

MEMENTO RECHERCHE SUR INTERNET

La recherche sur Internet n'est pas chose aisée et ce pour de multiples raisons : abondance de l'information, hétérogénéité et fragmentation des sources et des contenus, renouvellement permanent (caractère volatile de l'information), multilinguisme, fiabilité de l'information... L'enjeu majeur sur le web est de capter et de garder l'attention des usagers : en quelque sorte, les internautes paient avec le temps qu'ils passent sur les contenus, c'est de ce temps que veulent les annonceurs. Il convient alors de faire la différence entre les liens naturels et les liens sponsorisés. Derrière Internet, il y a d'énormes enjeux économiques, culturels, cognitifs, politiques, sociétaux... liés aux outils et aux pratiques de recherche.

1-Méthodologie de la recherche sur Internet

Quelques règles à suivre et intégrer dans sa démarche de recherche :

MAÎTRISER LES OUTILS DE NAVIGATION	-maîtriser l'interrogation avancée, -utiliser différents outils.
SAVOIR QUESTIONNER	-identifier l'objet de la recherche, -bien cerner le sujet, -traduire les concepts en mots-clefs, -connaître les opérateurs de recherche (guillemets pour rechercher une expression, opérateurs booléens, filetype : pour trouver les pages créées dans un format de fichier particulier...°) -identifier les sources pertinentes.
CHOISIR LES BONS MOTS-CLEFS	

CHERCHER L'INFORMATION À LA SOURCE	-trouver les sites de référence, -privilégier les sources primaires.
TOUJOURS ANALYSER L'INFORMATION	-recouper, -critiquer, -comparer.
GARDER LES TRACES	-historique, -outils de gestion des signets (Delicious).
MAÎTRISER LE TEMPS	-faire un plan de recherche, -ne pas vouloir atteindre l'exhaustivité.
RESTER CLAIR SUR SES OBJECTIFS	
UTILISER DIFFÉRENTS TYPES D'OUTILS	moteurs, métamoteurs, annuaires...
ÊTRE « AGILE »	-lecture rapide, -rebondir d'un document à l'autre... sans non plus passer trop vite (zapping).

2-Panorama des outils disponibles

Il existe plusieurs catégories d'outils : annuaires thématiques, moteurs de recherche, métamoteurs, outils de veille, portails et sites de référence, signets...

Ces outils se répertorient suivant deux principales typologies :

-en fonction du mode de collecte et d'indexation des ressources, on distingue les outils « humains » (annuaires, sélections de signets, plates-formes de partage de signets, portails de fils RSS) des outils automatisés (moteurs de recherche, métamoteurs).

-En fonction des ressources recensées, on distingue les outils généralistes qui renvoient à l'ensemble d'Internet des outils spécialisés qui renvoient à un sous-ensemble d'Internet.

On peut repérer aussi trois grands critères de spécialisation de ces outils :

-selon les supports internet : moteurs de blogs, de forums, de listes de diffusion, de FAQ, de wikis, de fils RSS...

-selon la nature du contenu : moteurs scientifiques, d'actualité, de personnes, de produits...

-selon la nature du média : moteurs d'images, de vidéos, de podcasts, de photos...

Les outils spécialisés sont donc très importants pour faire des recherches précises.

-Les annuaires thématiques

Outil de recherche (appelé aussi index ou *répertoire*), qui classe un ensemble de sites web suivant une thématique arborescente. Les annuaires sont les premiers outils du web : Virtual Library en 1991, Yahoo en 1994. Il existe plusieurs types d'annuaires :

- des annuaires généralistes (Yahoo) ou spécialisés (UREC),
- des annuaires privés (Yahoo) ou publics (Virtual Library),
- des annuaires collectifs (Open Directory) ou individuels (Quartier Français...).

Quel que soit le type d'annuaires, ils s'appuient sur des principes communs :

- Sélection *humaine* de ressources,
- Recensement de *sites* web (et non de pages web),
- Pré-catégorisation* des ressources,
- Arborescence,
- Description manuelle des sites,
- Moteur de recherche interne à l'annuaire.

Ce type d'outil pose le problème important du classement des sites, de la représentation des connaissances et des ressources. Depuis leur apparition, on peut noter un certain déclin des annuaires généralistes au profit du développement des annuaires spécialisés.

Utilité des annuaires: recherche d'information sur un sujet large, permet une première approche d'un domaine, permet de rechercher des sites de référence (Signets de la BNF).

Pour résumer les annuaires sont utiles pour défricher un domaine, pour effectuer les premières recherches sur un sujet. Ces outils sont complémentaires avec les moteurs de recherche.

-Les moteurs de recherche

Outil automatisé d'indexation et de recherche des ressources du web. Ils fonctionnent grâce à trois modules autonomes : un robot collecteur qui collecte des données, un module d'indexation (base de données du moteur) et un module de requête (gestion des requêtes et des résultats).

