

Support de cours de l'intervenant

INTERNET

**Rechercher l'information sur Internet :
approfondissement des méthodes**

6, 7 octobre 2005

et **[28 octobre 2005](#)**

Animateur

Stéphane COTTIN

L'ASSOCIATION DES PROFESSIONNELS DE L'INFORMATION ET DE LA DOCUMENTATION

25, rue Claude Tillier – 75012 PARIS – Tél. : 01 43 72 25 25 – Télécopie : 01 43 72 30 41 – SIRET 784 263 311 00036 –
APE 913 E

E-mail : adbs@adbs.fr – <http://www.adbs.fr>



Formations. 2005

Rechercher l'information sur Internet : approfondissement des méthodes

(anciennement Internet : recherche avancée et outils de veille)

Autres formations : Initiation à l'Internet
Recherche documentaire sur Internet

Recherche documentaire juridique sur Internet

Contenu du CD Rom d'accompagnement

En cliquant sur les boutons à gauche de l'écran, choisissez le thème :

- Les [supports des formations](#) (initiation Internet, pratique de la recherche documentaire sur Internet, Internet juridique)
 - Des [logiciels](#) en version complète ou en démonstration
 - De la [documentation générale](#)
 - [Une liste de signets mondiale](#) et une [page de signets français](#)
-

Avertissements importants :

**1. Le CD a été vérifié par les dernières versions de deux anti-virus (McAfee et Norton)
Il est conseillé de faire une nouvelle vérification avant toute installation**

2. Toute installation de logiciel sur un poste relié à un réseau doit se faire avec l'accord du responsable de ce réseau

Toute information complémentaire :

envoyer un mail à : stephane.cottin@gmail.com

Internet : recherche avancée et outils de veille

Présentation (0)

Des rappels (Les corpus disponibles)

(1)

Les outils à disposition : Les moteurs et les répertoires (2)

Des méthodes et des stratégies (3)

La veille quotidienne (4)

Boîte à outils (5)

Autoformation (6)

Internet : recherche avancée et outils de veille

Présentation (0)

Stéphane Cottin. stephane.cottin@wanadoo.fr page perso /
weblog : <http://www.servicedoc.info>

Actuellement chef de service informatique - greffe au Conseil constitutionnel

Ancien Documentaliste de la société Lexbase SA. Spécialisée dans la conception de moteurs et de bases de données juridiques (intra- et extranets)

Ancien chef de service Documentation du Conseil constitutionnel, créateur du site Internet du Conseil constit (<http://www.conseil-constitutionnel.fr>). Collaboration active dans la conception des bases de données juridiques publiques et dans la mise en place de plusieurs systèmes d'information (Sénat, élection présidentielle, Légifrance, Conseil d'Etat...)

J'ai une connexion Internet depuis 1994. J'utilise majoritairement un PC (sous windows XP Pro) et Firefox 1.0 comme explorateur Internet plutôt qu'Explorer 6. Mes outils préférés sont Google (<http://www.google.com>) en moteur, l'Open directory (<http://dmoz.org>) en répertoire, Wticker comme lecteur de fils RSS, Wysigot comme agent d'alerte (j'ai longtemps été fan de C4U) (http://www.agentland.fr/Download/5Monitoring_agents.html) . Pour info, j'ai abandonné Copernic (<http://www.copernic.com>) comme assistant logiciel à la recherche depuis 2001.

Depuis 2003, j'utilise de façon professionnelle la technologie des [weblogs](#), à la fois pour diffuser de l'information (en interne comme en externe) et pour en recevoir : mon weblog perso est sur <http://www.servicedoc.info>

Je vous remercie de vous présenter, en indiquant quelques informations techniques :

- le type de matériel utilisé et son système d'exploitation : PC, Mac, Station, ... sous Windows, Linux, Unix...
- le logiciel de navigation préféré : Netscape, Explorer, Mozilla, autre, ...

et en particulier, en guise de mini sondage, quel(s) moteur(s) ou outil(s) de recherche utilisez-vous le plus fréquemment, et nous dire en quelques mots ce que vous en pensez (éventuellement indiquez si vous avez récemment abandonné un outil pour un autre)

VOTRE SYSTEME ET VOTRE MATERIEL	LOGICIEL DE NAVIGATION	MOTEUR de RECHERCHE FAVORI	OUTIL D'AIDE A LA RECHERCHE FAVORI
(PC, MAC, Station...) / (Windows XP, 2000,... Linux; Unix...)	Internet Explorer, Netscape, Opera, Mozilla...	Google, Yahoo, Dmoz, Altavista, ...	Copernic, Agrégateur, Météamoteur, Mouchard, Aspirateur...

Objectif: .Approfondir une méthodologie de recherche spécifique à Internet.
Exploiter des fonctionnalités de recherche avancée.
Identifier les outils de veille disponibles.

Sommaire:

Rappels et approfondissements

- les robots de recherche, les catalogues, comparaison et approfondissement
- métamoteurs, bases de données en ligne, portails et sites de référence
- les réseaux humains : newsgroups, listes de discussion

Méthodologie spécifique

- en quoi une recherche sur Internet est-elle différente d'une autre recherche ?
- quels outils pour quoi faire ?
- comment identifier de sources d'information sectorielle ou spécialisée, les "experts" à contacter, les pages de liens...
- au delà des outils : stratégies et tactiques de recherche
- recherche linéaire, recherche multidimensionnelle
- apprendre à gérer le temps et l'espace de la recherche

Veille sur Internet : stratégie et outils

- le push et les channels,
- les agents intelligents : quelle réalité ?
- les agents d'alerte,
- les aspirateurs de site,
- les métamoteurs clients,
- les logiciels de cartographie de l'information

Paramétrages avancés des logiciels de navigation et de messagerie...

Gestion avancée des signets

Les problèmes posés par l'information sur Internet

- l'évaluation de l'information : comment faire ?
- la rediffusion de l'information
- gérer la confidentialité de ses recherches et de sa veille

Autoformation

- les indispensables de l'autoformation à la recherche,
- mettre en place une veille sur la veille.

INTRODUCTION

L'Internet est devenu au fil des années un outil indispensable, mais certainement pas exclusif, dans mes recherches documentaires. Il doit être relativiser, démythifier, dompter,... pour en retirer le meilleur, au plus vite et au plus sûr. Cela nécessite non seulement une pratique quotidienne, mais aussi une remise en question permanente des outils et des usages.

C'est pourquoi je vous présenterai en dernier lieu une partie consacrée à l'autoformation à mes yeux nécessaire à l'utilisation de tous les outils mis à la disposition des internautes avertis, que nous verrons dans les quatrième et cinquième parties.

Auparavant, il aura été question de traiter d'un petit peu de méthodes et de théories de l'utilisation des outils de recherche, spécifiquement sur Internet. Mais, pour cela, il faut d'abord définir clairement les typologies d'outils de recherche, et en guise de préliminaires, revoir avec soin ce sur quoi on travaille !

