



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

ADBS

Support de cours de l'intervenant

INTERNET

Rechercher l'information sur Internet :
approfondissement des méthodes

12, 13 et 14 juin 2006

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>

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)

Travaux pratiques (7)

L'ensemble du support de cours est à suivre en ligne sur :
<http://www.servicedoc.info/730/>

Ce support de cours de l'intervenant est le fil conducteur de ces trois jours de formation. Il s'appuiera sur l'ouvrage « Net Recherche » de Véronique Mesguich et Armelle Thomas.

Tout au long de ce support de cours, les références à l'ouvrage en question seront notées
(voir NR p. xx)

Issue de leurs expériences cumulées à l'occasion notamment des [stages de formation ADBS "Recherche avancée sur Internet \(approfondissement des méthodes\)" \(n° 730\)](#), deux des formatrices attitrées, [Véronique Mesguich](#) et [Armelle Thomas](#), proposent aux éditions de l'ADBS, une version très largement améliorée, luxueuse et modernisée du support de formation.

Net recherche : le guide pratique pour mieux trouver l'information utile

Véronique Mesguich et Armelle Thomas

Sciences et techniques de l'information (ISSN 1762-8288)

2006 - 148 pages - ISBN 2-84365-083-6

Résumé

Qu'elle réponde à des besoins professionnels ou personnels, la recherche d'information représente une grande part de l'activité des internautes. Le fonctionnement des moteurs de recherche est pourtant encore mal connu, et les utilisateurs sont souvent perdus face à des résultats trop nombreux et imprécis, à des contenus structurés ou non, validés ou non, gratuits ou payants... L'art de la recherche d'information consiste pour une grande part à déterminer les outils, généralistes ou spécialisés, les plus adaptés aux différents types de requêtes et à les utiliser avec maîtrise.

Cet ouvrage a pour objectif d'offrir à toute personne amenée à effectuer des recherches sur Internet un panorama des outils et méthodes existant à ce jour. Il est destiné tant aux professionnels de l'information (documentalistes, bibliothécaires, veilleurs) qu'aux enseignants, chercheurs, étudiants, confrontés à la complexité sans cesse grandissante des données disponibles en ligne. Ils y trouveront des conseils méthodologiques ainsi que des « trucs et astuces » destinés à optimiser le processus de recherche ; des informations précises sur les évolutions actuelles ainsi que des présentations d'outils et des listes d'adresses utiles.

[Introduction \(pdf\)](#)

[Table des matières \(pdf\)](#)

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 1.5 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 (mais j'avoue qu'avec Wticker, je ne l'utilise plus, et 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 (PC, MAC, Station...) / (Windows XP, 2000,... Linux; Unix...) | LOGICIEL DE NAVIGATION Internet Explorer, Netscape, Opera, Mozilla... | MOTEUR de RECHERCHE FAVORI Google, Yahoo, Dmoz, Altavista, ... | OUTIL D'AIDE A LA RECHERCHE FAVORI Copernic, Agrégateur, Métamoteur, 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

- 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)

(voir NR p. 9)

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)

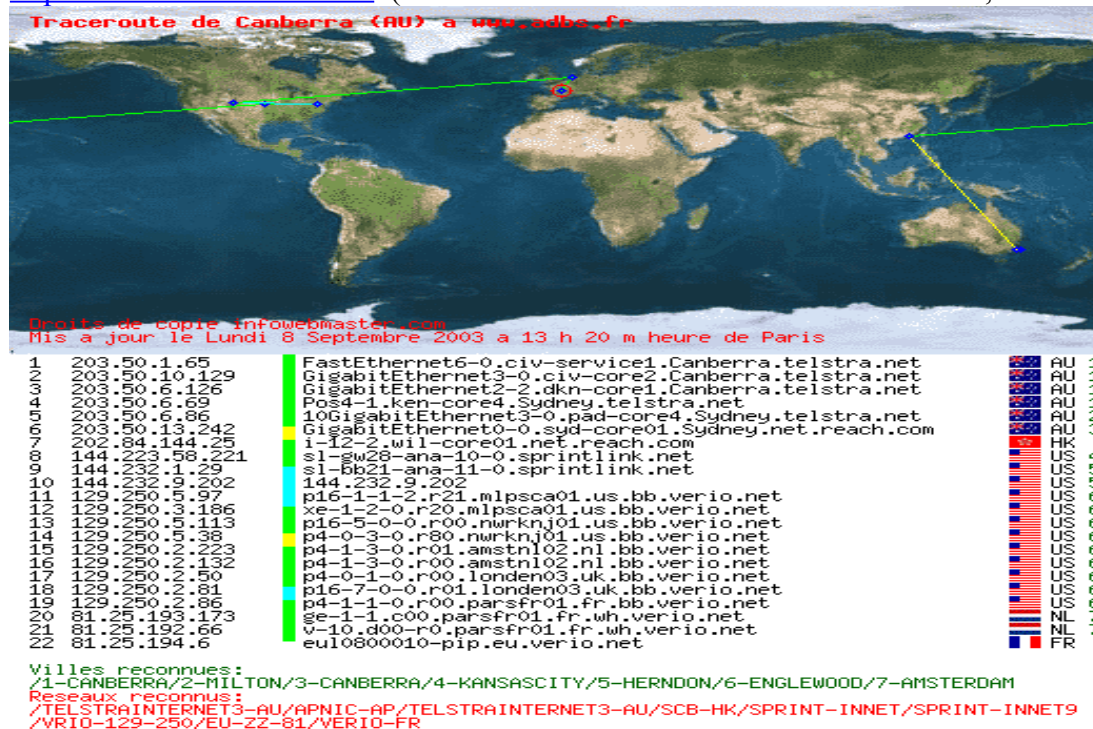
<http://www.chez.com/bibelec/links/com/internet.html> (liste de liens chez Bibelec)