Ces moteurs proposent :

- différentes fonctionnalités de recherché: mode simple et mode avancé
- différentes syntaxes de recherche : opérateurs de recherche (booléens, parenthèses), options linguistiques (casse, troncature, recherche d'une expression)

-des options de filtrage des requêtes.

Il est important de faire attention à la manière dont sont classés les résultats. Il existe deux grandes méthodes de classement :

-un **tri par indice de pertinence** : calculs statistiques sur la fréquence des termes...

-un **tri par popularité** qui s'appuie sur l'indice de **popularité** (calculé en fonction du nombre de liens qui pointent vers le site sélectionné). C'est de cette manière que Google présente ses résultats. Il faut aussi signaler que cette question du classement des sites l'est aussi pour les annonceurs et les entreprises qui souhaitent être classés à la meilleure position possible. Il existe donc de plus en plus de liens sponsorisés.

Ces outils proposent aussi différentes présentations des résultats : affichage des résultats, description des documents, exploitation qui peut en être faite...

Exemples de moteurs de recherche généralistes :

- Google,
- Yahoo Search,
- Bing,
- Orange,
- Exalead...

Exemples de moteurs de recherche spécialisés :

- Google scholar
- In-Extenso (sciences humaines et sociales)
- Spinoo (ressources éducatives)
- Truveo (moteur de recherche vidéos)...

-Les métamoteurs

Un métamoteur est un outil de recherche qui **interroge en parallèle plusieurs outils** (moteurs et annuaires), **rapatrie leurs réponses et les organise**, selon des méthodes de classement spécifiques, pour **fournir aux utilisateurs une présentation structurée** des résultats.

Leur intérêt principal est de simplifier la recherche en faisant une seule requête pour plusieurs outils. Ils permettent en outre un élargissement de la recherche, ils offrent la possibilité d'interroger plusieurs sources (d'une dizaine à un millier avec Copernic). La **pertinence des résultats** s'en trouve ainsi renforcée car la sélection des documents plus poussée (calcul de pertinence opéré sur plusieurs outils).

Les métamoteurs sont très utiles pour faire le panorama des ressources sur un thème, ils permettent une exploitation parfois sophistiquée des résultats. A signaler aussi, ce sont des outils qui peuvent être **paramétrables et personnalisables**

Exemples de métamoteurs :

- Ixquick,
- Kartoo,
- Seek

-Les outils de syndication de contenus

Ces outils permettent d'agréger des contenus divers sur un même navigateur, ils offrent la possibilité de publier automatiquement sur un site web des informations issues d'autres sites web.

Les fils RSS (*Rich Site Summary*, ou *Really Simple Syndication*) ont plusieurs fonctions :

- Suivre l'actualité d'un site : réception de toutes les nouveautés,
- Se tenir informé : sur l'actualité (tous les journaux proposent des fils RSS), sur un domaine (suivi de blogs et de sites spécialisés).

Il est aussi possible de créer des **pages personnalisables** qui offrent la possibilité de gérer et d'organiser soi-même ses flux. Ces pages web sont organisées en onglets et boîtes, leur utilisation est gratuite.

Exemples :Netvibes, PagesFlakes...

3-Évaluation de l'information sur Internet

L'information sur Internet se caractérise en particulier par la surabondance, la liberté totale de publication, l'absence de validation préalable. Beaucoup de fausses nouvelles, chaînes, pétitions, alertes aux virus ou promesses d'argent circulent aussi par messagerie. Divers sites répertorient ces "hoax". Par exemple, Hoaxbuster (<http://www.hoaxbuster.com>). Il importe donc de bien vérifier la validité et la qualité d'une information, avant de la citer ou de la diffuser.

Démarche en deux temps :

- identification de la ressource,
- évaluation de la ressource.

-L'identification d'une ressource électronique

Il faut **repérer, décrire et analyser les différents éléments constitutifs** du document. L'identification porte sur les caractéristiques extérieures, objectives de la ressource :

<p>ORGANISATION RESPONSABLE, ÉDITEUR DU SITE</p>	<ul style="list-style-type: none"> -nom de l'organisation? -"nationalité" du site? -nature du site? Universitaire, institutionnel, commercial, personnel... -statut ? Public ou privé, gratuit ou payant ? -objectifs ? Production-diffusion d'information, expression littéraire personnelle, vente... -public visé ? Tous publics, spécialisé... -notoriété du site, de l'organisation ? Quels sites "pointent" vers le site observé ?.
<p>AUTEUR DU DOCUMENT</p>	<ul style="list-style-type: none"> -auteur du document ? Est-il joignable ? -son statut, son domaine de compétence ? -sa notoriété ? Est-il cité par d'autres ? Si oui, par qui ? -A quels auteurs, à quels documents fait-il référence ? -Quel est son "point de vue" ? S'exprime-t-il en son nom propre ou au nom de l'organisation qui l'héberge ?
<p>DOCUMENT</p>	<ul style="list-style-type: none"> -"l'étendue" du document ? Un site entier, une partie d'un site, un document autonome... -type de support documentaire ? Site web, base de données, revue, forum... -dates du document ? Date de création, de mise à jour -nature du document ? Scientifique, de vulgarisation, d'expression... -domaine de connaissance, ou la discipline ?

