Cours de technologie du web: LA RECHERCHE SUR INTERNET

MIAS: Technologie Web

Université Badji Mokhtar ANNABA

Chapitre 3: La recherche sur internet

Aide-mémoire des techniques et **outils de recherche d'informations** classiques sur le World Wide Web, et ce indépendamment du navigateur utilisé.

1- Comment se présente un outil de recherche

Une recherche s'initie à partir de la page d'accueil d'un **outil de recherche**, page accessible par son adresse web (URL). Il vous est donc conseillé de garder parmi vos signets les URL de vos outils de recherche préférés.
Les pages d'accueil varient grandement dans leur contenu et leur présentation. Après une époque où ces pages se sont chargées de plus en plus, la tendance est de revenir à plus de sobriété pour plus de clarté. Un outil de recherche présente en général les catégories suivantes :

- Une boite d'interaction permettant de saisir la requête de recherche (cas des moteurs)
- Options et préférences qui permettent de configurer le fonctionnement de la recherche, la présentation des résultats, …
- · Liste des sujets principaux (cas des annuaires)
- · Une liste de sites populaires permettant d'accéder soit à des pages, soit à des services qui ont du succès (durée éphémère). L'outil de recherche ne remplit ici pas sa fonction première, mais plutôt celle d'un passage obligé vers l'Internet (portail)
- · Section d'aide et de réponses à des questions fréquemment posées (FAQ, Frequently Asked Questions)
- · Publicité, Nouvelles, Promotions en tous genres

2- Differents Outils et Méthodes de recherche

Il existe essentiellement 4 types d'outils de recherche et donc de méthodes de recherche associées, décrits brièvement ci-après.

2.1 Annuaires

Les annuaires (en anglais, directory search) permettent une recherche par sujet. Cela consiste en une recherche hiérarchique débutant par un sujet général que l'on affine au fur et à mesure. Il existe des annuaires généraux et des annuaires thématiques.

- · *Pour quel besoin* : information générale et pertinente sur un sujet donné.

- · *Avantages* : relativement facile à utiliser ; l'information répertoriée dans l'annuaire a été sélectionnée par des humains donc les informations sont pertinentes.
- · *Inconvénients* : Comme l'indexation est longue car manuelle, la quantité d'information répertoriée ainsi que la fréquence de mise à jour sont faibles.

2.2 Moteurs

Les moteurs de recherche (en anglais, search engine) permettent une recherche par mot clé. Leur réponse est une liste de références (hits) triée par ordre décroissant de pertinence par rapport à la formulation de la requête. L'alimentation de la base d'information est faite par des robots logiciels qui scrutent le web en permanence. La technique de rangement ou classement dans la base s'appelle l'indexation

- · *Pour quel besoin* : besoin d'information spécifique et parfois aussi, exhaustive.
- · *Avantages* : La quantité d'information répertoriée couvre une grande proportion des pages sur le web (par exemple de l'ordre de 50 %) et les mises à jour sont fréquentes.
- *Inconvénients* : La recherche d'une information précise dans l'immense base de données peut s'avérer délicate (nécessite de l'exactitude dans la formulation via par exemple, des expressions booléennes) et la technique des mots clés a ses limites.

2.3 Annuaires + Moteurs

Les annuaires connectés à des moteurs de recherche (en anglais, directory with search engine) permettent une recherche combinée par sujet et par mot clé. A chaque étape de raffinement du sujet, il est possible de passer en mode de recherche par mot-clé et vice versa, pour trouver de l'information répertoriée dans l'annuaire.

- · *Pour quel besoin :* lorsque on ne sait pas à l'avance laquelle des méthodes par sujet ou par mot-clé donnera les meilleurs résultats.
- · *Avantages :* à la fin de la liste des résultats de la recherche dans l'annuaire, il est souvent possible d'enchaîner avec une recherche classique par mot-clé, via la connexion à un moteur de recherche classique.
- · *Inconvénients :* la technique peut s'avérer inefficace pour des recherches complexes à exprimer

2.4 Meta-moteurs

Les méta-moteurs (en anglais multi-engine) font appel à un certain nombre de moteurs de recherche en parallèle, sans être eux-mêmes des moteurs.