Et oui, au fait, que trouve-t-on sur Internet ? Ce que l'on y cherche ? Pas si sûr ...

Internet : recherche avancée et outils de veille

Rappels (1)

Il est évident qu'avant de savoir bien chercher, il faut savoir ce que l'on peut trouver sur Internet. Il est donc important de connaître :

1) [L'histoire et la philosophie du réseau des réseaux](#)

et 2) [Les corpus à disposition et leur mode de constitution](#)

pour finir : 3) [ce qui est important à retenir pour éviter bien des déconvenues](#)

- l'histoire, la philosophie et la topologie d'Internet

Il n'est pas question ici de refaire un cours sur le sujet. Aller voir par exemple

http://agora.qc.ca/rech_int.html (J. Duquesne)

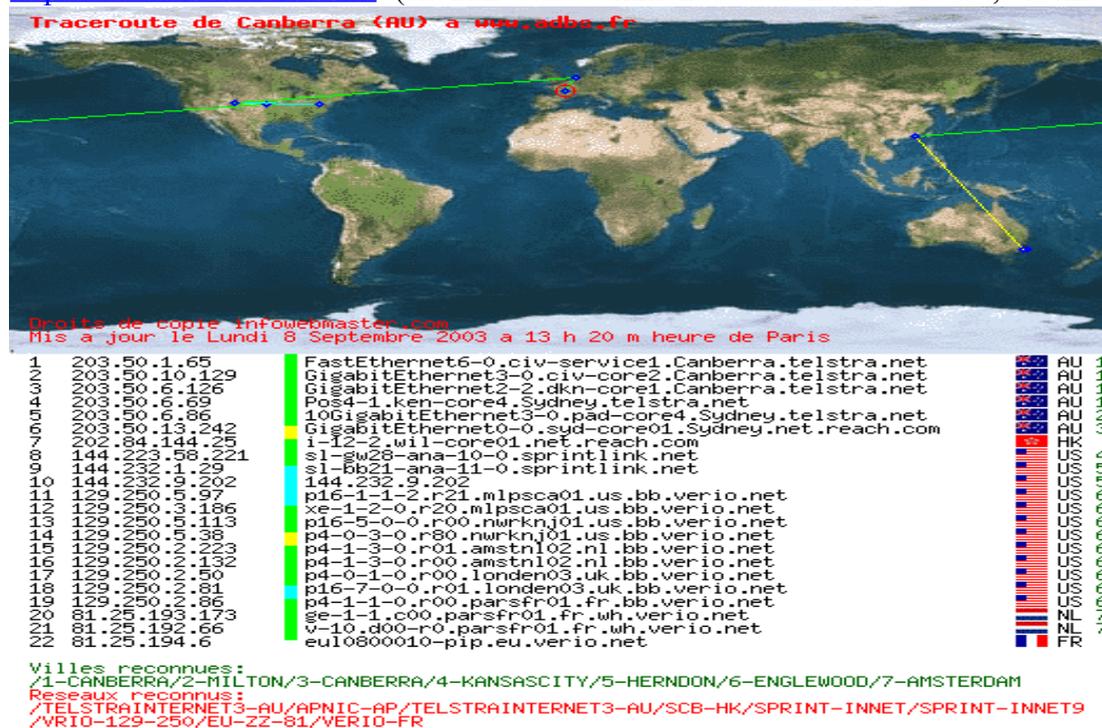
<http://www.chez.com/bibelec/links/com/internet.html> (liste de liens chez Bibelec)

On pourra retenir les aspects techniques sur les bases du TCP/IP (faire un traceroute sur son poste (Executer/command puis tracert [nom de domaine ou numéro IP] ou à

<http://www.visualware.com/indexm.html>)

application pratique et visuelle :

<http://www.visualware.com/> (chemin d'une commande de Canberra à l'adbs, via infowebmaster)



Votre message passe d'un endroit à un autre, et est conservé pour des raisons techniques à chaque passage d'où l'importance de la cryptographie !

A propos, qui est présent sur Internet ? Lire les dossiers remis à jour en permanence sur le Journal du Net : <http://www.journaldunet.com/chiffres-cles.shtml> (voir plus loin sur la taille du web)

Points de repères sur Internet - les corpus à disposition

Rappels indispensables pour savoir quels sont les documents disponibles sur le réseau des réseaux, en connaître a priori leur statut et leur histoire. Par ordre d'apparition, on a :

1°) Le Mail

L'Internet est né pour cela : et du mail sont nées les listes de diffusion (voir plus loin ou voir <http://listes.cru.fr/sympa/> ou <http://www.arobase.org/listes/index.htm>) = certainement les bases de connaissance les plus importantes en quantité et en qualité, en flux (réactivité) et en stock (ancienneté de certaines d'entre elles). Mais elles sont sous-utilisées et méconnues.

2°) L'Usenet (Newsgroups : les groupes de discussion)

L'Internet a été connu pour cela (les "autoroutes de l'information") : savoir que cela existe et ce que cela vaut. **S'en méfier en général** et tester les quelques moteurs de recherche spécialisés pour s'en servir (ou les éviter selon son goût) au mieux : <http://www.deja.com/usenet/> (ancienne adresse qui pointe désormais vers le nouveau propriétaire depuis 2002 de DejaNews : Google = <http://groups.google.com/> = on y testera notamment la commande **author** :

Usages préconisés = recherches techniques très pointues (essentiellement informatiques) ; contrôle de rumeurs.

3°) Le Web

Le coup de génie de Tim Berners-Lee.

Difficile évidemment d'en faire le tour, mais il faut en connaître :

1) les techniques (les codes HTML, et notamment les balises META : c'est ce que les moteurs lisent : http://www.abondance.com/docs/meta_1.html)

2) son évolution et ses pièges :

Taille du web : <http://www.strategic-road.com/recherch/mesurefr.htm>

2000 : les chiffres sont contradictoires mais la plupart tournent autour de 1,2 Milliard de pages statiques au début 2000, avec 100 millions de pages nouvelles par mois et surtout, une estimation d'à peu près autant de pages dynamiques (le "web invisible" voir <http://www.invisibleweb.com>) voir http://www.journaldunet.com/cc/03_internetmonde/intermonde_sites.shtml ou <http://www.nic.fr/statistiques/>

