

Recherche et veille sur la toile (web)

Inventaire raisonné des moyens utilisables, gratuits et payants

Leur utilité et leur insertion dans l'enseignement

Sommaire

| | |
|---|-----------|
| I - Inventaire..... | 2 |
| 1 Annuaire..... | 2 |
| 2 Moteurs de recherche..... | 2 |
| 3 Métamoteurs..... | 3 |
| 4 Les outils spécialisés | 4 |
| 5 La toile profonde ou invisible et les bases de données | 4 |
| 6 Les blogs..... | 6 |
| 7 Les travaux de groupes | 6 |
| 8 Les agents dits intelligents (cf. introduction)..... | 7 |
| 9 Trouver le bon outil et rester informé de l'évolution des autres..... | 7 |
| II - Interpréter les résultats..... | 8 |
| 1 Validité de l'information sur Internet | 8 |
| 2 Les critères d'évaluation | 8 |
| 3 Les outils d'analyse et de synthèse textuelle ou cartographique | 9 |
| III - Pérenniser la recherche par la veille | 11 |
| 1 Pour les pages nouvelles dans un site encore inconnu | 11 |
| 2 Pour les pages modifiées ou ajoutées dans un site déjà connu | 11 |
| 3 Les listes de diffusion et les groupes de nouvelles (<i>newsgroups</i>) | 11 |
| 4 Discipline personnelle pour une veille efficace | 12 |
| IV - Recommandations pour l'enseignement | 13 |
| V - Références générales..... | 14 |

Mise en œuvre d'une veille scientifique et technique dans le cadre d'enseignements à distance

La toile ou web contient environ 8 milliards de pages. Leur nombre croît sans cesse.

Des catégories nouvelles apparaissent comme les blogs personnels, puis les blogs politiques, enfin les blogs d'entreprises.

Les outils de fouilles sont très nombreux, environ 2 000 moteurs, agents, méta moteurs...

L'aspect raisonné de notre inventaire porte sur la typologie des outils. Leur liste exhaustive est hors d'atteinte dans le cadre de cette étude.

On trouvera ci-dessous les grandes catégories d'outils pour la recherche et la veille et quelques exemples pour chacun.

Le lecteur qui n'est pas du tout au fait de ces choses lira utilement le § V « Références générales » page 14.

La quatrième partie de ce texte contient nos recommandations pour l'introduction de ces matières dans l'enseignement, page 13.

I - Inventaire

1 Annuaires

Un annuaire (ou répertoire) recense des sites et fournit une fiche sur chacun : titre, adresse (URL), descriptif de 15 à 25 mots.

Chaque site est inscrit dans une ou plusieurs catégorie(s) ou rubrique(s).

Ces outils peuvent ainsi être considérés comme les pages jaunes de la toile. Lorsque l'on met en ligne un site que l'on désire voir apparaître dans un annuaire il faut donc s'y inscrire. Le travail d'indexation est manuel et fait par des spécialistes du domaine concerné, ce qui est un gage de qualité.

L'intégralité des documents n'est pas indexée et la mise à jour des index est lente.

Le nombre de sites pris en compte par les annuaires est petit mais ils sont parmi les plus pertinents.

Quelques annuaires

Dans la suite, les dates entre parenthèses sont celles de la dernière vérification du site.

Yahoo <http://www.yahoo.com/> (12/12/2005)

Yahoo http://fr.dir.yahoo.com/references_et_annuaires/portails_et_annuaires/ (12/12/2005)

Webbrain <http://www.webbrain.com> (12/12/2005)

Nomade <http://nomade.aliceadsl.fr/> (12/12/2005)

Dmoz <http://dmoz.org/> a une représentation graphique <http://maps.map.net/start> (12/12/2005)

Voilà <http://guide-voila.search.ke.voila.fr/> (12/12/2005)

Note : Il existe des annuaires d'annuaires :

www.lesannuaires.com/ (13/12/2005)

<http://annuairedesannuaires.maxibottin.com/> (13/12/2005)

2 Moteurs de recherche

Un moteur de recherche est constitué

- d'une base de données ;
- d'un outil de recherche dans la base ;
- d'un robot (spider ou crawler en anglo-saxon) qui parcourt la toile de lien en lien pour alimenter la base de données.

Les informations trouvées sont sauvegardées et indexées automatiquement au fur et à mesure. Les délais de passage du robot sont variables.

