot en minuscule et en majuscule ;

– écrivez sans faire de phrases.

Évitez les déterminants (les articles). Cependant, si vous cherchez une expression particulière, mettez le groupe de mots entre guillemets.

## Les opérateurs booléens

Vous allez vite vous rendre compte que le moindre mot-clé rapporte des milliers de pages et parfois des millions. Il est donc impossible de trouver l'information la plus pertinente car elle est noyée dans un océan d'informations inutiles (on appelle ça « le bruit » par opposition au « silence » qui est une requête qui ne rapporte rien quand le mot-clé n'est pas adéquat).

Nous devons donc préciser notre requête.

Si nous inscrivons plusieurs mots-clés séparés par un espace, le moteur rapportera, le plus souvent, les pages contenant chacun des mots-clés.

Par exemple, pour *maladie iatrogène*, le moteur va rapporter toutes les pages contenant le mot « maladie » (il y en a beaucoup !) et toutes celles contenant le mot « iatrogène » !

Nous verrons plus loin que chaque espace entre les mots est interprété comme un opérateur « OR ».

Les moteurs de recherche ne comprennent pas la syntaxe des phrases que nous utilisons. Quand nous devons procéder à une requête ciblée, il est nécessaire d'employer une syntaxe accessible au matériel informatique : **les équations booléennes**.

Il s'agit d'un langage rudimentaire associant nos mots-clés et des opérateurs logiques en anglais (sur le site *Adminet* et sur d'autres moteurs francophones, les opérateurs sont en français).

Les plus employés de ces opérateurs sont : **AND, OR, NOT, NEAR**.

Il est préférable de se positionner sur les fenêtres de mode de requêtes avancées pour utiliser les opérateurs booléens.

Certains moteurs n'acceptent qu'un nombre réduit de types d'opérateurs. Vous le constaterez dans l'expression qu'ils donnent de votre requête.

Bien que la casse des opérateurs, préconisée par les moteurs de recherche, ne soit pas identique, nous ferons le choix d'essayer d'abord de mettre les opérateurs en majuscule (et les mots-clés en minuscule).

La plupart des moteurs de recherche acceptent aussi l'utilisation des signes + (à la place de AND) et – (à la place de NOT) entre les mots-clés.

Un + devant un mot, en début d'expression de plusieurs mots, lui donne une priorité dans la recherche.

Il n'y a pas d'espace entre les mots et les signes mais il y a un espace entre les mots et les opérateurs.

Voici la signification générale de différents opérateurs (il faut savoir que certains moteurs ne disposent pas de toutes les fonctions booléennes décrites) :

– **AND** : les mots-clés unis par AND doivent obligatoirement, tous les deux, figurer dans la page recherchée ;

– **OR** : les pages doivent contenir l'un ou l'autre des mots-clés (ou bien les deux) ;

– **NOT** ou de préférence **AND NOT** : le mot qui suit est obligatoirement absent des pages recherchées ;

– **NEAR** : les deux mots-clés sont proches de dix mots ou moins (c'est une garantie de relation entre eux). Par exemple : hanche NEAR fracture, fournit les fractures de la hanche.

Sur certains moteurs la distance maximum entre les mots est de 50 au lieu de 10 ;

– **NEAR/n** (n est un nombre inférieur ou égal à 1 000) : les deux mots sont proches de n mots ou moins ;

– **ADJ** : les deux mots-clés sont à deux mots ou moins, l'un de l'autre ;

– \* : remplace une chaîne de caractères manquant (informati\* = information, informatique, etc.).

Les terminaisons n'étant pas identiques en français et en anglais, l'opérateur \* peut servir à effectuer une recherche simultanée dans les deux langues (scan\* trouvera toutes les occurrences de scanner et scanning) ;

– ? : remplace un caractère manquant (psycholog?e = psychologue, psychologie)

L'utilisation de **parenthèses** permet de regrouper des phases de traitement de la recherche.

Dans l'exemple suivant, nous recherchons les pages parlant de la schizophrénie et du traitement par psychotropes ou neuroleptiques. Nous excluons les pages qui traitent de la forme borderline de la schizophrénie :

(psychotrope OR neuroleptique) AND schizophrénie AND NOT borderline

Nous recherchons les pages traitant des consensus sur les lombalgies et celles qui traitent du même sujet sur les sciatiques :

consensus NEAR (lombalgie OR sciatique)

Sur leur moteur de recherche, les éditions Masson ont proposé cet exemple, pour rechercher les documents écrits par les Professeurs Dupont et Jones traitant de fracture :

professeur AND (Dupont NEAR Jones) AND ((Dupont OR Jones) NEAR fracture).

Pour ceux qui préfèrent une explication visuelle, le CHU de Rouen image l'usage des opérateurs booléens.



Figure 5 : L'usage des opérateurs booléens dans les pages d'aide du CISMEF-CHU de Rouen.

## Quelques outils de recherche dans le domaine médical

### Français

CISMEF (CHU de Rouen) : <http://www.chu-rouen.fr/ssf/ssf.html>

Atmédica : <http://www.atmedica.com/>

Medisite : <http://www.medisite.fr/>

Dentaire-village : <http://www.dentaire-village.com/>

### Anglo-saxons spécialisés dans les bases de données scientifiques

<http://www.ussc.alltheWeb.com> (Fast serait un des plus performants).

<http://galaxy.einet.net/GJ/medical.html>

<http://www.medmatrix.org/Index.asp>

<http://healthWeb.org/>

<http://www.mic.ki.se/Diseases/index.html>

<http://www.oshu.edu/clinWeb>

<http://omni.ac.uk/search/>

<http://www.arcade.uiowa.edu/hardin-www/md.html>

<http://www.docnet.org.uk/>

<http://www.emory.edu/WHISCL/search/AT-MedWebquery.html>

<http://www.emory.edu/WHISCL/search/AT-MedWebquery.html>

<http://www-informatics.ucdmc.ucdavis.edu/>

<http://smartsearch.uthscsa.edu/searchform3.htm> (orientation EBM).

<http://www.anesthescience.com/index.phtml?idb=17 & lg=fr>

<http://www.medicinescience.com/index.phtml?idb=5 & lg=fr>

<http://www.orthoscience.com/index.phtml?idb=9 & lg=fr>

**Suisse :**

[http://www.hon.ch/MedHunt/MedHunt\\_f.html](http://www.hon.ch/MedHunt/MedHunt_f.html)