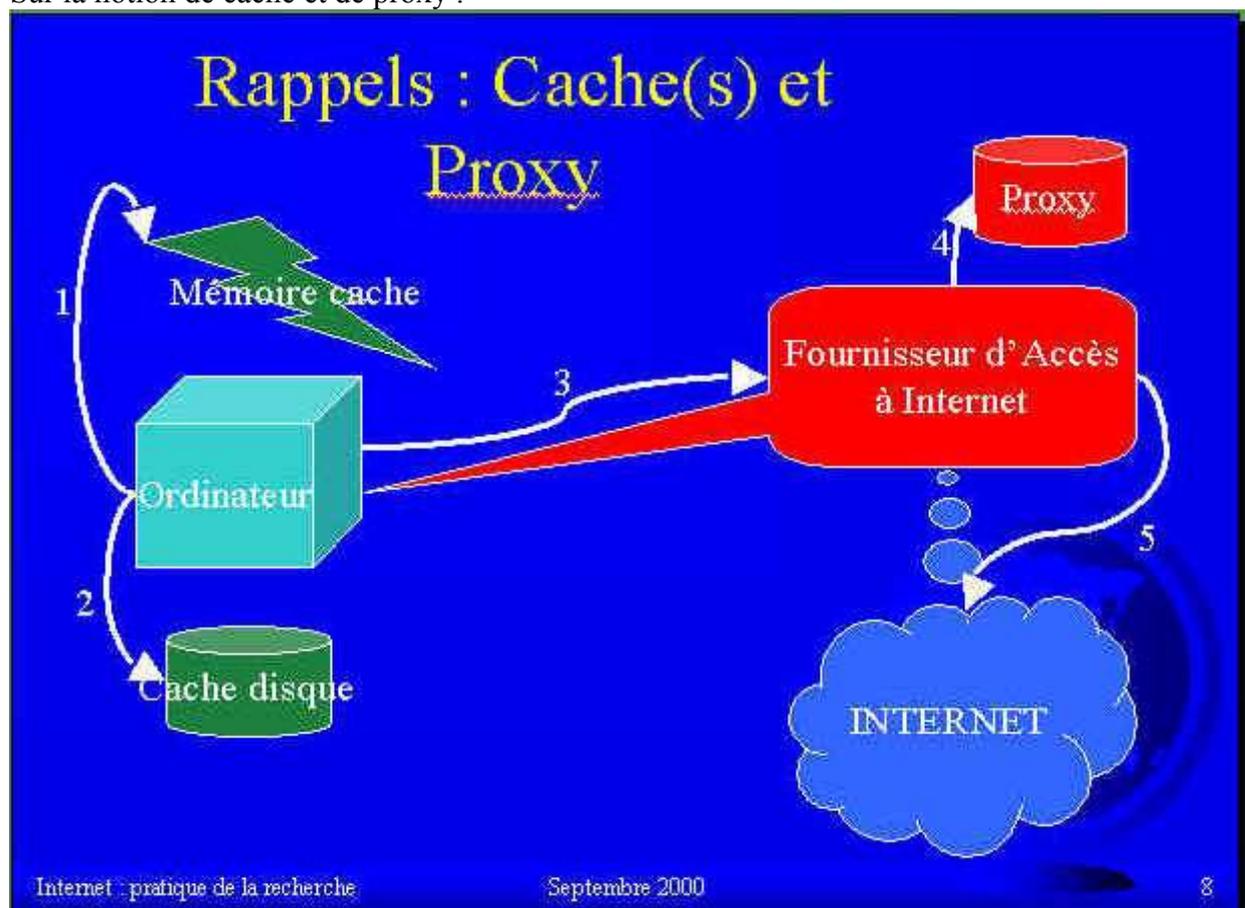
2003 : on a arrêté de chiffrer. Les moteurs de recherche se battent à coup de milliards de pages référencées. Voir sur Abondance des éléments de réponse ici :

<http://docs.abondance.com/question10.html> et sur le nombre de noms de domaines : http://www.journaldunet.com/cc/03_internetmonde/intermonde_domaines.shtml

Pages statiques / pages dynamiques : voir par exemple <http://www.legifrance.gouv.fr> ou <http://admi.net/jo/>

- à retenir, ces quelques éléments techniques

Sur la notion de cache et de proxy :



Les risques à connaître :

Le SPAM : publicité non voulue : <http://www.cypango.net/~spam/whatisspam.html>

Les HOAXES : les faux-bruits et les rumeurs <http://www.hoaxbuster.com/>

Les vrais VIRUS mais aussi le **scam** (la recherche systématique de contacts en posant des fausses questions sur les listes de messagerie, avec sa variante la plus célèbre : l'escroquerie connue sous le nom de **nigerian scam**), le **spamdexing** (tromper les moteurs de recherches en mettant de fausses balises meta), les **pages satellites** (leurrer pour les moteurs de recherche), le **spoofing** (leurrer le destinataire du message en modifiant l'entête de l'expéditeur)...

Quelques techniques à connaître =
savoir installer un logiciel (notion d'administrateur)

exercice = installer Wysigot sur le poste et mettre des mouchards sur des pages pour le lendemain (voir plus loin)

La construction des URL (uniform resource location = l'adresse des pages web)

Connaître les protocoles : http:// (hypertext transfer protocol) ; https:// (http sécurisé) ; ftp:// (file transfer protocol); nntp:// (news network...) ; smtp (sendmail...) etc.

Pour les noms de domaines : connaître les TLD : Top Level Domains, avec les ccTLD (country code TLD) à deux lettres, et les gTLD (generic TLD) à trois ou plus.

Les gTLD historiques : .com ; .net ; .org // .int // .edu ; .gov ; .mil

Les outils à disposition : Les moteurs et les répertoires (2)

- [les robots de recherche, les catalogues, comparaison et approfondissement](#)
- [métamoteurs, bases de données en ligne, portails et sites de référence](#)
- [les réseaux humains : newsgroups, listes de discussion](#)
- [le phénomène weblog et les fils RSS](#)

- les robots de recherche, les catalogues, comparaison et approfondissement

Définitions et mise en garde : **MOTEUR <> RÉPERTOIRE !**

Les moteurs **Indexation mécanique de pages web**

Évidemment sur Abondance : <http://outils.abondance.com/moteurs.html>
et leur comparatif : <http://outils.abondance.com/comparatif.html>

Historique des moteurs :

http://cbs.marketwatch.com/newsimages/misc/search_engines_timeline.pdf

(voir aussi le papier d'Alexandre Serres sur SavoirsCdi du CNDP :

<http://savoirscdi.cndp.fr/CulturePro/actualisation/Serres/Serres.htm>)

ACTUALITES 2004 : il semble que trois moteurs se dégagent. Evidemment GOOGLE toujours premier, mais qui voit son hégémonie battue en brèche par Yahoo Search (à ne pas confondre avec le répertoire Yahoo) déjà en ligne sur <http://search.yahoo.fr> et par MSN Search, (Microsoft), pas encore tout à fait opérationnel. Comparatif pénétration =

http://www.journaldunet.com/cc/03_internetmonde/intermonde_moteurs.shtml

Pensez à utiliser plusieurs moteurs, et à les comparer fréquemment sur des requêtes tests. Voir par exemple ce message de Christophe Asselin sur la liste MOTRECH d'août 2004:

Date : Tue, 24 Aug 2004 09:49:18 +0200

De : "Christophe ASSELIN - DIGIMIND" <christophe.asselin@digimind.fr>

Objet : Jux2 interroge Google et Yahoo et compare les résultats

Jux2 est un métamoteur qui permet de rechercher sur Google ET Yahoo!Search et de comparer les résultats. Dans les "bulles" oranges en haut à droite, vous pouvez voir les résultats communs entre les 2 moteurs. Où l'on se rend compte qu'en cherchant sur un seul moteur, même un Google, on rate beaucoup de chose.