Mise en œuvre d'une veille scientifique et technique dans le cadre d'enseignements à distance

Les index contiennent des milliards de pages. Elles sont classées par ordre de pertinence, selon un ordre et un algorithme basé sur certains critères de tri, propre à chacun.

Le moteur de recherche fait donc les recherches sur des pages lisibles alors que les annuaires proposent des sites. On ne peut pas comparer les résultats fournis par les deux types d'outils.

Qu'est ce que les robots utilisent pour indexer une page ?

Il utilise ce qu'il a «vu» c'est-à-dire :

- le contenu des pages web, textes, légendes lisibles par le lecteur;
- des éléments d'indexation, attachés à des balises insérées dans des pages de type html, non visibles à l'écran.

le titre Ex. : <title>Plasturgie : le site de tous les plastiques</title>

le titre visible est «Plasturgie : le site de tous les plastiques» ;

une description ou **une phrase** Ex. : <meta name= « description » content= «le site qui traite de l'actualité des plastiques » >;

censée décrire le mieux possible les sujets traités par le site

des mots clefs caractérisant le mieux possible le site.

Ex. <meta name= « keywords » content= « plastiques, plastique, plasturgie, plastic »>

des noms d'images ou de documents multimédias.

Certains moteurs ou versions de moteurs n'utilisent que ces mots-clefs pour faire l'indexation.

Une variante de Google recherche les titres.

Certains moteurs sont spécialisés dans les sites personnels ou «blogs»

<http://www.mylinea.com/> (12/12/2005)

<http://www.google.fr/blogsearch?hl=fr> (16/12/2005)

<http://technorati.com/> (15/07/2006)

Le concepteur d'un site prend en compte ces éléments. Pour cela, on peut consulter la FAQ :

http://www.abondance.com/docs/faq_promo.html (15/12/2005)

Ce que le robot ne voit pas

Les pages dynamiques dans les formats asp, jsp, php, sont peu, voire jamais indexées car leur contenu change selon les réponses de la base de données à laquelle elles sont couplées.

Les parts de marché de moteurs et la liste des moteurs les plus utilisés est visible dans :

<http://www.barometre-referencement.com/> (14/12/2005)

3 Métamoteurs

Les métamoteurs relaient les interrogations vers des moteurs ou des annuaires.

Ils sont en ligne (accessibles par un navigateur) ou sous la forme d'un logiciel chargé sur l'ordinateur du client.

Ces outils donnent des réponses rapides, un panorama global des résultats. On commence une recherche avec un métamoteur.

Ils ont deux inconvénients :

Leurs réponses sont souvent les plus « rapides » mais non les plus pertinentes ;

Leurs réponses dépendent de la qualité des moteurs et annuaires qu'ils utilisent

Métamoteurs en ligne

Vivísimo <http://vivisimo.com/> (11/12/2005)

IxQuick <http://ixquick.com/> (11/12/2005)

Kartoo <http://www.kartoo.com/> outil graphique (Flash obligatoire) (13/12/2005)

Metacrawler <http://www.metacrawler.com/> recherche sur les textes ainsi que sur les noms des fichiers images, multimédias, audio et news. (14/12/2005)

EO http://eo.st/index_eo_net.htm (14/12/2005)

Jux2 <http://www.jux2.com> (15/12/2005) utilise Google, Yahoo et Ask Jeeves. Il compare séparément ce que chacun d'eux a trouvé et que les deux autres n'ont pas trouvé. Il est un bon outil d'enseignement pour contrebattre la tendance à n'utiliser que le moteur que l'on connaît bien.

Mise en œuvre d'une veille scientifique et technique dans le cadre d'enseignements à distance

Une étude de *Dogpile.com* au printemps 2005 a établi que les résultats des premières pages fournies par les 3 moteurs *Google*, *Yahoo! Search* et *Ask Jeeves* sont de plus en plus différents, 3% vus par les trois, 12 % par deux des trois, 85 % par un seul.

Un méta-moteur original <http://www.goshme.com> (16/12/2005) Sa particularité est de chercher au moyen d'un nombre considérable d'annuaires, de moteurs spécialisés ou généraux, le tout à la demande. Il annonce un millier de moteurs. Les résultats de chaque outil sont donnés séparément avec un commentaire sur les particularités de chacun. C'est un très bon outil pédagogique.

