

Méthodes de recherche sur Internet

Septembre 2004

© GLS Réseaux



Table des matières

OBJECTIF DE L'ATELIER.....	3
À LA FIN DE L'ATELIER LES PARTICIPANTS :	3
UTILISER EFFICACEMENT INTERNET À DES FINS PROFESSIONNELLES ET ACADÉMIQUES.....	3
COMPÉTENCES NÉCESSAIRES À LA RECHERCHE ET L'ANALYSE D'INFORMATION.....	4
ÉTAPES DANS UN PROCESSUS FORMEL DE RECHERCHE ET DE VEILLE	5
MÉTHODES DE COLLECTE D'INFORMATION	6
TYPES D'INFORMATION.....	7
TYPES DE SOURCES	7
TYPES DE RECHERCHES	8
SOURCES D'INFORMATION ET TYPES DE RECHERCHE.....	8
OUTILS ET SOURCES D'INFORMATION SUR INTERNET	9
ANNUAIRES DE RECHERCHE	10
MOTEURS DE RECHERCHE	10
RÉFÉRENCEMENT	10
CONSEILS POUR UNE UTILISATION MAXIMALE DES MOTEURS DE RECHERCHE.....	11
POPULARITÉ DES ENGINES DE RECHERCHE	11
SITES ET PORTAILS SPÉCIALISÉS.....	12
BASE DE DONNÉES ET WEB INVISIBLE	12
OUTILS DE RECHERCHE DANS LES BASES DE DONNÉES	13
LISTES ET FORUMS DE DISCUSSION	13
COMMUNIQUÉS DE PRESSE ET NOUVELLES EN CONTINU.....	14
CARNET - BLOG (WEBLOG)	15
LOGICIELS DE RECHERCHE ET AGENTS DE VEILLE.....	16
AGENTS DE RECHERCHE	16
AGENTS D'ALERTE.....	16
OUTILS DE CARTOGRAPHIE	17
ASPIRATEURS.....	17
AGRÉGATEURS DE FILS RSS	17
AUTRES OUTILS ET SOURCES D'INTÉRÊT	18



Objectif de l'atelier

À la fin de l'atelier les participants :

- Seront sensibilisés au contexte d'Internet comme sources d'information;
- Connaîtront les principales méthodes de recherche d'information;
- Seront aptes à utiliser efficacement les options de recherches avancées des moteurs de recherche;
- Seront avisés des différentes sources d'information sur Internet;
- Pourront utiliser une variété d'outils permettant d'optimiser la recherche, l'indexation et le partage d'information

Utiliser efficacement Internet à des fins professionnelles et académiques

Une utilisation optimale d'Internet permettra à différents types d'organisation d'intervenir avec efficacité sur une variété d'aspects :

Type d'organisation	Thèmes de veille
Entreprise	État et suivi de la compétition, nouveaux concurrents, parts de marchés, tendances de consommation, canaux de distribution, client potentiels, identification d'opportunités, fournisseurs, prix de produits/ services, publicité, appels d'offres, etc.
Gouvernement	Politiques publiques, stratégies, programmes, lois et réglementation, Règlement et normes, etc.
Laboratoire, département de R&D	Nouvelles technologies, nouveaux matériaux, percées scientifiques, nouveaux produits, brevets, propriété intellectuelle, déclaration d'invention, etc.



Compétences nécessaires à la recherche et l'analyse d'information

Patient chercher l'information désirée peut prendre plus de temps que l'on pense.

Curieux l'information de se trouve pas que sur Internet.

Intelligent la capacité à créer des renseignements.

Créatif la capacité à penser hors des schèmes communs.

Coopératif partager l'information avec votre réseau.

Avisé agir au bon moment auprès de la bonne personne.

Et surtout, la maîtrise des langues...