-L'évaluation d'une ressource électronique

Elle consiste à **attribuer une valeur** (faible, moyen, bien, très bien) aux différents éléments de la ressource, selon des critères prédéfinis. On peut distinguer au moins cinq parties différentes, comportant chacune plusieurs critères :

<p>SOURCE, AUTEUR</p>	<ul style="list-style-type: none"> -Clarté, facilité d'identification de l'organisation, de l'auteur -Sérieux, fiabilité de l'organisation, de l'auteur -Degré de compétence dans le domaine -Notoriété
------------------------------	---

CONTENU	<ul style="list-style-type: none"> -Degré de fiabilité de l'information -Niveau de précision, d'exhaustivité, d'exactitude -Degré de nouveauté, d'originalité -"Fraicheur" de l'information, actualisation -Clarté d'indication et accessibilité des sources d'information -Pertinence, richesse des liens externes -Qualité de la langue (expression écrite, traduction) -Clarté de présentation de l'information -Utilité, pertinence des illustrations (schémas, images...)
STRUCTURATION DU DOCUMENT	<ul style="list-style-type: none"> -Logique de l'organisation : cohérence, clarté de la structure -Facilité d'orientation : sommaire, plan du site... -Facilité de navigation : aller-retour, retour page d'accueil, ascenseur... -Facilité de lecture des pages intérieures : sommaires internes, boutons de retour... -Qualité et efficacité du moteur de recherche interne
MISE EN FORME	<ul style="list-style-type: none"> -Design du site, choix des couleurs -Homogénéité de la mise en forme -Lisibilité du texte -Qualité des illustrations
ACCESSIBILITÉ, ASPECTS TECHNIQUES	<ul style="list-style-type: none"> -Rapidité de chargement du site et des pages, des illustrations -Accessibilité du site (référencement dans les outils de recherche)

► A chaque étape de l'évaluation, on vérifie la pertinence de la ressource, c'est-à-dire la correspondance entre le document et la requête : l'information trouvée correspond-elle au sujet de recherche ?

-Quelques conseils

-Déchiffrer l'adresse de la ressource repérée : la structure d'une adresse Internet ou URL (Uniform Resource Locator) est souvent source de renseignement.

URL d'une page d'accueil : Ministère de la culture : <<http://www.culture.gouv.fr>>

http : protocole technique de communication propre au web (*HyperText Transfert Protocol*)

www : Indicatif qui précise que le site est placé dans le *World Wide Web*

culture : nom de sous-domaine

gouv : nom de sous-domaine (organisme gouvernemental français)

fr : domaine - abréviation du nom du pays (France)

URL d'un document particulier : < <http://www.culture.gouv.fr/culture/bdd/index.html>

<http://www.culture.gouv.fr/> : adresse du site

culture/ : nom d'un répertoire

bdd/ : nom d'un sous-répertoire dédié aux bases de données

index : nom du document

.htm : extension du fichier qui renseigne sur le format du document page en langage htm

Les différents éléments de l'URL, tels que les **noms de domaines et sous-domaines**, peuvent contribuer à l'identification d'un site. En voici quelques uns :

-fr, .ca, .de, .it... abréviation du nom de pays (France, Canada, Allemagne, Italie...)

-.edu université américaine - ex : <http://www.georgetown.edu/>

-.com site commercial - ex: <http://www.lesbelleslettres.com/>

-.gov organisme gouvernemental américain -ex: <http://catalog.loc.gov/>

-.univ-fr université française - ex: <http://www.mshs.univ-poitiers.fr>

-.asso.fr association française - ex: <http://www.societes-savantes-toulouse.asso.fr/>

-Croiser ses sources : consulter plusieurs sources documentaires, en les comparant, en identifiant leur positionnement idéologique ou politique.

-Utiliser les signets sélectionnés par des professionnels (bibliothèques universitaires, Bibliothèque Nationale, institutions de recherche, etc ...) afin d'éviter des sites dépourvus de toutes garanties.

Pour aller plus loin :

-V. Mesguich et A. Thomas, *Net recherche 2009*, ADBS, 2009.

-Évaluer des sites Internet. CERISE [en ligne]. Paris : URFIST de Paris. [réf. Du 24 juillet 2013]. Disponible sur : < <http://urfist.enc.sorbonne.fr/anciensite/cerise/index.htm> >

-Méthodoc: Rennes : SCD de l'université Rennes 2. [réf. Du 24 juillet 2013]. Disponible sur : <<http://methodoc.univ-rennes2.fr/>>

-Sapristi ! Lyon: SCD Doc'INSA. [réf. Du 24 juillet 2013]. Disponible sur <<http://sapristi-docinsa.insa-lyon.fr/>>