Logiciels métamoteurs

Copernic (Fr) <http://www.copernic.com/> a une version gratuite (14/12/2005)

WebFerret (An) <http://www.ferretsoft.com/download.htm> a une version gratuite (14/12/2005)

4 Les outils spécialisés

Les moteurs et annuaires spécialisés dans un domaine

Ils sont spécialisés dans un ou plusieurs domaines. Ils sont donc pertinents. Ils sont souvent associés à des portails., sauf pour Goshme méta-moteur cité plus haut.

Ces outils fournissent souvent une information pertinente et à jour. Ils peuvent être aussi spécialisés dans la recherche d'un média en particulier.

Deux outils atypiques

Décisionnel <http://www.decisionnel.net/internet/support.htm> (12/12/2005) recherche séparément adresse, norme, photo, salon, logiciel, brevet.

Pour ce qui relève de la recherche de renseignement économique <http://epices.univ-tln.fr/index.php> (12/12/2005)

Pour rechercher des sites spécialisés

Bonweb <http://www.bonweb.com/> (06/12/2005)

Pour rechercher des moteurs spécialisés

BigSearchEngine <http://www.search-engine-index.co.uk/> (14/12/2005)

Des automates de recherche

<http://www.dsi-info.ca/moteurs-de-recherche.html> (12/12/2005)

Une analyse des principaux moteurs de recherche

<http://www.allhtml.com/analysemoteurs/index.php> (12/12/2005)

5 La toile profonde ou invisible et les bases de données

5.1 La toile dite profonde

Les données ou les pages auxquelles on n'accède pas par les annuaires, les moteurs, etc. sont dites appartenir au "web invisible" ou "web profond".

Elles ne sont pas accessibles parce que :

- 1) trop récentes, le ou les moteurs ne les ont pas encore trouvées ;
- 2) aucun lien ne pointe vers la page ;
- 3) ce sont des pages dynamiques construites à la volée après fourniture d'une requête. C'est le cas des bases de données. Un moteur de recherche ne sait pas fournir une requête ;

Mise en œuvre d'une veille scientifique et technique dans le cadre d'enseignements à distance

- 4) l'accès est bloqué par la fourniture d'un identifiant et d'un mot de passe ;
- 5) ces pages sont dans un format inhabituel alors que les moteurs sont écrits au départ pour lire des pages html. Certains lisent du pdf, d'autres du word, aucun ne « lira » une image.
- 6) une marque a été apposée dans la page pour interdire sa lecture par un robot ;
- 7) la politique du moteur est de ne pas indexer telle page en raison par exemple de son url ;
- 8) elles sont dans une partie non explorée d'un site trop « gros » pour le moteur, ou d'une page trop grande ;
- 9) parce que l'accès à la page est conditionné par un choix dans une liste de mots ou par la frappe d'un mot, semblable aux pages asp, php, etc. voir http://www.abondance.com/docs/faq_promo.html;
- 10) parce que certains pays interdisent la consultation de sites (France, Chine).

Environ 95 % du web profond est accessible en ce qu'il est gratuit et n'est pas réservé à certains utilisateurs.

Pour en savoir plus sur le web invisible :

<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html> (14/12/2005)

<http://www.digimind.fr/> (03/02/2006)

Des sites donnent accès à des parties plus ou moins invisibles du web.

<http://dadi.enssib.fr/> (16/12/2005)

http://signets.bnf.fr/html/categories/c_025invisible.html (16/12/2005)

<http://infomine.ucr.edu/> (11/12/2005)

<http://www.rdn.ac.uk/> (11/12/2005)

Remarque. Un site référencé en aucun annuaire, non déclaré à des moteurs et vers lequel aucun lien ne pointe restera totalement invisible.

Les bases de données en ligne ne sont pas à proprement parler des composantes du web, certaines sont accessibles depuis longtemps par telnet (Dialog, Questel...). Le web ne sert que d'interface. Elles font partie du **web profond** parce que les moteurs ne les indexent pas. Elles constituent plus de la moitié de ce web profond.

Comment localiser une base de données ?

CompletePlanet (An) <http://www.completeplanet.com/> (14/12/2005) annonce 70 000 bases de données et engins spécialisés.

Pro Fusion (An) <http://www.profusion.com/> fait une recherche par thèmes (02/02/2006)

5.2 Typologie du contenu des sites du web profond

Bright Planet a étudié 17 000 sites du web profond et établi une typologie de leur contenu en 12 classes.

1. Les bases de données spécialisées : Ce sont des agrégateurs d'informations, interrogeables, spécialisés par sujet. Exemples : les bases de données médicales, de chimie, de brevets.

