

À propos de la recherche sur le web

En documentation, on estime qu'une recherche est performante si elle renvoie toutes les informations cherchées et uniquement ces informations. La recherche sur le web répond rarement à ces critères : le résultat comporte souvent beaucoup de **bruit** (on obtient une foule de réponses non pertinentes), ou à l'inverse un taux de **silence** qui peut être trompeur (on ne trouve pas forcément tous les sites pertinents).

On obtiendra de meilleurs résultats en choisissant un type d'outil et une stratégie d'interrogation adaptés à la recherche à faire. Quelques notions pour vous y aider :

Moteurs de recherche : le principe

Les moteurs de recherche ne cherchent jamais "en direct" sur le web mais dans des bases de données préalablement constituées et régulièrement mises à jour : le résultat de la recherche dépend donc de ce qui a été référencé ou non dans la base.

Des réponses extrêmement inégales selon le type de moteur utilisé :

- Les robots Bases de données énormes, constituées automatiquement par des programmes informatiques
→ *réponses abondantes, mais beaucoup de bruit*
- Les index humains Bases de données de taille plus réduite, mais les informations sont sélectionnées et généralement classifiées par des documentalistes.
→ *moins de réponses (risque de silence), mais globalement plus pertinentes*
- Afin de mieux cibler l'information, les moteurs les plus efficaces allient de plus en plus souvent les deux méthodes : ils confrontent les résultats extraits d'une base de données vérifiée et classifiée par l'homme (répertoire) et d'une autre constituée par un robot. C'est le cas de Google, Yahoo et Altavista.

Les modes de recherche

■ Navigation dans un répertoire thématique (ou "catalogue" ou "annuaire")

Les systèmes d'annuaire présentent les ressources classées par catégories et sous-catégories : cela permet un ciblage progressif de l'information, par sélection de catégories de plus en plus fines. La plupart du temps, le classement est thématique.

Les répertoires sont très utiles pour une recherche sur un thème vaste ou quand les mots-clés qui le caractérisent sont d'usage fréquent, ou polysémiques, ou encore si l'un de leurs contextes possibles est très en vogue mais non pertinent pour vous...

Exemple : "feuille de styles", renverra une masse de sites spécialisés dans la conception de pages web, mais très peu de choses sur les styles en traitement de textes.

■ Interrogation par mots clés : les différents modes

- *Le mode "recherche simple"* : c'est le mode le plus facile mais le moins efficace car il ne permet guère de moduler la requête. Dans le meilleur des cas, on disposera de quelques opérateurs basiques :

-mot mot interdit

+mot mot exigé

"chien et chat" les guillemets forcent la recherche de l'expression littérale au lieu des mots isolés.

La plupart des moteurs ignorent les "mots vides" (articles, pronoms, prépositions...) sauf s'ils sont précédés du signe +, et ne tiennent pas compte des signes de ponctuation à l'intérieur des guillemets,

En général, ils renvoient en priorité les pages où tous les mots cherchés sont présents et assez rapprochés, puis celles où ils sont présents mais assez éloignés.

- *Le mode "recherche avancée"*

Il permet d'exprimer des requêtes plus complexes – donc plus précises – basées sur les opérateurs logiques ET, OU, NON dits "opérateurs booléens". La syntaxe varie selon les moteurs.

- *Restriction du champ de la recherche :*

Par défaut, les mots clés sont cherchés à la fois dans le texte des pages, dans leur url (adresse) et dans le titre de la fenêtre. Certains moteurs permettent de ne les chercher que dans l'une ou l'autre de ces zones.

■ Recherche par mots clés dans un répertoire thématique

Quand le moteur offre à la fois une recherche par mots clés et un répertoire, on peut souvent restreindre la portée d'une interrogation par mots clés à une catégorie du répertoire. C'est une bonne méthode pour cibler un thème pour lequel on ne dispose pas d'un vocabulaire spécialisé.

■ La règle d'or...

Consulter les pages d'aide sur la recherche et sur la recherche avancée pour connaître les possibilités et la syntaxe propres au moteur de recherche utilisé.