- · *Pour quel besoin* : accélérer le processus de recherche et éviter les doublons.
- · *Avantages* : Plus tolérant qu'un seul moteur de recherche face à des requêtes imprécises. Fournit moins de hits de plus grande pertinence.

- · *Inconvénients* : Pas aussi efficace que l'utilisation d'un seul moteur de recherche lorsque la requête est complexe.

3- Guide pour la recherche par mot-clé

Règles à connaître pour exprimer une requête de recherche à base de mots clés. Seules seront listées les adresses des pages web pour lesquelles la requête correspond.

La présence de Majuscules/Minuscules influent sur la recherche.Pour un mot entièrement écrit en minuscules (ex, mot), le moteur recherchera les mots-clé mot, Mot, MOT, MOt, MoT, mOT, mOt, moT… en bref la recherche ne tiendra pas compte de la casse. Dès qu'une majuscule apparaît (ex, Mot), alors le moteur considère le mot-clé tel quel .

De même , la langue et l'alphabet peuvent être pris en compte. De nombreux moteurs de recherche permettent de ne considérer que les pages écrites dans une ou plusieurs langues que l'utilisateur aura préalablement précisées. Idem concernant l'alphabet. Par défaut, toutes les pages, quelque que soit la ou les langues dans lesquelles elles sont rédigées sont considérées.

La troncature s'utilise lorsque l'on ne connaît pas exactement la terminaison d'un mot clé. On ajoute le symbole joker (en anglais, wildcard) représenté couramment par le caractère * (ex, Mot ou Mots pourra être décrit grâce à Mot* mais attention, vous obtiendrez aussi Moto, Motard,…)

Une phrase, un proposition, ou tout autre regroupement de mots peut être donnée. Une expression constituée de plusieurs mots s'entoure de guillemets par exemple « Un Mot ». Le moteur recherche non pas l'occurrence de 'Un' , ni de 'Mot', mais celle de la chaîne « Un Mot » exactement.

Forcer la présence d'un mot-clé s'exprime en faisant précéder le mot-clé de + (attention, pas d'espace entre le symbole et le mot). Exclure les pages dans lesquelles apparaît un mot-clé s'exprime en précédant le mot-clé de –

En l'absence d'un + ou d'un -, chaque moteur a son propre comportement par défaut, il vaut donc mieux préciser

4- Opérateurs booléens

Ils servent à connecter des termes (mots ou chaînes de mots), à la manière des expressions logiques, l'utilisation de ( ) peut être nécessaire. Ce style de recherche rentre souvent dans la catégorie de recherche avancée il faut donc d'abord se positionner dans ce mode particulier

- · t1 AND t2 : les 2 termes t1 ET t2 doivent être présents dans la page pour qu'elle soit sélectionnée
- · t1 OR t2 : au moins un des 2 termes t1 OU t2 doit être présent
- · NOT t : le terme t ne doit pas être présent pour que la page soit sélectionnée
- · t1 NEAR t2 : les 2 termes t1 et t2 doivent se trouver A PROXIMITE l'un de l'autre dans le texte (par exemple, distants d'au plus 10 mots). L'opérateur est commutatif.

5- Position dans la page

Un terme peut apparaître à différents endroits dans une page : dans le texte lui-même, ou bien dans l'adresse (URL) de la page, ou bien encore dans le titre, etc…

Il suffit de faire précéder le terme par l'emplacement suivi de : (ex, url :mot sélectionne les pages dont l'url contient mot). Le moteur altavista est le plus complet et fournit les possibilités ci-dessous

Emplacements	Signification
anchor:texte	Le texte d'un hyperlien contient le texte indiqué
domain:nomdomaine	Les pages sur les sites dont l'adresse internet termine par nomdomaine sont considérées
host:nom	Ne recherche que sur des pages hébergées sur les machines dont l'adresse internet contient nom
image:nomfichier	Sélectionne les pages contenant une image stockée dans le fichier nomfichier
link:urltext	Sélectionne les pages qui ont un lien pointant vers l'url désignée par urltext
text:texte	Sélectionne les pages qui contiennent le texte, en excluant les zones correspondant à une url, à un lien, à un titre d'image
title:texte	Sélectionne les pages dont le titre contient le texte
url:texte	Sélectionne les pages dont l'adresse web contient texte.