National Library of Medicine interrogeable via *PubMed*, *Esp@ceNet*, base de données européenne de brevets, *GlobalSpec*, base dédiée à l'information technique et d'ingénierie.

2. Les bases de données internes à des sites web de grande taille. Exemple : la base de connaissance des sites *Microsoft*.

3. Les bases de publications, interrogeables via un moteur interne donnant accès aux articles, à des extraits de livres, des thèses. Exemple : *FindArticle*.

4. Les sites de vente en ligne ou d'enchères. Exemples : eBay, Fnac, Amazon.

5. Les sites de petites annonces. Exemples : *particulier à particulier*, *VerticalNet* place de marché industrielle.

6. Les portails sectoriels. Ils agrègent plusieurs types d'information : articles, publications, liens, forums, annonces interrogeables via un moteur et une base de données.

Exemple : *PlasticWay*, portail de la plasturgie.

7. Les bibliothèques en ligne. Elles sont universitaires ou nationales. Exemples : *Bibliothèque du Congrès*, *BNF-Gallica*, bases recensées par l'*OCLC*

8. Les pages jaunes et blanches, les répertoires de personnes morales et physiques. Exemples : *Pages Jaunes*, *ZoomInfo* répertoire des dirigeants et salariés d'entreprise.

Mise en œuvre d'une veille scientifique et technique dans le cadre d'enseignements à distance

9. Les calculateurs, simulateurs, traducteurs. Ils contiennent des tables de données pour calculer et afficher leurs résultats. Exemples : Traducteur *Systran*, *Babelfish*, *le Grand Dictionnaire Terminologique*, *le Trésor de la langue française*.

10. Bases de données d'emploi et de CV. Exemples : *Monster*, *APEC*, *Cadresonline*...

11. Site de forums ou de chat. Exemples : journaux quotidiens, *Le Figaro*, *Dernières nouvelles d'Alsace*, etc.

12. Bases de données de recherche généralistes. A la différence des bases de données spécialisées, elles recensent des thèmes éclectiques. Exemple : *Weborama*.

5.3 Utilité et pertinence du web profond

Cette partie du web étant constituée de ressources écrites le plus souvent par des spécialistes est de bien meilleure qualité que le web visible.

5.4 Quelques outils

<http://www.goshme.com> déjà cité.

<http://www.completeplanet.com/> déjà cité

<http://oaister.umd.umich.edu/o/oaister/> de l'université du Michigan, répertoire d'articles et de documents. (03/02/2006)

<http://www.archive.org/> archive des sites disparus depuis 1996 (03/02/2006)

5.5 Les bibliothèques

800 bases de données par l'URFIST DE Lyon : <http://dadi.enssib.fr/> (03/02/2006)

La Bibliothèque Nationale de France (BNF) <http://www.bnf.fr/pages/catalogues.htm> (02/02/2006)

<http://www.libdex.com> répertoire d'environ 18 000 bibliothèques. (03/02/2006)

<http://www.objectifgrandesecoles.com/pro/secteurs/index.htm> portail vers 1000 portails d'affaires sectoriels.

**L'hétérogénéité du web profond et l'absence d'outils généraliste
rendent son exploitation difficile et longue.**

6 Les blogs

Le nombre de blogs ou sites personnels d'humeur croît très vite.

Certains estiment qu'il y en avait 55 millions fin 2005.

Des entreprises en ont créé à la manière des tableaux d'affichage libres.

Des hommes politiques en font un outil de diffusion de leurs idées.

Leur foisonnement les rend difficilement accessibles et leur intérêt est aléatoire.

Les moteurs classiques commencent à les parcourir.

<http://www.terrar.com/> (14/12/2005)

<http://www.davidjanes.com/blogosphere/> (14/12/2005)

<http://www.technorati.com> (15/07/2006)

<http://www.blogpulse.com/> (14/12/2005)

http://c.asselin.free.fr/french/weblogs_moteurs.htm (15/07/2006) entretient une liste des moteurs de recherche dans des blogs.

7 Les travaux de groupes

Les outils liés aux travaux de groupes (groupware)

Ils apparaissent sous les formes de forums, de listes de diffusion, de listes de discussion et des groupes de nouvelles (newsgroups). L'information qu'on y trouve est toujours très récente et parfois pertinente. Les moteurs de recherche indexent parfois des informations en provenance des forums des sites qu'ils explorent.

Mise en œuvre d'une veille scientifique et technique dans le cadre d'enseignements à distance

Un outil

Francopholistes <http://www.francopholistes.com/> (14/12/2005) est le grand annuaire des listes en français., plus de 6000.