***Tout ce qui se conçoit bien s'énonce clairement,
Et les mots pour le dire viennent aisément.***

Nicolas Boileau

Pour être un chercheur efficace, un veilleur aguerri ou tout simplement un bon utilisateur d'Internet à des fins professionnelles, il est important de faire une veille sur la veille. Les veilleurs doivent être au fait et à jour sur :

- les nouveaux outils et produits de veille;
- l'évolution et les fonctionnalités des outils existants;
- les trucs et astuces de recherche;
- l'application des techniques d'analyse;

Pour ce faire, nous recommandons aux individus de s'abonner à différentes sources d'information sur la veille, les moteurs de recherche et les nouveautés Internet. Ce travail peut être effectué par tous et chacun ou une personne peut être en charge de cette tâche et en aviser ses collègues en temps et lieux.

Étapes dans un processus formel de recherche et de veille

*J'écoute et j'oublie;
je lis et je me souviens ;
je fais et je comprends.*

PROCESSUS FORMEL DE RECHERCHE	
Étape	Description
1. Penser	Définir les objectifs, les besoins en information, identifier et localiser les sources d'information formelles et informelles (publications, articles, bases de données)
2. Collecter	Identifier et collecter toutes les sources d'information dans les secteurs d'emploi identifiés
3. Classer	Classer et inscrire les sources d'information dans les agents d'alerte ou les agrégateurs de nouvelles et organiser la collecte des sources <i>push</i>
4. Extraire	Extraire le contenu provenant de la mise à jour des sources d'information
5. Traiter	Mettre en forme et homogénéiser les informations (inscrire dans les bases de données, hiérarchiser et organiser l'information)
6. Éditer	Vérifier, consolider et fiabiliser les informations
7. Diffuser	Diffuser et partager les informations auprès des utilisateurs
8. Valider	Consulter et valider les statistiques de consultations et ajuster la stratégie
9. Maintenir	Renouveler et ajouter de nouvelles sources d'informations



Méthodes de collecte d'information

Recherche formelle (forward chaining)

Ce type de recherche se présente lorsqu'à partir d'une référence quelconque, une recherche est amorcée.

Ex. Un communiqué de presse diffuse les intentions du gouvernement du Québec de mettre en place de nouvelles structures de gouvernance en région. Pour obtenir davantage d'information sur ce plan de gouvernance, il suffit de naviguer jusqu'au site officiel du gouvernement du Québec et d'identifier la bonne section pour consulter le rapport. Dans certains communiqués annonçant la parution de rapports, l'adresse de référence est souvent fournie.

Recherche intuitive (Backward chaining)

La recherche intuitive consiste à naviguer à partir d'un site reconnu et pertinent, et ensuite à suivre les liens et ressources proposées par ce site jusqu'à épuisement des sources d'information.

Cette méthode de recherche est la plus pertinente pour structurer des cellules de veille, pour mettre en place un système de veille à long terme ou pour élaborer un dossier documentaire. Elle évite les pertes de temps dans les annuaires, BD et les moteurs de recherche.

Ex. En plaçant une requête (Gouvernance) dans un moteur de recherche on obtient une liste de résultats. En consultant ces résultats, on peut notamment trouver des sites de grandes organisations internationales (OCDE) et généralement, ces sites fourniront une liste d'organisations ayant des activités similaires.

Types d'information

Les types d'information à colliger dans une démarche de recherche dépendent des objectifs et de la méthodologie privilégiée. On peut distinguer trois types d'information :

Information blanche : information directement et librement accessible.

Information grise : information que l'on peut acquérir de manière indirecte ou détournée, mais d'accès légal.

Information noire : information obtenue de manière illégale par l'espionnage industriel.

La collecte d'information peut être légale ou illégale.