<http://www.jux2.com/index.php>

NOUVEAUTE 2005 : Twingine (ex Yagoohoogle) (comparaison Yahoo! Google) voir aussi http://del.icio.us/cottinstef/comparateur_de_moteur

Nouveau : Les moteurs à suivre <http://tinyurl.com/2cmcb>

Google en Chiffres : <http://tinyurl.com/64vng>

1. <http://www.beaucoup.com>/**Beaucoup** Liste mondiale des moteurs et annuaires de recherche - - <http://www.beaucoup.com> - (très américano-centriste)
2. <http://www.searchenginecolossus.com/> Annuaire international de moteurs de recherche
3. <http://www.enfin.com/>**Enfin** Liste des moteurs et annuaires francophones – ex *IDF.NET* - <http://www.enfin.com> et son weblog : <http://joueb.com/influx/> à suivre sur <http://c.asselin.free.fr/french/faitsmarquants.htm>
4. SearchEngineWatch <http://www.searchenginewatch.com/>
5. **Searchtools** - <http://www.searchtools.com>
6. <http://www.notess.com/http://searchengineshowdown.com/>**Tests, actualités des moteurs de recherche** - <http://www.notess.com> ou <http://searchengineshowdown.com/>
7. <http://www.abondance.com/>Abondance Actualités et fiches techniques sur les moteurs de recherche - *Olivier Andrieu* - <http://www.abondance.com>

	<i>Moteurs</i>	<i>Répertoires</i>
Archétype	Google	Yahoo!
(mais aussi...)	Yahoo Search Technology (YST), MSN ; anciennement Altavista, Overture, Alltheweb, Hotbot... (tous racheté par Yahoo).	Open Directory (DMOZ)
Unité doc	Page web	Site web
Analyse	Robot	Humain
Rafraîchissement de l'index	Quotidien à au pire un mois (pour Google)	Quasiment jamais effectué
Indexation	Automatique, limitée à la page d'accueil (profondeur d'indexation de google = 4 liens et 100 ko par page)	Humaine = analyse par des documentalistes de l'ensemble du site

Les répertoires

Analyse humaine de sites Web



L'archétype : <http://www.yahoo.fr> . Savoir s'en servir : <http://fr.docs.yahoo.com/aide/>

Le concurrent injustement méconnu : <http://dmoz.org> .

- métamoteurs, bases de données en ligne, portails et sites de référence

Rappel : le web invisible. (voir sur le site de l'Urfist de Lyon à la page d'accueil <http://urfist.univ-lyon1.fr/>)

Archétype : <http://www.strategic-road.com> (mayaconcept)

Bases de données gratuites sur Internet : <http://urfist.univ-lyon1.fr/gratuits.html> (normalement arrêté)

Les portails :

nouvelle notion de "vortails" : portails verticaux

exemple de site de référence : <http://www.droit.org>

exercice : trouver d'autres sites de référence dans d'autres domaines

- les réseaux humains : newsgroups, listes de discussion, bookmarks et logiciels sociaux (CMS et del.icio.us)

From: Marie-Luce PARANQUE-THOMAS <paranque@ocirp.fr>

To: "'Liste de diffusion ADBS'" <adbs-info@cru.fr>

Subject: Synthèse des réponses sur la création d'une liste de diffusion

Date: Mon, 4 Oct 1999 12:24:31 +0200

exercice. Trouver ce message de la liste de discussion ADBS-Info.

Bonjour,

Tout d'abord merci à tous pour vos réponses rapides et intéressantes...

Inscrite depuis peu sur la liste, je suis déjà impressionnée par l'efficacité de ce système et tout cela grâce à la participation de chacun ! Je vous fais part des adresses qui m'ont été communiquées :

<http://www.arobase.org/listes/automats.htm> , pour s'informer sur la marche à suivre pour créer une liste de diffusion, et/ou pour faire héberger une liste (ce qui en simplifie la gestion). Voir aussi le leader <http://fr.groups.yahoo.com/>

Les solutions existantes sont donc la sous-traitance, l'hébergement gratuit (cf adresses ci-dessus) ou l'acquisition d'un logiciel spécifique. Tout dépend des moyens dont on dispose et du temps que l'on veut consacrer ensuite à l'administration de la liste. D'après une synthèse adressée à la liste adbs le 27/7/99, il faut réfléchir à certaines questions : public concerné et ses besoins, moyens et ressources dont on dispose (connaissances techniques entre autres), solutions existantes (sous-traitance, gestion interne, serveur associatif (globenet), services gratuits (type yahoo, simple et efficace, mais avec pubs)), liste modérée, gestion d'archives, suivi et animation de la liste. Le temps passé à l'administration variera en fonction de tous ces éléments. Pour débiter, on peut utiliser le logiciel de messagerie en sélectionnant autant de destinataires qu'on le souhaite. Pour apprendre à utiliser sa messagerie : <http://www.memoclic.com>

Marie-Luce Paraque, OCIRP-Paris 8 (Union d'institutions de prévoyance régie par le code de la Sécurité Sociale), <http://www.ocirp.fr>, paraque@ocirp.fr ,

Des listes de listes : francophones = <http://www.francopholistes.com/>

Mondiales : <http://www.tile.net/> [google it via dmoz !]

Le phénomène Weblogs et les fils RSS

Les **weblogs**, ou 'blogues' sont nés de la rencontre du phénomène de simplification des techniques de publication sur Internet, et de celui de la volonté toujours présente de partager ses informations avec le plus grand nombre. Ce dernier phénomène, principe de base de l'Internet des premiers temps, connaît un jour nouveau avec ces nouveaux moyens de publication rapide, simple, souple... (voir http://www.servicedoc.info/article.php3?id_article=28)

Intimement liée aux weblogs, mais sans en être une caractéristique, la **syndication** est une technique permettant d'afficher des données provenant (et offertes) d'autres sites, dans son propre site. C'est l'archétype de la gestion de contenu : c'est de l'information venant d'ailleurs, mise à disposition (éventuellement filtrée, reconfigurée...) du plus grand nombre.