8 Les agents dits intelligents (cf. introduction)

Un agent est idéalement

- autonome : son fonctionnement est automatique.
- communicant : il échange des données avec d'autres programmes ou des hommes.
- perfectible : il est capable de réagir, de s'adapter aux circonstances, de prendre une décision, d'améliorer son comportement, sur la base de ses observations

On les utilise pour les veilles.

Ils diminuent la masse d'information.

Ils sont voulus exhaustifs tout en restant pertinents.

Ils économisent du temps pour le balayage des serveurs.

Ils fabriquent des bases de données thématiques consultables hors ligne.

Quelques agents

WebSeeker, fait des recherches sur les moteurs francophones, payant, version d'essai temporaire gratuite <http://www.bluesquirrel.com/products/seeker/> (12/12/2005)

Web Ferret à essayer, a une version gratuite <http://www.ferretsoft.com/> (12/12/2005)

Copernic propose un méta moteur et agent de recherche. Plus de 30 sources d'information (module de base) regroupées en trois domaines: le web, groupes de discussion et adresses de courrier. <http://www.copernic.com/fr/> (14/12/2005)

Strategic Finder La version gratuite interroge plus de 200 moteurs de recherches sur Internet. La version payante ajoute de nombreuses sources professionnelles. <http://www.strategicfinder.com/> (12/12/2005)

Le blog *Outils Froids* est un site utile <http://www.outilsfroids.net/texts/InfosOutilsFroids.shtml> (12/12/2005)

Deux adresses de recensement d'agents

BotSpot <http://www.botspot.com/> (14/12/2005)

AgentLand <http://www.agentland.fr/> (14/12/2005)

9 Trouver le bon outil et rester informé de l'évolution des autres

La recherche du bon outil sera aidée par :

<http://www.enfin.com/> (12/12/2005) est une base de données internationale d'annuaires et de moteurs

<http://www.indicateur.com/> (12/12/2005)

<http://www.searchenginecolossus.com/> (12/12/2005) répertoire international des outils de recherche.

II - Interpréter les résultats

L'information trouvée n'est pas forcément pertinente. De plus, il faut tenir compte de l'arrivée de l'information payante. Exemple : *Northern Light* <http://www.northernlight.com/> (12/12/2005) indexe environ 350 millions de pages. En plus, il a une sélection payante de 50 millions de pages spécialisées, composée d'articles de journaux scientifiques et financiers.

1 Validité de l'information sur Internet

Données générales

La qualité des documents est très variable.

Tout le monde peut écrire n'importe quoi ...

Les auteurs ou émetteurs sont parfois difficiles à identifier.

L'objectif réel des sites peut rester caché, leur information n'est-elle pas biaisée ?

Les enjeux économiques et politiques influencent les contenus, les idéologies les modèlent.

Démarche à suivre

Il faut rester soupçonneux et critique. Le recoupement des sources est nécessaire mais il faut prendre garde aussi à la même information plusieurs fois reprise par des « copier-coller ». On devra procéder par évaluations successives.

Attention à la signification des mots : la qualité de l'information n'est pas son exactitude ni sa complétude.

2 Les critères d'évaluation

La crédibilité du site

On distinguera d'abord les trois grands types de sites :

- site d'information ou commercial ;
- site de communauté virtuelle entre amateurs ou professionnels ;
- site très personnel dit « blog ».

La crédibilité pourra être appréciée par :

- la conformité du site aux règles usuelles : identification du webmestre, nom et adresse du responsable de la publication, si l'on peut s'inscrire, quels engagements sont pris en matière de confidentialité, de non divulgation d'adresse, de possibilités de radiation ?
- l'environnement publicitaire ;
- la notoriété du site, le fait qu'il ait ou non un comité de lecture, le fait qu'un forum est modéré ou pas ;
- combien de liens aboutissent :
 - à un site à partir de sites et annuaires, francophones ou autres : <http://www.soumission-manuelle.com/> (12/12/2005)
 - à une page d'un site <http://www.linkpopularity.com/> (12/12/2005) pour les seuls moteurs *Altavista*, *Google* et *MSN*.
- La qualité graphique du site et la navigation dans le site
 - Le texte est-il aisément lisible ?
 - La navigation est-elle aisée ?
 - Le chargement est-il rapide ?
 - Des modules additionnels (« plug-ins ») sont-ils nécessaires ?