➤ **Légales (90%)**

Informations ouvertes (70%)
Sources secondaires

Informations fermées (20%)
Sources primaires

➤ **Illégales (10%)**

Espionnage industriel

Types de sources

PRINCIPALES SOURCES D'INFORMATION	
Institutions	Administrations, organisations internationales, associations, centres de recherche, instituts, fondations, universités,...
Sources fortuites	Train, avion, la rue, stagiaires, ...
Groupes partenaires	Sous-traitants, fournisseurs, clients, concurrents, groupements professionnels, financiers,...
Nouveaux réseaux	Moteur de recherche, listes de discussion, BD, carnets...
Réseaux personnels	Clubs, collègues, confrères, ...
Fournisseurs d'information	Presse, librairies, bibliothèques, serveurs et courtiers, banques de données, centres d'information, ...
Manifestations	Colloques, salons, voyages d'études, ...
Consultants et experts	Écoles d'ingénieurs, universitaires, experts, centres techniques, consultants, bureaux d'études, ...
Internes	... encore fallait-il y penser !

Types de recherches

Veille technologique

Recherche fondamentale et appliquée, usines, processus, brevets, normes

Veille commerciale

Fournisseurs et sous-traitants, clients, acheteurs et distributeurs, y compris marché du travail

Veille concurrentielle

Concurrents, prix, produits substituts

Veille environnementale et sociétale

Réglementaire, financière, fiscale, économique, politique, sociale, ressources humaines

Sources d'information et types de recherche

Quelles sources pour quelle veille ?

	Sources formelles										Sources informelles
	Revue spécialisée	Presse généraliste	Sites Internet	Brevets, marques	Normes lois	Études publiques	Infos légales	Rapports annuels	Livres	Autres médias	
Veille concurrentielle	X	X	X	XX		X	XX				X
Veille de marché, marketing	X	X	X			XX	X	X		X	X
Veille technologique	XX		X	XX	X	X			X		X
Veille réglementaire	X	X	X		XX	X					X
Veille environnementale	X	X	X		X	X					XX
Sujets généraux	X	XX	XX						X	X	X

On rappelle que les sources informelles comprennent : les concurrents, les fournisseurs, les clients, les banques, les experts, les associations, les événements, les formations, etc. Les sources informelles sont utiles pour toutes les veilles, et sont intéressantes mais nécessitent plus de validation.

Outils et sources d'information sur Internet

- Annuaires
- Moteurs de recherche
- Sites et portails spécialisés
- Bases de données et web invisible
- Listes et forums de discussion
- Communiqués de presse et nouvelles en continu
- Carnets (Blogs)

➤ **Annuaires et moteurs de recherche**

Une distinction importante est celle entre annuaires et moteurs de recherche :

ANNUAIRES	MOTEURS
Indexation de sites - par des documentalistes	Indexation de mots - par des robots
<i>Recherche</i> - sur des sites - et des catégories	<i>Recherche</i> - en texte intégral - sur des pages web
<i>Avantages</i> - choix des informations - classement raisonné par catégories et sous-catégories	<i>Avantages</i> - plus d'exhaustivité, - mise à jour plus rapide
<i>Inconvénients</i> - moins d'exhaustivité, - mise à jour moins rapide	<i>Inconvénient</i> capture de pages web sans classement raisonné
<i>A retenir</i> L'exploration des catégories s'avère souvent plus fructueuse que celle des sites.	<i>A retenir</i> La recherche par mots clés donne de meilleurs résultats sur les moteurs

Source : EducNet

Annuaire de recherche

Un annuaire (ex. Yahoo!) a pour mission de lister des sites web et de les placer dans des rubriques, un peu comme un annuaire de type pages jaunes. On ne trouve donc que des informations très générales sur le thème principal d'un site. Lorsque l'on effectue une recherche par mot-clé, c'est uniquement dans le titre du site et dans sa description (qui varie entre une et trois lignes) que la recherche s'effectue.

ANNUAIRES DE RECHERCHE	
Yahoo	http://www.yahoo.com
Open Directory Project	http://www.dmoz.org
Google Directory	http://directory.google.com

Moteurs de recherche

Un moteur de recherche (ex. Google) inspecte le contenu complet du site (en théorie) et localise tous les mots dans le contenu effectif du site pour ensuite renvoyer une liste complète de tous les documents publiés sur Internet contenant les mots-clés de la recherche.