La technique utilisée est issue du XML, mais très simplifiée : elle est d'ailleurs nommée RSS pour Really Simple Syndication, en fait une version dépouillée de la norme RDF. On peut en profiter directement si on utilise un CMS (Content management system, comme SPIP par exemple) qui en tient compte, mais il est aussi possible d'insérer un simple code java script (voir http://www.servicedoc.info/article.php3?id_article=57)

Les fils RSS servent alors soit à afficher dans un intranet ou dans un autre site internet, les infos publiées sur le weblog, un peu comme une fenêtre d'actualité, soit à être collectés via des RSS-aggregators, des lecteurs de fils. De la même façon que l'on ouvre un utilitaire de messagerie ou un lecteur de news, on peut "s'abonner" à tel ou tel fil et lire en direct les infos provenant de ressources diverses (voir http://www.servicedoc.info/article.php3?id_article=100)

Très anecdotique, notamment en Europe, jusqu'en 2002, cette méthode, pourtant ancienne (les premiers weblogs et fils RSS datent de 1997) a récemment explosé, tant et si bien qu'elle a sinon révolutionné le circuit de l'information dans certains secteurs (par exemple celui de l'information et de l'auto-formation à la recherche documentaire), au moins influer très sensiblement sur la politique d'indexation des gros moteurs de recherche.

Exemples de fils RSS sur la recherche documentaire (en général, le weblog associé est la racine du site hébergeant le fil):

<http://www.librarystuff.net/>

<http://www.biblioacid.org>

<http://google.blogspot.com/>

<http://www.figoblog.org/>

<http://blogokat.canalblog.com/>

<http://kesako.canalblog.com/>

<http://www.ouilsfroids.net/>

<http://urfistinfo.blogs.com/>

<http://inforizon.blogs.com/veille/>

Installer un lecteur de fil : Wticker <http://www.wticker.org>
(présent aussi sur le CD Rom dans le sous-répertoire logiciels)

Des méthodes et des stratégies (3)

Les dix règles d'or de la recherche d'information sur Internet

1. "Affiner" sa question (type de recherche, sujet précis et objectif, étude des concepts, recherches préliminaires éventuelles), choisir ses stratégies de recherche.
2. Maîtriser les outils de navigation et de recherche : gestion des signets, récupération des données, répertoires, moteurs et méta-moteurs. Pour les moteurs, utiliser au moins deux moteurs ayant des approches différentes et complémentaires.
3. Trouver de bons points de repère : annuaires et "bons sites" dans un domaine :
 - Retrouver les équivalents de ses sources habituelles
 - Compléter avec les sources originales
 - Trouver les répertoires et "méta-pages" spécialisées.
4. Toujours analyser l'information : recouper l'information, faire preuve d'esprit critique, évaluer rapidement
5. Utiliser en cours de recherche son carnet d'adresses pour garder trace des sites ou pages intéressants mais momentanément hors sujet, et "noter" rapidement les ressources enregistrées.
6. Savoir se limiter dans le temps : ne pas se rendre esclave d'une recherche d'exhaustivité à tout prix, ne pas s'obstiner en vain. Internet contribue souvent à répondre à la question "où trouver" (chercher l'info qui conduira à l'info).
7. Choisir les bons mots-clés
8. Rester clair sur ses objectifs, sa stratégie et ses critères de choix établis auparavant face à "l'hyper-choix". Rester vigilant sur la trajectoire parcourue et celle qui reste à parcourir.
9. Conjuguer harmonieusement recherche dans les outils classiques, web invisible, presse et actualité et navigation hypertexte : la recherche d'information sur Internet est un processus itératif qui oblige à passer par différents modes d'accès à l'information.
10. Etre "agile" : Développer une lecture rapide, lancer plusieurs recherches à la fois, savoir rebondir d'une information à l'autre, d'un outil à l'autre, d'un article à une institution. Se souvenir qu'il n'existe pas de méthode infaillible et que chercher l'information sur Internet, c'est avant tout un état d'esprit.

1. "Affiner" sa question (type de recherche, sujet précis et objectif, étude des concepts, recherches préliminaires éventuelles), choisir ses stratégies de recherche.

Choisir des outils

Questions à se poser	Réponses
Quel catalogue ? Quel moteur ?	Vous devez connaître les outils qui sont susceptibles de vous être utiles (veille sur la recherche d'info.).
Quel moteur de newsgroups ? Quelle liste de liste ?	en guise de questions : savez-vous ce qui circulent sur les newsgroups ou sur les listes ? Avez-vous le temps de l'évaluer, de le recouper ? Avez-vous le temps (et le pouvoir hiérarchique), d'initier des discussions utiles pour vous ? Quelles sont vos marges de manoeuvres en matière de divulgation d'information
Quel moteur utiliser ?	Celui que vous maîtrisez le mieux. Fondamentalement, les gros moteurs indexent tous la même chose.

Sur le site web de la BnF un nouvel outil en ligne le Guide de recherche en bibliothèque

(<http://grebib.bnf.fr>) esquisse les grandes lignes d'une recherche d'information à la Bibliothèque nationale de France, et plus particulièrement dans les collections imprimées et audiovisuelles. Destiné avant tout aux lecteurs débutants, il peut également être utile à des habitués par sa rubrique "Miscellanées BnF" qui détaille des types ou des fonds de documents plus complexes à trouver que de simples ouvrages. Deux types de navigation sont proposés :

- une navigation pas à pas qui permet de progresser logiquement d'étape en étape, de la formulation du sujet à la rédaction du texte et de la bibliographie, en passant par les phases de recherche dans les bibliographies, les catalogues et la localisation des documents,

- une navigation libre qui permet d'aller directement à la rubrique souhaitée. Bien que certains paragraphes illustrent la complexité des recherches dans une bibliothèque nationale, la démarche présentée dans le GREBIB peut s'appliquer à toute autre bibliothèque et être utile particulièrement aux usagers des bibliothèques universitaires. La méthodologie repose en partie sur une méthode de résolution de problème d'information développée aux Etats Unis par Mike Eisenberg et Bob Berkowitz sous le nom de Big 6.

Les pages sont générées à partir d'un référentiel XML et d'un moteur XSL-T. Ce guide est la version entièrement mise à jour et ergonomiquement adaptée au web du guide imprimé diffusé dans les salles de lecture depuis 2002 et correspondant lui-même à une première version accessible en ligne sur le site de la BnF de 1999 à 2003.

Ce nouvel outil peut être utilisé à des fins pédagogiques de formation des lecteurs sans but lucratif et avec mention "© Bibliothèque nationale de France".

Raymond Josué Seckel , Catherine Bonhomme
Bibliothèque nationale de France
Direction des collections
Département de recherche bibliographique

Voir aussi :

<http://www.philb.com/whichengine.htm>

<http://library.albany.edu/internet/choose.html>

<http://www.noodletools.com/debbie/literacies/information/5locate/advicengine.html>

<http://www.infopeople.org/search/chart.html>

<http://searchenginewatch.com/facts/article.php/2156031>

ajout fin 2005 : le site REPERE de l'ENSSIB = <http://reperere.enssib.fr>

2. Maîtriser les outils de navigation et de recherche : gestion des signets, récupération des données, répertoires, moteurs et méta-moteurs. Pour les moteurs, utiliser au moins deux moteurs ayant des approches différentes et complémentaires.

- quels outils pour quoi faire ?