Dans un site sans plan de navigation, un outil comme *SiteMapper* payant peut s'avérer utile : <http://www.trellian.com/mapper/> (12/12/2005)

L'appréciation du contenu

Le contenu proprement dit

- Le contenu du site montre-t-il un thème traité avec une profondeur suffisante?

Mise en œuvre d'une veille scientifique et technique dans le cadre d'enseignements à distance

- Le niveau est-il approprié à votre travail ? L'expression est-elle trop compliquée, trop spécialisée ou trop simple au contraire?
- Est-ce que le site articule de l'information de type "recherche", "éducation", "savante ou scientifique" avec un produit commercial ou un service de promotion -vente (publicité)?
- L'information n'est-elle pas, périmée, obsolète?

D'autres critères :

- la qualité de la langue ;
- l'identification du ou des auteurs à l'intérieur du source de la page. Vérifier pour cela la balise auteur. (Ex. meta name= « author » content= « jean Dupont » indique que Jean Dupont est l'auteur des pages) ;
- la distinction entre contenu et publicité est-elle nette ?
- les mises à jour sont-elles fréquentes ? On peut obtenir, mais pas toujours, la date de dernière mise à jour d'une page. On utilise des outils qui s'intègrent aux navigateurs pour aider cette vérification : exemple les bookmarklets gratuits <http://www.bookmarklets.com/> (12/12/2005) ;
- combien de liens partent de cette page <http://www.bookmarklets.com/> (12/12/2005) .

On opérera ensuite sensiblement comme pour un document écrit.

- les sources originales sont-elles citées ?
- le site est-il mis à jour ?
- la page datée et signée ?
- vérifier aussi ces éléments dans la bibliographie citée. Y-a-t-il suffisamment de références à d'autres travaux ?
- s'agit-il d'une source stable, produite et mise à jour par un organisme reconnu, une université, une bibliothèque ?
- les auteurs du site sont-ils des chercheurs ou des spécialistes dans ce domaine? Appartiennent-ils à des sociétés savantes, des institutions, des organismes reconnus? Peut-on vérifier leurs qualités et identifier leur établissement de rattachement ?
- l'information est-elle structurée ?

3 Les outils d'analyse et de synthèse textuelle ou cartographique

En synthèse textuelle, l'exemple du site *Tropes* est significatif. Il propose ou s'engage à :

- accélérer énormément votre vitesse de lecture,
- analyser en profondeur et de façon objective,
- extraire l'information pertinente,
- classer automatiquement, donc structurer rapidement l'information.

Tout cela au moyen de quatre logiciels *Tropes* <http://www.acetic.fr/> (12/12/2005)

En cartographie, le site de *Kartoo* <http://www.kartoo.com/index.php3?langue=fr> (12/12/2005) est un bon exemple. A chaque résultats trouvé, est associé un jeu de domaines voisins issus de l'analyse de deux textes et s'il y a lieu les sites trouvés correspondants

Le site <http://www.wwf.be/> déclaré pertinent a pour domaines connexes : « plastiques » lui-même, « polymères » et « chimique » pour <http://www.metaux.com/>, « vinyl » sans site etc.

Les résultats sont affichés au moyen de flashplayer et donc non directement copiables ici.

Agents de recherche avancée

Leur fonction est de trouver "toutes" les pages contenant la requête, l'enrichir, l'analyser et filtrer les pages trouvées pour rapatrier uniquement les bonnes pages.

La bibliométrie

Cette technique d'analyse statistique porte sur les noms, les références et est extensible aux mots. Elle sort du champ de notre document.

Une bonne référence se trouve à :

<http://www.pasteur.fr/infosci/biblio/formation/metrie/> (12/12/2005)

Mise en œuvre d'une veille scientifique et technique dans le cadre d'enseignements à distance

Quelques agents

DigOut4U est un agent spécialisé dans l'analyse textuelle

<http://planete.qc.ca/logiciels/description.asp?cle=712> (12/12/2005)

Umap examine les convergences de liens et détecte les sites fédérateurs

<http://www.trivium.fr/fr/index.htm> (12/12/2005)

Ressources

On trouvera un panorama des méthodes et outils d'exploitation dans *"outils de veille, typologie "* de Claire François

<http://www.inist.fr/rencontresIST/docpdf/4veille.PDF> (12/12/2005).

Un domaine nouveau est celui de la gestion des connaissances (knowledge management).

