

La recherche documentaire sur Internet

Introduction :

Pour se repérer dans l'énorme masse de documents accessibles sur Internet, différentes sortes d'outils sont disponibles comme les moteurs de recherche. Ils sont très efficaces, mais ils n'explorent qu'une partie des ressources, laissant de côté le « Web Invisible » (sites accessibles avec un mot de passe, certaines listes de diffusion...) Ce sont aussi maintenant, des outils commerciaux, non contrôlés par des instances scientifiques ou démocratiques. Il est donc important de connaître les autres instruments de recherche que sont :

Les répertoires ou annuaires : appelés aussi "répertoires", ils proposent une **sélection de sites Web** (et non de pages). Les sites sont décrits (titre, phrase de présentation, adresse URL) et organisés de manière **hiérarchique en catégories et sous-catégories thématiques** selon le principe de l'arborescence. Ce travail est réalisé par des **humains**, et non des machines, ce qui augmente la **pertinence** de la recherche (par rapport à un moteur) mais entraîne également un **manque d'exhaustivité** et des **misés à jour peu rapides**. <http://www.aef-dmoz.org/>

Les portails : il s'agit d'une porte d'entrée sur Internet, constituée d'une **page d'accueil d'un site** mettant à disposition un **large ensemble de ressources et de services**, intérieurs ou extérieurs au site. Il existe différentes catégories de portails : généralistes ou spécialisés, professionnels ou grand public, émanant d'entreprises ou d'institutions publiques. Ex. : le portail de l'Union européenne : http://europa.eu/index_fr.htm

Les sites et catalogues de bibliothèques : ces sites donnent accès à des ressources souvent invisibles pour les robots, mais surtout à des ressources validées, sélectionnées par des spécialistes, selon des procédures offrant des garanties scientifiques.

Les différents moteurs de recherche intéressants à utiliser :

Exalead <http://www.exalead.fr/search>

Ce moteur français propose des mots-clés complémentaires (termes associés) et renvoie vers des annuaires, des fichiers multimédias... Grâce à cette organisation de mots-clés, le bruit documentaire (surabondance de d'informations peu ou pas pertinentes) est limité.

.Mozbot <http://www.mozbot.fr>

Il a le même index que Google, soit environ 10 milliards de pages, mais en plus de Google, il a une fonctionnalité particulière : le lien « info » situé sous l'adresse du site trouvé, qui permet d'accéder à toute une série d'informations sur l'auteur et le site lui-même. Il est donc intéressant quand il s'agit de vérifier la fiabilité d'une source.

Kartoo <http://www.kartoo.com/flash.php3> (métamoteur)

Les métamoteurs sont des « super moteurs » qui **interrogent simultanément plusieurs moteurs et annuaires** à partir d'une même requête. Ils éliminent ensuite les doublons et organisent généralement les résultats selon des modes spécifiques : sous forme de **liste (Ixquick)** ou sous **forme cartographique (Kartoo)**.

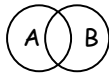
Très original, ce métamoteur affiche les résultats sous la forme d'un réseau cartographique. en revanche, il offre une lisibilité assez limitée...A voir tout de même car il relie les mots-clés à des notions équivalentes

Savoir utiliser un moteur de recherche : la recherche avancée :

Proposée par plusieurs moteurs de recherche, elle permet de restreindre, d'affiner la recherche (quand les résultats sont trop nombreux par exemple) et donc d'avoir des résultats plus pertinents. La structure même de la recherche avancée sous-tend l'usage de 3 connecteurs logiques (booléens) : ET, OU, SAUF. On peut ainsi multiplier le nombre de mots clés, les combiner

entre eux. Elle offre aussi d'autres possibilités de précision comme le choix de la langue, le format de fichier, la date de mise à jour...

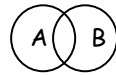
La combinaison des mots clés



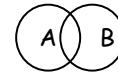
ET (intersection)

SAUF (exclusion)

Recherche à la fois sur A et sur B



sur A, sur B ou sur les deux à la fois



OU (réunion)

sur A sauf sur B

Evaluer la qualité de l'information

➤ Lire et analyser la page de résultats :

Dans un moteur de recherche, avant même d'ouvrir une page Web, il est nécessaire d'analyser la page de résultats : quels sont les sites susceptibles d'être visités pour ma recherche, quels sont ceux que j'élimine d'office. Pour effectuer ce travail d'analyse, je prélève un certain nombre d'indices dans les titres, les extraits et les adresses URL.

L'adresse URL du site peut vous aider à identifier sa nature : une adresse URL se compose notamment d'un nom de domaine, qui comprend un suffixe. Par exemple dans l'adresse « <http://www.liens-socio.org/> », « .org » est le suffixe. Chaque suffixe correspond à un type de site. Cependant, les règles qui régissent l'attribution des noms de domaine évoluent et cette correspondance est de moins en moins vraie.

Il existe deux types de suffixe :

le suffixe géographique composé de deux lettres identifiant un pays : « .fr » signifie que l'on va consulter un site français (ce sera .uk pour l'Angleterre, .ca pour le Canada...). Attention certains pays comme la Russie n'ont aucune exigence en ce qui concerne le contenu d'un site

le suffixe générique : .com : sites à vocation commerciale

.org : sites d'organismes à but non lucratif (associations, musées, organisations...)

.edu : sites du domaine de l'éducation

.gouv.fr : sites gouvernementaux français

.info : sites à vocation informationnelle

Les adresses URL contenant « .pagesperso-orange.fr » ou « .perso.free.fr », par exemple, renvoient à des sites personnels dont la fiabilité n'est pas assurée. Il faut alors rechercher le métier et les compétences de l'auteur.

➤ Evaluer la fiabilité de l'information sur le Web

Avant de sélectionner une information sur une page Web, il faut non seulement vérifier qu'elle correspond bien à ce que l'on recherche (actualité de l'info, lisibilité...) mais aussi s'assurer de sa qualité, de sa fiabilité. Une information fiable est une information dont on connaît obligatoirement l'auteur, ses compétences et son intention.

Quelques astuces pour vous aider :

➤ L'auteur d'un site : trouver l'auteur n'est pas toujours évident. Quelques pistes :

- Sue le site lui-même, consulter les rubriques « Qui sommes-nous ? », « présentation » ou « copyright » de la page d'accueil

- Les sites présentent parfois une rubrique « Contact », avec le mail de l'auteur.

- On peut aussi faire « affichage », « sources » dans la barre d'outils

- Sur un moteur de recherche, faire une recherche en plein texte sur l'auteur lui-même

- Dans Google, rechercher tous les sites qui contiennent un lien vers le site en question : link :http:// + adresse. (dans Yahoo, linkdomain : http:// + adresse)

- Faire des recherches sur des sites spécialisés : <http://www.afnic.fr/outils/whois>

➤ La date de publication et la date de mise à jour du site :

- La date de publication peut se situer en bas de page ou dans la rubrique « copyright »

- Pour avoir une idée de la date de mise à jour, on peut vérifier à quelle date a été publié le dernier article

➤ La nature d'un site et l'intention de son auteur :

Il existe différents types de site : personnel, institutionnel, commercial, collaboratif, associatif, universitaire, éducatif...