MOTEURS DE RECHERCHE	
Google	http://www.google.com
Fast	http://www.alltheweb.com
Teoma	http://www.teoma.com

Référencement

Le référencement est un ensemble de moyens qui permet de donner une meilleure visibilité à un site web. Précisément, le référencement permet de faire afficher un site dans les premiers résultats des moteurs et annuaires. Le référencement est souvent biaisé et c'est pourquoi il est parfois si difficile de trouver ce qu'on cherche sur Internet.

Il existe différentes méthodes de référencement :

- Balises (*Meta-tags*);
- Liens (*Google Bombing*);
- Production de contenu;
- Nombre de clics;
- Référencement payant;

Deux outils pour vérifier la popularité de vos liens :

Market Leap

<http://tools.marketleap.com/publinkpop/>

Outiref

<http://www.abondance.com/audit/index.html>

Conseils pour une utilisation maximale des moteurs de recherche

Types de requêtes

Les moteurs ont tous leurs particularités propres. Ils ont tous leur façon de comprendre et d'interpréter les mots avec lesquels les requêtes sont effectuées. Pour plus d'information sur les syntaxes de saisie de ces différents moteurs : <http://www.abondance.com/outils/moteurs.html>

- Requête en langage naturel;
- Requête simple à un mot ou à plusieurs mots;
- Requête complexe à plusieurs mots;
- Utilisation des opérateurs booléens;
- AND, ALL, OR, ANY, NEAR, NOT, “ ”, *, [,], +, -

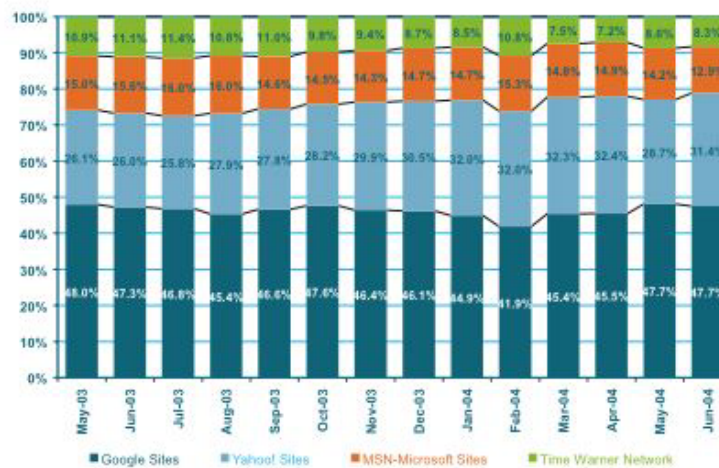
Fonctions de recherche avancée

Certains moteurs offrent des fonctions de recherche avancée permettant plusieurs possibilités. Il est possible de rechercher seulement dans le nom de domaine, le titre ou le texte. On peut même vérifier par qui un site est référencé. Les fonctions de recherche avancée sont les plus importantes à retenir et ce sont elles qui orienteront le plus efficacement les recherches sur Internet.

FONCTIONS DE RECHERCHE AVANCÉE	
Fast	http://www.alltheweb.com/advanced
Google	http://www.google.com/advanced_search
Search Engine Watch	http://www.searchenginewatch.com/facts/assistance.html

Popularité des engins de recherche

Worldwide Search Big 4 Monthly Search Share



Source: Majestic Research and comScore Networks



Sites et portails spécialisés

Il y a sur Internet des millions de sites personnels, corporatifs, récréatifs, etc. Un site portail est une "porte d'entrée" sur l'Internet et un « agrégateur de contenu ». Il regroupe à lui seul tout un ensemble de sites et de services dans un ou des domaines spécifiques.