Une étude détaillée des types d'outils spécialisés se trouve dans *"Analyse comparée de méthodes de Gestion des connaissances pour une approche managériale"*, de Françoise Barthelme-Trapp et Béatrice Vincent :

<http://www.strategie-aims.com/quebec/web/actes/f-113-cd.pdf> (12/12/2005).

III - Pérenniser la recherche par la veille

La veille est une « Activité continue et en grande partie itérative visant à une surveillance active de l'environnement technologique, commercial, etc., pour en anticiper les évolutions ».

Norme expérimentale XP X 50-053, « Prestations de veille et prestations de mise en place d'un système de veille », Afnor, 1998

Des informations deviennent caduques, d'autres apparaissent. Il faut pouvoir apprécier les évolutions qui nous intéressent. C'est la fonction de veille.

La veille ne diffère de la recherche que par la question : « quoi de nouveau ? ».

Par ailleurs on peut vouloir, non seulement savoir ce qui a été écrit mais tenter de connaître l'avenir, même imparfaitement.

Cette veille peut être menée à la demande de l'employeur ou comme fournisseur de service. Mais chacun devrait en mener une sur son ou ses domaines d'excellence.

Où peut-on trouver « du nouveau » ?

- dans des pages d'un site encore inconnu;
- dans des pages modifiées ou ajoutées à un site connu.

1 Pour les pages nouvelles dans un site encore inconnu

Le plus simple et le plus efficace est de les chercher par un processus classique de recherche. On vérifiera tout de même leur date de dernière mise à jour, <http://www.bookmarklets.com/> (12/12/2005).

La recherche peut être aidée par des outils spécialisés dans la veille.

Exemple : l'outil en ligne <http://www.googlealert.com/> (12/12/2005) gratuit.

2 Pour les pages modifiées ou ajoutées dans un site déjà connu

On met en place une surveillance (monitoring) en vue de surveiller les modifications d'un site ou d'une page pour suivre la concurrence et automatiser la recherche d'information.

On peut utiliser les outils en ligne gratuits que sont les traqueurs (trackers). On donne au traqueur une ou plusieurs adresses à surveiller. Il ira, à intervalles définis, détecter les changements intervenus sur la page ou le site. Les paramètres peuvent être définis, par exemple un prix pour surveiller un concurrent.

L'utilisation de RSS (sommaire de site enrichi (rich site summary) ou une syndication vraiment simple (really simple syndication) date de décembre 1997. Elle fournit ses résultats dans le format xml. Très brièvement, le RSS est un fichier descripteur des contenus et des modifications apportées à un site. Sa consultation au moyen d'un outil lecteur RSS, dit aussi agrégateur de nouvelles, évite l'examen détaillé du site d'origine quand on y recherche les modifications et les nouveautés.

Il en existe malheureusement des versions nombreuses qui nuisent à sa bonne utilisation.

http://www.pointblog.com/abc/rss_et_syndication_1.htm (12/12/2005) contient une description détaillée du RSS, des liens vers les spécifications et vers les outils de RSS.

News Interceptor <http://www.newsinterceptor.com/> (15/10/2006) paraît être le logiciel le plus général.

3 Les listes de diffusion et les groupes de nouvelles (newsgroups)

Les listes de diffusion ont déjà été citées comme outils de travail collectif. Elles sont particulièrement utiles pour la veille, <http://www.francopholistes.com/> (12/12/2005).

Leur intérêt majeur est que l'on participe à une veille collective et non plus solitaire. L'efficacité en est largement augmentée.

Les groupes de nouvelles sont accessibles à partir des navigateurs (dans Netscape : « fenêtre », puis « messagerie et forums »).

Leur intérêt est très variable mais il faut les essayer.

4 Discipline personnelle pour une veille efficace

La veille ne se fait pas en attendant de trouver la pépite qui contiendra les réponses complètes et définitives. Il est indispensable de rédiger une fiche pour chaque résultat. Cette fiche sera ensuite intégrée dans une base de données (voir annexe).

Des outils

Un outil logiciel *CopernicAgentProfessionnal* <http://www.copernic.com/> (12/12/2005) payant.

La plate-forme documentaire « Pertinence Information Network » apporte une veille automatisée (essai gratuit) <http://www.pertinence.net/pin/index.jsp> (12/12/2005), sur abonnement.

Les outils de *Tracerlock* <http://www.tracerlock.com/> (12/12/2005) payant.

KBCrawler payantx <http://www.beaconseil.com/products/KBCrawl/contenu.php?Id=1> (11/12/2005).