Les moteurs pour les recherches lourdes, sans but précis, juste pour vérifier la présence (ou l'absence) d'une notion, d'un fait,... sur le réseau. Savoir alors repérer des notions-clefs, si possible discriminantes, pour éviter le bruit (par exemple numéros de loi, dates, noms de personnes rares...)

Pour les recherches "fines", les annuaires pour trouver l'adresse d'un site web sur lequel, par expérience, on est sûr de pouvoir trouver ce que l'on cherche.

Trucs divers (voir [boîte à outils](#))

- Gérer les signets (utilitaires)
- Paramétrages des navigateurs
- Enregistrer des pages
- Edition-Rechercher (CTRL + F)
- Mémoire cache
- Adresse mel de complaisance

3. Trouver de bons points de repère : annuaires et "bons sites" dans un domaine :

- Retrouver les équivalents de ses sources habituelles
 - Compléter avec les sources originales
 - Trouver les répertoires et "méta-pages" spécialisées.
- Utiliser la commande related: de google
- Si on y a accès, lire les logs de connexion d'un site d'information, et notamment de ses referers (ceux qui pointent vers ces informations ont forcément quelque chose à voir avec ces informations = notion de « cosmos » voir <http://www.technorati.com>)

4. Toujours analyser l'information : recouper l'information, faire preuve d'esprit critique, évaluer rapidement (connaître les principes de bases des constructions d'URL – noms de domaines, savoir rechercher les entités responsables d'une page web ou d'un site, connaître en gros la structuration des corpus disponibles, leur âge, leur redondance...)

- comment identifier de sources d'information sectorielle ou spécialisée, des "experts" à contacter, des pages de liens ?

Recouper. Utiliser plusieurs outils. Ne pas hésiter à lancer un fil de discussion sur une liste, tout en respectant la netiquette. (voir aussi www.arobase.org) Faire des synthèses à partager avec les autres.

5. Utiliser en cours de recherche son carnet d'adresses pour garder trace des sites ou pages intéressants mais momentanément hors sujet, et "noter" rapidement les ressources enregistrées. Utiliser le bloc note (notepad) = faire un lien sur son bureau vers le programme notepad.exe, et apprendre à manipuler les combinaisons de touches ALT+TAB (basculer d'une fenêtre à une autre), CTRL+C (copier) et CTRL+V (coller). Apprendre surtout à sélectionner sur les documents complexes (type HTML avec ou sans cadres, le PDF...).

Utiliser l'historique de son navigateur (au besoin, en augmenter son volume)

6. Savoir se limiter dans le temps : ne pas se rendre esclave d'une recherche d'exhaustivité à tout prix, ne pas s'obstiner en vain. Internet contribue souvent à répondre à la question "où trouver" (chercher l'info qui conduira à l'info).

Méthode : gardez le fil

- Ne vous laissez pas emporter par la vague. Gardez le fil de votre recherche (éviter le phénomène Dictionnaire) (a contrario, ne pas négliger le phénomène dit de « serendipité » (l'heureux hasard)
- Représentez-vous mentalement vos étapes de recherche et votre logique
- Ne remettez pas en cause toute votre tactique à cause d'une étape (i.e : remettez juste en cause l'étape !)

Aidez-vous avec l'historique et les signets (CTRL + H) (finalement, le papier a aussi du bon ; apprenez à utiliser NOTEPAD et les combinaisons de touches)

Méthode : la gestion du temps

Dire qu'on n'a pas le temps de chercher, c'est mal poser le problème.

Vous ne pouvez évaluer le temps d'une recherche qu'après avoir déterminé une stratégie de recherche.

Ce temps dépend surtout de votre culture du sujet car on perd énormément de temps à trouver le bon terme, la bonne expression = commencer par le trouver ; utiliser des dictionnaires

(<http://www.granddictionnaire.com> : très bon pour les synonymes et la traduction en anglais - et en latin !)

Ne vous surestimez pas. Si vous ne connaissez pas bien le domaine, ajoutez une étape "recherche de terminologie" dans laquelle vous allez vous former à la terminologie du domaine. Utilisez les catégories yahoo, une recherche vague sur Google et analyser les premières réponses, les documents d'encyclopédie en ligne, les articles de presse, etc.

7. Choisir les bons mots-clés

Méthode : analysez l'échec d'une étape

L'échec d'une étape de recherche est le plus souvent dû :

- à des termes ne correspondant pas à ceux qui sont réellement utilisés (connaître les bons outils linguistiques : www.granddictionnaire.com (français technique, anglais, latin !) <http://atilf.atilf.fr/tlf.htm> (le fameux Trésor de la langue française) et les ressources de traduction de l'union européenne http://europa.eu.int/comm/dgs/translation/index_fr.htm
- à des outils ne permettant pas de trouver l'information (Ex. : altavista ne cherche pas dans les bases de données)
- à une direction erronée (ne cherchez pas d'info. sur l'ethnologie des mineurs du Kouzbass dans DisneyChannel).
- avez-vous pensé à retirer le filtrage ? : voir <http://cyber.law.harvard.edu/people/edelman/google-safesearch/> et faites le test sur <http://asp-cyber.law.harvard.edu/filtering/google-safesearch/test.asp> ou sur <http://www.buzztoolbox.com/google/domainpur.shtml> ou <http://www.researchbuzz.org/archives/001411.shtml> (ma google key est 3wJB0/JQFHIjLag9z1eBxH2WVqKVTqcZ)

8. Rester clair sur ses objectifs, sa stratégie et ses critères de choix établis auparavant face à "l'hyper-choix". Rester vigilant sur la trajectoire parcourue et celle qui reste à parcourir.
- en quoi une recherche sur Internet est-elle différente d'une autre recherche ?

Problème : Internet n'est pas conçu comme un outil de recherche. Ce ne sont que des rustines qui permettent de "faire des recherches" sur des corpus non stables, non structurés et non validés.
Risques : manque de fraîcheur ; "flicage" ; publicité intempestive ; faux-bruits ; bruit tout court.

Avantages : fréquente gratuité ; rapidité de réaction ; immensité du corpus ; international et à la fois très "local".

9. Conjuguer harmonieusement recherche dans les outils classiques, web invisible, presse et actualité et navigation hypertexte : la recherche d'information sur Internet est un processus itératif qui oblige à passer par différents modes d'accès à l'information.

- Méthodes : Recherches passives/actives

Une évolution surprenante (un peu artificielle) du monde Internet a fait apparaître une rupture entre deux méthodes :

la première consiste à tirer (pull) de l'existant des données et des infos : cela constitue la base de la navigation

la seconde consiste à se voir rapatrier par des agents externes (qui "poussent" vers vous) ces mêmes données

- **Les méthodes "pull"**

les plus classiques : navigation hypertexte, recherches via moteurs d'indexation, participation passive à des newsgroups, des listes de messagerie (recherche dans leurs archives par exemple), recherches sur des sites spécialisés (comme sur des bases de données en ligne)

- **Les méthodes "push"**

Les nouveaux "info-agents", limiers, traqueurs de nouvelles...