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.

(voir NR p. 98)

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.

(voir NR p. 99)

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 :

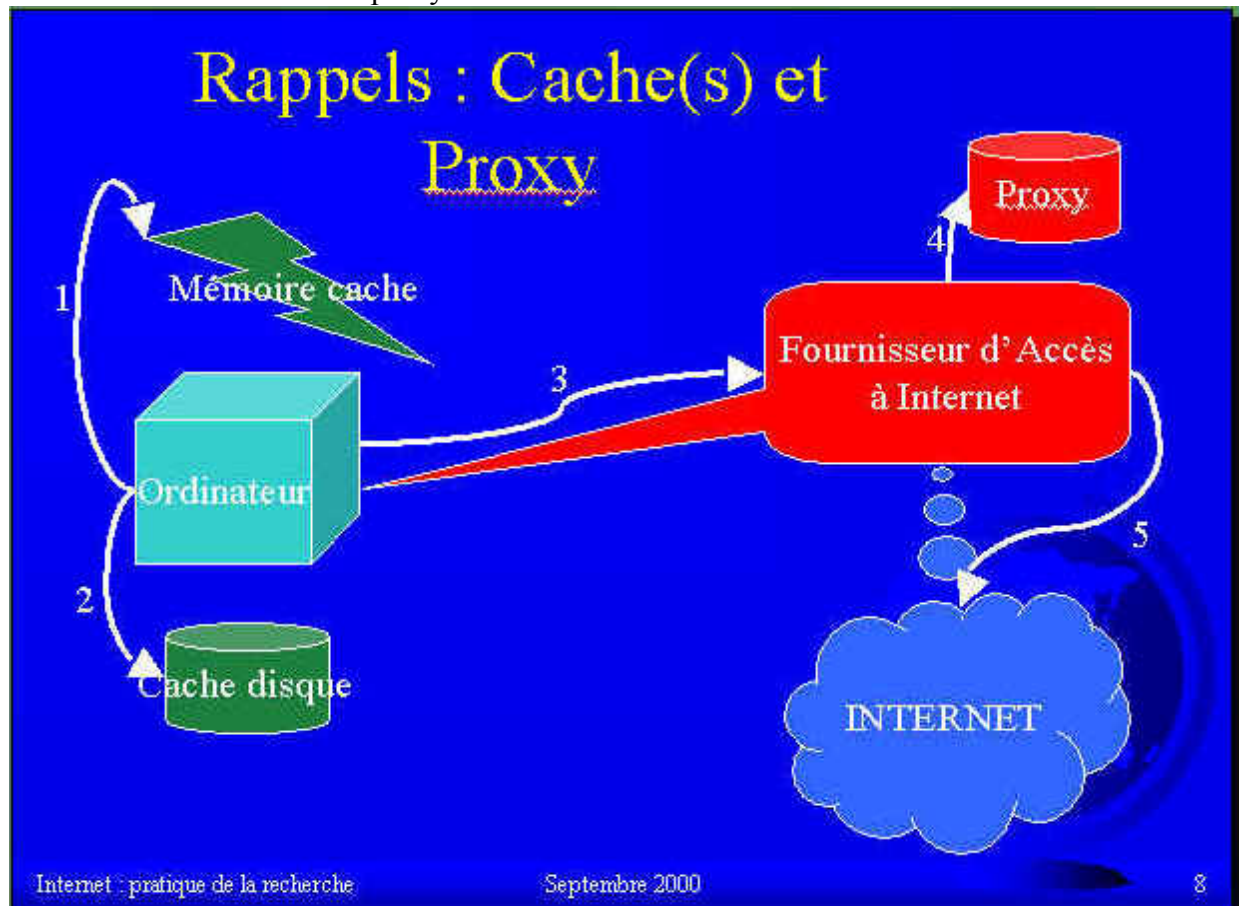
http://www.journaldunet.com/cc/03_internetmonde/intermonde_domaines.shtml

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

(voir NR p. 25)

- 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

(voir NR p. 82)

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>

(voir NR p. 23)

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 : trois moteurs se sont dégagés. 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

A ces trois moteurs (Google, Yahoo, MSN, on a tendance à ajouter un challenger français : Exalead)

(voir NR p. 31)

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>

http://del.icio.us/cottinstef/comparateur_de_moteur

(voir NR p. 87 et p. 66 : la « guerre des moteurs »)

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

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

(voir NR p. 36 : les tendances à suivre)

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) - <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 (ex-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> .

