

# LA RECHERCHE SUR INTERNET

## 1. Les moteurs de recherche

### 1.1 Qu'est-ce qu'un moteur de recherche?

En anglais « Search engine », un moteur de recherche est un outil qui recense automatiquement des ressources se trouvant sur Internet. Par opposition aux répertoires, un moteur offre une base de données référencant des pages en texte intégral.

### 1.2. Fonctionnement général


Les moteurs de recherche fonctionnent à partir d'index construits automatiquement à partir de la collecte de robots (en anglais robots, spiders ou encore crawlers) qui parcourent le Web en permanence et engrangent les résultats d'exploration des pages Web.


### 1.3. Les robots

L'ensemble des pages Web est visité de manière régulière par des logiciels, appelés robots, qui les répertorient. À partir de cette collecte, les gestionnaires des moteurs constituent des bases de données qui sont ensuite traitées par des logiciels d'indexation. L'indexation est effectuée sur le texte intégral. Elle peut porter sur les titres seuls, sur l'ensemble de la page, sur les premiers paragraphes, sur les « meta tags », sur les en-têtes de paragraphes...

Un robot a pour seule mission d'explorer le net. Il y tourne 24 heures sur 24. Un robot international tel que celui d'Altavista (robot nommé Scooter) fait le tour de l'Internet en six semaines.

### 1.4. Quelques moteurs de recherche généralistes

 **GOOGLE** (<http://www.google.fr>)  
N'est plus à présenter, créé en 1998, Google n'a pas cessé d'être le meilleur depuis qu'il est lancé.

 **SEEKS** : un moteur de recherche libre (<http://www.seekr.fr>)  
Un moteur de recherche open source et anonyme  
Seeks est en fait un méta moteur de recherche. Cela signifie qu'il envoie ses requêtes (ce qu'on lui demande de chercher) à d'autres moteurs de recherche, comme Google ou Bing, puis retourne les résultats obtenus à l'utilisateur. Notez que les doublons sont inexistant : un même résultat trouvé sur Google ou Yahoo! Search ne sera affiché qu'une seule fois.

## 2. Les Annuaire ou Répertoires

Le premier des grands outils de recherche, dans l'ordre logique (historiquement, ils sont apparus en même temps que les moteurs), ce fut l'annuaire.

Deux étudiants américains se sont mis en tête de répertorier l'ensemble des sites Web. À l'origine, il était possible de rêver de couvrir artisanalement et manuellement l'ensemble des sites de la toile, alors qu'il ne s'en trouvait que par milliers. C'est ainsi que « Yet another hierarchical officious oracle » est né, en d'autres termes Yahoo! (Le point d'exclamation fait partie de la marque.)

### 2.1 Comment fonctionne ce type d'outil ?

Il s'agit d'une base de données recensant des références de sites.

La première caractéristique est donc que les répertoires recensent des sites et non des pages. L'autre caractéristique est que chaque site est en fait signalé au moyen d'une notice descriptive.

C'est donc une base de références de sites, tout ce qu'il y a de plus documentaire ou bibliothéconomique. Ce type d'outil s'oppose donc presque terme à terme aux moteurs de recherche.

### 2.2 Comment rechercher dans un répertoire ?

Pour s'y retrouver dans cette base elle-même il existe deux possibilités.

D'une part on peut naviguer au travers d'une classification appelée catégories, en anglais. Chaque site est rangé dans autant de catégories ou sous-catégories qu'il traite de sujets appartenant aux dites catégories. C'est le système documentaire le plus classique qui soit et, au passage, on peut s'amuser, alors qu'on nous prédit la mort des professions de l'I-D face à l'Internet, de voir que les bonnes vieilles

techniques de classification sont ici remises à l'honneur...

L'autre possibilité est de rechercher par un des mots qui sont présents dans les fiches descriptives. C'est donc de l'informatique documentaire de base... Une fenêtre de recherche est toujours disponible en tête de la page d'accueil du répertoire.

### 2.3 Quelques annuaires généralistes

<http://www.google.fr/dirhp?hl=fr> L'annuaire de Google

très complet, de nombreuses rubriques classées par thématiques et sous-catégories.

## 3. Les critères de la recherche.

Une bonne recherche se doit d'être efficace. Vous disposez de certains outils pour vous simplifier la vie, voici quelques conseils et astuces :

- Chaque terme de recherche compte
- La recherche ne tient jamais compte de la casse
- En règle générale, les signes de ponctuation, notamment les caractères @#\$%^&\*()=+[]\ et les autres caractères spéciaux, sont ignorés.
- Utilisez des mots descriptifs, le moins de termes possibles.
- Restez simple, et ordonnez en fonction de la priorité désirée

Pour la recherche avancée voici les clefs :

- recherche de synonymes : placer un ~devant le mot (CTRL+ALT+2 et taper le mot accolé au signe)
- les pages contenant au moins un des noms écrits : mot1 + mot2 + mot3 etc...
- les pages contenant l'un OU l'autre des mots : mot1 OR mot2 (à écrire en capitales)
- les pages contenant tous les noms : +mot +mot +mot
- rechercher dans un site précis : mot site:www.domain.com
- rechercher une phrase exacte : "mettre entre guillemet"
- rechercher une définition : define:mot
- rechercher quelquechose mais pas autre : mot -mot (par exemple : virus -computer)
- utilisation de la troncature (\*) : tous les termes commençant par mot\* (résultat : motivé, motard, moto, etc...)

### Récapitulatif :

Opérateur	Résultat
<b>+nom1 +nom2</b>	Renvoie les documents contenant les 2 mots cherchés
<b>nom1 + nom2</b> ou <b>nom1 nom2</b>	Renvoie les documents contenant au moins un des 2 mots cherchés (ou les 2)
<b>+nom1 -nom2</b>	Renvoie les documents ne contenant pas le mot qui suit l'opérateur -
<b>nom*</b>	Renvoie les documents contenant les mots proches du mot cherché
<b>guillemets</b>	Renvoie les documents contenant la phrase entière
<b>Essentiel (+)</b>	Le symbole "+" classifie un mot comme essentiel
<b>Exclusion (-)</b>	Le symbole "-" exclut un mot de la recherche