Application : Wysigot

Avant même l'apparition de tels outils, et pour que l'information soit réellement partagée, le fait de participer activement aux newsgroups, aux mailing lists, de créer ses pages web... correspondait déjà à de l'externalisation de votre info ; et... en contrepartie à amener (façon « veille ») de l'info externe vers vous.

- au delà des outils : stratégies et tactiques de recherche

1) Méthodes de long terme

- Méthode du 'lurker' = participation passive à des forums de discussions

- Mini-veille = installation d'outils minimalistes de surveillance de pages stratégiques

- Méthode de la chèvre = attirer l'information à vous = devenez source d'info !

2) Méthodes de court terme

- Méthode de la bouteille à la mer = usage inconsidéré de newsgroup ou de liste de discussion (à éviter sauf...)

- La "personne ressource" (parfois développé comme "business plan" dans certains cas comme AskJeeves <http://www.ask.com/> , ou sur Google Answers : <http://answers.google.com/>)

10. Etre "agile" : Développer une lecture rapide, lancer plusieurs recherches à la fois, savoir rebondir d'une information à l'autre, d'un outil à l'autre, d'un article à une institution. Se souvenir qu'il n'existe pas de méthode infaillible et que chercher l'information sur Internet, c'est avant tout un état d'esprit.

Méthode : gestion de l'espace de la recherche

- Apprenez à lire un écran rapidement. Éduquez votre œil.
- Apprenez à mentaliser le document dans son ensemble (et pas juste ce qui est affiché)
- Apprenez à maîtriser vos clics de souris. Chaque clic sur un lien doit être pesé, rationnel (pensez à utiliser plutôt le clic droit et ouvrir dans une nouvelle fenêtre, puis à [manipuler vos fenêtre avec ALT TAB](#)). Pas un clic ne doit être impulsif. Il n'est pas interdit de surfer "au hasard", mais dans une logique de veille, pas dans une logique de recherche chronométrée.

Lors d'une recherche, ne faites pas n'importe quoi. Quand vous cherchez hors d'Internet, vous ne faites pas n'importe quoi. Alors ayez conscience de l'économie de votre recherche

La veille quotidienne (4)

- le push et les channels, ancêtres des fils RSS
- les agents intelligents : quelle réalité ?

<http://www.strategic-road.com/recherch/agents/agentsfr.htm>

<http://www.agentland.fr/> (attention, plein de pub, et gros risques de spywares)

http://www.cybion.fr/pages/technologie/labo_agents.html



http://www.cybion.fr/pages/technologie/labo_agents.html

- les agents d'alerte,

des gratuits ... Google Alert = <http://www.googlealert.com/> (site privé non relié à Google) / à ne pas confondre avec les Google web Alerts <http://www.google.com/webalerts> (site « officiel » de Google)

TRACERLOCK : <http://www.tracerlock.com/> (test gratuit pour 30 jours)



<http://www.watchthatpage.com/>



<http://www.infominder.com/webminder/index.jsp>



<http://www.trackengine.com/>

... ou peu onéreux :

WebSite-Watcher *Save Time,
Stay Informed*

Website Watcher : <http://www.aignes.com/index.htm> (version démo pendant 30 jours, 30 \$ version perso)

TRACERLOCK

<http://www.tracerlock.com/> (4 \$ / mois)

<http://www.wysigot.com/fr/> Wysigot est le logiciel idéal pour capturer, surveiller, gérer et archiver tous les documents, sur internet, votre réseau local ou votre ordinateur. (version gratuite limitée, payante à 31 €)

- les aspirateurs de site, (mot clef anglais : offline browser)

<http://dmoz.org/Computers/Software/Internet/Clients/WWW/Browsers/>

http://dir.yahoo.com/Business_and_Economy/Business_to_Business/Communications_and_Networking/Internet_and_World_Wide_Web/Software/Development/Offline_Web_Agents/



WEBZIP : <http://www.spidersoft.com/>

- les métamoteurs clients,

COPERNIC : <http://www.copernic.com>

- les logiciels de cartographie de l'information

UMAP : <http://www.trivium.fr/fr/products.htm>

<http://c.asselin.free.fr/french/carto.htm>

Boîte à outils (5)

Paramétrages avancés des logiciels de navigation et de messagerie

- <http://www.indexa.fr/ConfigNav/index.html> (un peu dépassé mais encore utile)

- reconnaître les extensions des fichiers (savoir les afficher sur Windows : elles sont cachées par défaut), puis la retrouver sur <http://webperso.easynet.fr/~cibderf/> ou sur <http://www.commentcamarche.net/systemes/format.php3> (moins riche en extensions mais plus précis)

Google it !

- <http://www.dsi.cnrs.fr/RMLR/textesintegraux/volume4/415-cirdu21-01-2002.htm> **Circulaire du 23 mai 2001 relative au choix des formats pour la réalisation de documents constitués de données numériques**

- Jargon :

- en français : <http://www.linux-france.org/prj/jargon/>

Gestion avancée des signets

hébergement d'annuaires (partagé ou non)

Les logiciels sociaux et le « social bookmarking »: <http://del.icio.us>



<http://www.mylinea.com/>

sélection de sharewares : http://download.com.com/3150-2370-0.html?tag=stbc_gp (bookmark managers)

Webnotes : <http://www.webnotes.org/> (français, version 6.3 freeware, 7.0 shareware (20 €))

Powermarks : <http://www.kaylon.com/power.html> (shareware : 25 \$)

Powermarks™

autres solutions : le faire soi-même :

- via un CMS (genre SPIP, par exemple avec le squelette Biospip)

- via un petit outil de gestion de signet sur une base de donnée (voir Hypathia.mdb sur le CD-Rom)
-

Les problèmes posés par l'information sur Internet

- l'évaluation de l'information : comment faire ?

- recouper et se méfier !
- connaître l'utilisation des WHOIS, des cadres "invisibles", des http-equiv, des reroutages...