Orbiscope <http://www.orbiscope.net/fr/> (12/12/2005) payant.

Note :

L'activité de veille, au contraire de la recherche documentaire, est parfois sous-traitée par les entreprises. À cause de cela, elle a été segmentée dans un but commercial.

On distingue dans la littérature :

La veille stratégique porte sur des points essentiels pour la survie de l'entreprise.

La veille concurrentielle ou commerciale porte son nom.

La veille technique ou technologique porte son nom.

Etc.

Forums sur la veille, les agents intelligents et la prospective.

<http://www.veille.com/listes/maillinglists.html> - veille (13/12/2005)

Liste des formations en veille et intelligence économique

<http://formation.veille.co.uk/> (13/12/2005)

IV - Recommandations pour l'enseignement

L'apprentissage préalable d'une méthodologie générale de recherche et de veille nous paraît hors de propos. Cette façon de procéder relève d'enseignements spécialisés à l'usage des bibliothécaires, documentalistes et professionnels du domaine.

Notre recommandation est de procéder en trois temps en faisant travailler en groupes de 3 à 5 élèves.

1^{er} temps Prise de contact avec la recherche sur la toile. (environ 1 mois)

Introduction aux mots clefs et à leurs connecteurs logiques : ET, OU, SAUF.

Définition d'un thème de recherche.

Utilisation d'un seul moteur, google par exemple avec plusieurs jeux de mots-clefs pour le domaine ci-dessus.

2^e temps Approfondissement des techniques de recherche et d'exploitation. (environ 1 mois)

Utilisation de deux métamoteurs en ligne capable de comparaisons d'efficacité : jux2 et goshme toujours avec les mêmes jeux de mots-clefs.

Utilisation d'un logiciel gratuit installable : copernic basic.

Affinage des jeux de mots-clefs.

Rédaction de deux synthèses :

l'une sur les aspects méthodologiques, ;

l'autre sur les résultats obtenus à partir de fiches individuelles de résultats.

2^e temps Apprentissage et installation d'outils de veille automatisée : (environ 1 mois)

Outil en ligne : Google alert

Outil logiciel local Copernic basic en utilisant les dates d'obtention des références de résultats

Fils RSS à installer

Abonnements aux listes de diffusion

Suivi de forums.

3^e temps Exploitation et synthèse (environ 3 mois)

Exploitation des résultats obtenus au fur et à mesure de leur arrivée.

Rédaction d'une synthèse sur le sujet.

Cette activité de recherche et de veille doit être menée dans la durée, cinq à six mois sont réalistes et non un semestre académique.

Elle ne doit en aucun cas consommer plus d'une heure par semaine en période chargée.

Cette formation a pour finalité de mettre les élèves en état de chercher et veiller pour leur propre compte et pour celui de leur entreprise.

Le passage à l'activité de production nécessite l'acquisition d'outils de veille et de gestion des connaissances.

V - Références générales

Ressources générales sur la recherche et sur la veille

Norme AFNOR XP X50-053 Avril 1998 Prestations de veille - Prestations de veille et prestations de mise en place d'un système de veille, disponible en ligne à

<http://www.boutique.afnor.fr/Boutique.asp?url=NRM%5Fn%5Fhome%2Easp&lang=French&btq=HOM&cookie%5Ftest=1>

Un site général sur la recherche, la veille, la gestion des connaissances <http://www.decisionnel.net/index2.html> (12/12/2005). Il n'est plus mis à jour mais reste pertinent sauf pour les blogs.

On trouvera un exposé des techniques employées dans les moteurs de recherche dans :

<http://cedric.bouttes.free.fr/RapportProbatoire.htm> (16/12/2005)

Un site de qualité <http://www.zeknowledge.com/> (13/12/2005)

Sur la veille

Sur les agents intelligents, la veille, l'actualité du net, l'actualité des outils de recherche

<http://www.asktibbs.com/> (13/12/2005).

Outils de veille et d'alerte accessibles gratuitement sur internet et conseils méthodologiques :

<http://www.msh-alpes.prd.fr/veille/> (13/12/2005).

http://www.veille.com/annuaire_veille/03Outils/ (13/12/2005).

Outils de recherche dans le cadre de la veille sur internet.

<http://www.decisionnel.net/recherche/outils.htm> (13/12/2005)

Annuaire des agents d'alerte intelligents.

http://www.agentland.fr/Download/5Monitoring_agents.html (13/12/2005)