SITES ET PORTAILS SPÉCIALISÉS	
Commission européenne	http://www.europa.eu.int
Service communautaire d'information sur la recherche et le développement (CORDIS)	http://www.cordis.lu/fr/home.html
Gouvernements dans le monde	http://www.gksoft.com/govt/en
Strategis	http://strategis.ic.gc.ca/frndoc/main.html

Base de données et web invisible

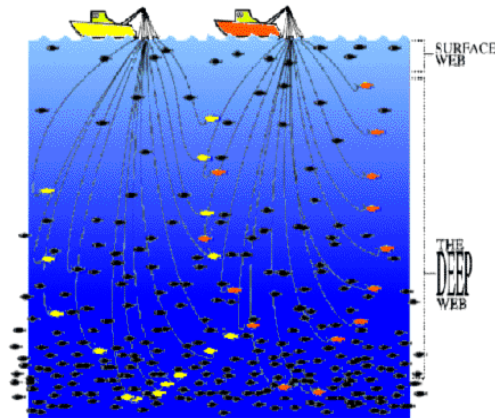
Une base de données est un ensemble structuré de fichiers reliés entre eux par des liens dynamiques pour rassembler des données destinées à une application précise. En recherchant dans des BD, on parcourt le Web invisible.

BASES DE DONNÉES ET WEB INVISIBLE	
Icriq	http://www.icriq.com
PubMed	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi
Hoover's	http://www.hoovers.com
Proquest	http://www.proquest.com

Outils de recherche dans les bases de données

Certains outils de recherche en ligne permettent une recherche simultanée dans plusieurs bases de données.

OUTILS DE RECHERCHE DANS LES BASES DE DONNÉES	
Profusion	http://www.profusion.com
Complete Planet	http://www.completeplanet.com




Source : CompletePlanet

Listes et forums de discussion

Une liste de discussion permet d'abriter des discussions sur un thème précis via le courrier électronique. Pour participer à ces discussions, il faut s'abonner à la liste. Dans un forum, tous les messages sont regroupés et archivés sur un serveur.

LISTES ET FORUMS DE DISCUSSION	
Listserv	http://www.lsoft.com/lists/listref.html
Google	http://groups.google.com
JISC mail	http://www.jiscmail.ac.uk



Communiqués de presse et nouvelles en continu

Les communiqués de presse diffusent des nouvelles ou des initiatives de grandes organisations. Les avantages de ces sources d'information sont qu'elles sont mises à jour régulièrement et qu'elles sont très riches en information. En revanche, on retrouve beaucoup de bruit sur ces sources, la valeur ajoutée est faible et il faut souvent effectuer d'autres recherches pour obtenir toute l'information.

COMMUNIQUÉS DE PRESSE	
Canada News Wire	http://www.cnw.ca
Business Wire	http://www.businesswire.com

Les sites de nouvelles en continu sont généralement produits par des groupes de presse faisant converger les médias (presse, télévision, Internet) pour offrir des services de nouvelles 24h sur 24h. Ces sources couvrent une variété de thèmes sur des sujets généraux. Ils sont utiles pour décoder les signaux faibles de l'environnement.

NOUVELLES EN CONTINU	
BBC	http://www.bbc.co.uk
Google news	http://news.google.com



Carnet - Blog (Weblog)

Un blogue est une page Web évolutive présentant des informations de toutes sortes, généralement sous forme de courts messages mis à jour régulièrement, affiché en ordre chronologiques inverses, et dont le contenu et la forme, très libres, restent à l'entière discrétion des auteurs." Les programmes RSS (**R**eally **S**imple **S**yndication) fonctionnant en XML sont des agrégateurs de nouvelles sur les blogues et permettent d'être averti automatiquement dès qu'un message est mis en ligne sur les blogues sélectionnés.