Connaître la syntaxe des URL pour mieux s'en sortir...

On appelle URL les adresses des ressources disponibles sur Internet. *Connaître la syntaxe des URL permet de mieux maîtriser la situation.*

Pour une page web ou document téléchargeable, la syntaxe de base est la suivante :

Protocole	Serveur .	Domaine	Chemin d'accès à la ressource dans le disque du serveur
<i>http://</i>	<i>sites</i>	<i>. univ-lyon2.fr</i>	<i>/lettres/infopratiq/navig.html</i>

Chemin d'accès à la ressource : cascade de dossiers jusqu'à celui où se trouve la ressource, suivie du nom de la ressource elle-même (ici le document "navig.html" rangé dans le dossier "infopratiq", lui-même situé dans le dossier "lettres").

L'extension attachée au nom de la ressource renseigne sur sa nature, donc sur ce qui se passera si l'on clique sur le lien. Cas les plus fréquents :

- `.html, .htm` → pages web,
- `.php, .asp, .js` → la plupart du temps le lien affiche une page web, mais l'affichage peut s'accompagner de l'exécution d'un script (petit programme informatique).
- `.pdf` → document qui sera téléchargé et/ou visualisé par Acrobat Reader – ou tout autre logiciel reconnaissant le format pdf.
- `.zip, .sit, .tar, .hqx` → documents compressés qui seront téléchargés sur le disque dur et nécessitent un logiciel de décompression (Unzip, Stuffit Expander, etc.).

Quand la ressource n'est pas un document stocké sur le serveur, mais résulte de l'interrogation d'une base de données, la fin de l'url comporte un point d'interrogation suivi des éléments exprimant la requête.

Exemple : <http://www.idlivre.com/523c-009/TheNews.cfm?Ref=1002>

Aspect pratique : les astuces

- ◆ *Un menu contextuel se cache sous chaque lien*, qui offre des commandes très utiles. Pour le dérouler :

- sur Macintosh : CTRL-clic
- sous Windows : bouton droit de la souris

- ◆ Quand on dépouille le résultat d'une requête, on gagne du temps en affichant les pages trouvées dans une nouvelle fenêtre de navigation :

- menu contextuel du lien, commande "Ouvrir dans une nouvelle fenêtre" (ou *approchante*).

- ◆ Pour récupérer l'adresse d'un lien par copier-coller sans avoir à afficher la page :

- menu contextuel du lien, commande "copier l'adresse du lien" ou "copier l'url"...

- ◆ D'une manière générale, *quand on pointe sur un lien (sans cliquer), on voit l'URL correspondante dans le bas de la fenêtre* : utile pour éviter de retourner sur un site que l'on a déjà exploré.

- ◆ Quand une adresse est périmée, il arrive souvent que la page soit toujours là, mais qu'elle ait été renommée ou déplacée dans le serveur. En examinant l'URL, on peut essayer de "remonter" à un niveau supérieur pour trouver un lien valide (voir ci-dessus *la syntaxe des Urls*).

Exemple : Si l'url "sites.univ-lyon2.fr/lettres/memosfac/ponctuation.html" ne fonctionne pas, on essaiera de remonter à "sites.univ-lyon2.fr/lettres/memosfac/", puis, en cas de nouvel échec, "sites.univ-lyon2.fr/lettres/"...

- ◆ Restriction de la recherche à une zone d'information

Google, Yahoo et Altavista permettent tout trois de restreindre le champ de la recherche de manière simple, en jouant sur les options de leur page "Recherche avancée".

Avec Google et Altavista, certaines restrictions peuvent aussi s'exprimer directement dans une recherche simple, via une syntaxe spécifique :

Sur Google :

link:url

link:www.bnf.fr trouve les pages qui comportent un lien vers la BNF

site:domaine ou site

site:scd.univ-lyon2.fr restreint la recherche aux pages du site "scd.univ-lyon.fr"

site:univ-lyon2.fr restreint la recherche aux pages des sites qui sont situés dans le domaine "univ-lyon.fr"