- la rediffusion de l'information :

- [Le droit des intranets \(intervention décembre 1999\)](#)
- Les recommandations du forum des droits sur l'internet : <http://www.foruminternet.org/recommandations/>

- gérer la confidentialité de ses recherches et de sa veille

- Des adresses email de complaisance
- Nettoyer ses traces (voir sur Agentland par exemple)
- Ne pas se faire blacklister

Autoformation (6)

- les indispensables de l'autoformation à la recherche,
- mettre en place une veille sur la veille

- les indispensables de l'autoformation à la recherche,

FORMATIONS à la recherche

Évidemment : <http://urfist.univ-lyon1.fr/risi/risi.htm> le RISI : Recherche d'Information Sur l'Internet de Jean-Pierre Lardy (Adbs) <http://urfist.univ-lyon1.fr/jpl.html>

- [Formist Réseau francophone pour la formation à l'usage de l'information dans l'enseignement supérieur-](http://formist.enssib.fr) <http://formist.enssib.fr>
- [Arobase](http://www.arobase.org/) Guide d'utilisation de la messagerie électronique - - <http://www.arobase.org/>
- [Mata Hari](http://www.thewebtools.com/) - <http://www.thewebtools.com/>
- [Search Engine Guide](http://www.searchengineguide.com) - <http://www.searchengineguide.com> -
- [Un nouveau guide Internet](http://guide.ungi.net/) - <http://guide.ungi.net/> (attention 99 !)
- [GIRI : Guide d'initiation à la recherche sur internet](http://www.bibl.ulaval.ca/vitrine/giri/) CREPUQ - Sous-comité des bibliothèques. Groupe de travail sur l'accès aux ressources documentaires. Sous-groupe de travail sur Internet - - <http://www.bibl.ulaval.ca/vitrine/giri/>  [CANADA](http://www.bibl.ulaval.ca/vitrine/giri/)
- <http://www.unil.ch/ci/cours/supports/search/>

Guides de recherches (généralistes)

C E R I S E

Conseils aux Etudiants pour une Recherche d'Information Spécialisée Efficace

CERISE : <http://www.ccr.jussieu.fr/urfist/cerise/>

Guides BNF : <http://grebib.bnf.fr/> et <http://signets.bnf.fr/>

Catégorie Yahoo consacrée à la Recherche sur Internet :

http://fr.dir.yahoo.com/Informatique_et_Internet/Internet/World_Wide_Web/Recherche_sur_le_Web/

équivalent en anglais :

http://dir.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Searching_the_Web/

SAPRISTI!= Sentiers d'Accès et Pistes de Recherches d'Informations Scientifiques et Techniques sur Internet consultable à : <http://docinsa.insa-lyon.fr/saprستي/index.php>

- mettre en place une veille sur la veille

Lire la lettre actu-moteurs d'Olivier Andrieu (site [Abondance](http://www.abondance.com)) : <http://lettres.abondance.com/>
(version gratuite hebdomadaire ou mensuelle payante 6€ / mois)

(voir en annexe un extrait des lettres "Recherche et Référencement")

S'abonner à des listes de messageries spécialisées :

MOTRECH : <http://motrech.free.fr/>

ou consulter régulièrement les sites web des responsables et des animateurs de la liste : Olivier Andrieu (<http://www.abondance.com>) ; Jérôme Charron : (<http://www.chez.com/jcharron/motrech/presentation.html>)

VEILLE : <http://fr.groups.yahoo.com/group/veille>

ou consulter régulièrement les sites Internet des responsables de la liste : Intelligence Stratégique sur Internet <http://www.cybion.fr> <http://www.veille.com>

Méthode :

Apports théoriques et pratiques.

Utilisation de logiciels spécifiques.

<p>Exercice dirigé (si le temps le permet) : création d'une liste de discussion sur Yahoo-Groupes. Gestion de la liste et apprentissage des principales fonctionnalités</p>

<http://fr.groups.yahoo.com/>

Extraits (avec l'aimable autorisation de Monsieur Olivier Andrieu © Abondance 2005) des lettres "R&R" (Recherche et référencement) de juin et juillet/août 2003

Présentation de la lettre d'information "Recherche & Référencement" du site Abondance **www.abondance.fr**

La lettre professionnelle "**Recherche et Référencement**" convient à tous ceux dont le métier consiste à **rechercher des informations** de la façon la plus efficace possible sur le réseau et à tous ceux qui désirent "créer du trafic sur leur site Web". Elle vous propose l'actualité et les nouvelles fonctionnalités des outils de recherche majeurs, les nouveaux moteurs, les nouvelles versions des logiciels, bref, toute l'information dont vous avez besoin pour mener vos recherches de la façon la plus efficiente possible. Elle vous propose également les astuces des moteurs, l'actualité fine du **référencement**, les nouveaux moteurs pour s'y référencer, etc.

Nom : "**Recherche & Référencement**" ou "R&R" pour les intimes ;-).

Périodicité : mensuelle (envoi aux environs du 15 de chaque mois).

Modalité d'envoi : le signalement de la disponibilité de la lettre est effectué par e-mail. La lettre est disponible sur le Web, dans la zone "abonnés" du site Abondance dès le jour de son envoi. Formats disponibles : HTML, Word (.doc), PDF (.pdf) ou par e-mail.

Le site Abondance propose 3 types de **premier abonnement** :

- Abonnement "Fidélité" pour **2 ans** (22 numéros de la lettre R&R, accès complet aux archives depuis 1998). **115 euros HT**
- Abonnement "Premium" pour **1 an** (11 numéros de la lettre R&R, accès complet aux archives depuis 1998). **65 euros HT**
- Abonnement "Découverte" pour **6 mois** (6 numéros de la lettre R&R, accès complet aux archives depuis 1998). *Disponible uniquement pour un premier abonnement (pas de réabonnement possible sur 6 mois).* **40 euros HT**

Et deux types de **réabonnement** :

- Réabonnement "Fidélité" pour **2 ans** (22 numéros de la lettre R&R, accès complet aux archives depuis 1998). **110 euros HT**
- Réabonnement "Premium" pour **1 an** (11 numéros de la lettre R&R, accès complet aux archives depuis 1998). **60 euros HT**

Actuellement : Sommaire de [la lettre du mois en cours](#).

Témoignages  : Lisez [plus de 40 témoignages](#) sur la lettre de la part d'abonnés actuels !

Abonnés  : Consultez [une liste de sociétés abonnées](#) à la lettre.

Archives : Toutes les archives de la lettre "R&R" depuis ses premiers numéros sont disponibles sur la [zone "abonnés" du site Abondance](#) (accès compris dans le prix de l'abonnement).

Les **numéros** sont également en **vente à l'unité** sur la [Boutique Abondance](#).

Suivi à distance

Exercices proposés :

Plateforme YahooGroupes : <http://fr.group.yahoo.com/groups/adbs730/>

1) Trouver le(s) responsable(s) d'un site (applications : WHOIS, notions : mentions légales, registrar, billing contact, administrative contact...)

ex: conseil-constitutionnel.fr
ex : www.adbs.fr
ex : servicedoc.info
ex : whitehouse.gov / whitehouse.com [attention, contenu "adulte"]
ex : legifrance.gouv.fr / legifrance.com

2) Utilitaires pratiques (facultatif) :

- Installer et utiliser une toolbar (googlebar, yahoo, copernic-meta..., ou NeedleSearch pour les utilisateurs de Mozilla)

- Installer un utilitaire "anti-pop-up"

installer la googlebar : critiques et appréciations

3) Comprendre le travail d'un editor de l'Open Directory (www.dmoz.org)

lire <http://www.dmoz.org/World/Français/about.html>

4) Installer et localiser des thèses (ou des travaux universitaires)

les fonds sudoc et abes = chercher des autres fonds

les cas particuliers, thèses en cours, thèses soutenues par matière, par université...

localisation, acquisition, texte intégral