CARNETS	
Mario tout de go	http://cyberportfolio.st-joseph.qc.ca/mario
ConstellationW3	http://www.constellationw3.com/carnet
Research Buzz	http://www.researchbuzz.com
Corporate blogging	http://www.corporateblogging.info
ERADC	http://www.eradc.org

PLATEFORMES DE PUBLICATION	
Movable Type / TypePad	http://www.movabletype.org
Blogger	http://www.blogger.com
Live Journal	http://www.livejournal.com
Radio Userland	http://radio.userland.com

Logiciels de recherche et agents de veille

Les "agents" sont des logiciels ou des programmes informatiques dont l'objectif est de faciliter la recherche et la gestion de l'information sur Internet.

On peut distinguer cinq types d'agents intelligents:

- Agents de recherche
- Outils de cartographie
- Agents d'alerte
- Aspirateurs de sites web
- Agrégateurs de fils RSS

Pour tout connaître sur les agents intelligents: <http://www.agentland.fr>

Agents de recherche

Les agents de recherche sont aussi appelés méta-moteurs. Ceux-ci peuvent interroger plusieurs moteurs simultanément.

AGENTS DE RECHERCHE	
Copernic	http://www.copernic.com
Strategic Finder	http://http://www.digimind.fr
Metacrawler	http://www.metacrawler.com

Agents d'alerte

Les agents d'alerte sont des logiciels qui actualisent les informations ajoutées à un site préalablement répertorié. Ce sont les principaux outils utilisés en veille pour être averti automatiquement des nouveautés sur un site.

AGENTS D'ALERTE	
Google newsalerts	http://www.google.com/newsalerts
Webspector	http://www.illumix.com
Website watcher	http://aignes.com
Vigilus	http://www.vigilus.ca
TimelyWeb	http://www.topshareware.com



Outils de cartographie

Les outils de cartographie permettent d'établir les liens qui existent entre les structures d'information à l'intérieur d'un site ainsi que les liens et corrélations existants entre différents sites web.

OUTIL DE CARTOGRAPHIE	
Kartoo	http://www.kartoo.com
Alexa	http://www.alexa.com

Aspirateurs

Un aspirateur de site Web est un logiciel qui se connecte à votre place sur un site et en "aspire" les pages, c'est-à-dire qu'il effectue une copie du site ciblé sur le disque dur pour permettre une lecture "hors-ligne" par la suite.

ASPIRATEURS	
Acrobat	http://www.adobe.com
Teleport Pro	http://www.tenmax.com/teleport/pro/home.htm
WebWhacker	http://www.bluesquirrel.com/products/whacker/index.html

Agrégateurs de fils RSS

Un agrégateur de fils RSS est un outil qui permet de collecter automatiquement les fils de nouvelles sur les sites qui diffusent leur contenu en format XML ou RSS.

AGGRÉGATEURS DE FILS RSS	
SharpReader	http://www.sharpreader.net
NewsGator	http://www.newsgator.com
Bloglines	http://www.bloglines.com
NetNewsWire	http://ranchero.com/netnewswire

Autres outils et sources d'intérêt

AUTRES OUTILS ET SOURCES D'INTÉRÊT	
Signets partagés – del.icio.us	http://del.icio.us
Copernic Desktop Search	http://www.copernic.com
Grand dictionnaire terminologique de l'OLF	http://www.granddictionnaire.com
Tendances sur les carnets - BlogPulse	http://www.blogpulse.com
Traducteur en-ligne de Reverso	http://www.reverso.net/textonly/default.asp
Google Guide	http://www.googleguide.com
Google Toolbar	http://toolbar.google.com
Society of competitive intelligence (SCIP)	http://www.scip.org
Merx	http://www.merx.com/



Merci de votre attention

Rémy Gauthier
GLS Réseaux
595 5ème rue
Québec, Qc
G1J 2S6
418-524-0222

Site corporatif : <http://www.glsreseaux.com>

Carnet : http://www.glsreseaux.com/carnet_rjg

Signets partagés : <http://del.icio.us/rjg>