(voir NR p. 52)

Métamoteurs, bases de données en ligne, portails et sites de référence

(voir NR p. 45)

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 :

(voir NR p. 51)

nouvelle notion de "vortails" : portails verticaux

(voir NR p. 79)

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

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

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

(voir NR p. 98)

Le phénomène Weblogs et les fils RSS

(voir NR pp. 15, 19, 57, 62, 104, 105)

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.php?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.php?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.php?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.

Pour en savoir plus sur les fils rss, voir le support de cours présent sur le site : <http://www.servicedoc.info/rss>

Installer un lecteur de fil : Wticker <http://www.wticker.org>

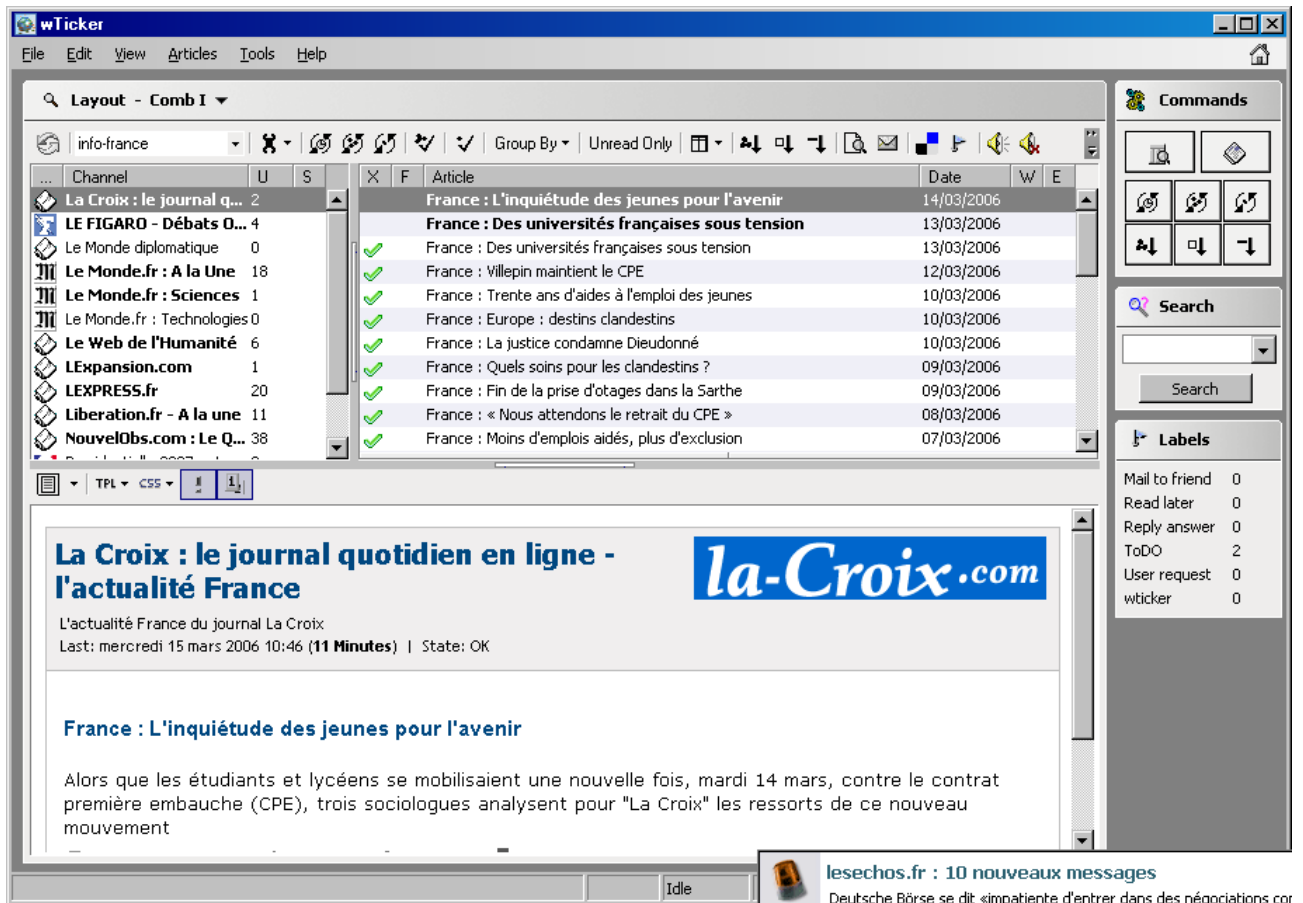
Wticker : Application : [Wticker.org](http://www.wticker.org)
Télécharger l'application sur <http://www.wticker.org> (section « Downloads »
<http://www.wticker.org/index.php?section=history>)

Mode d'emploi en ligne (en anglais : <http://www.wticker.org/index.php>)

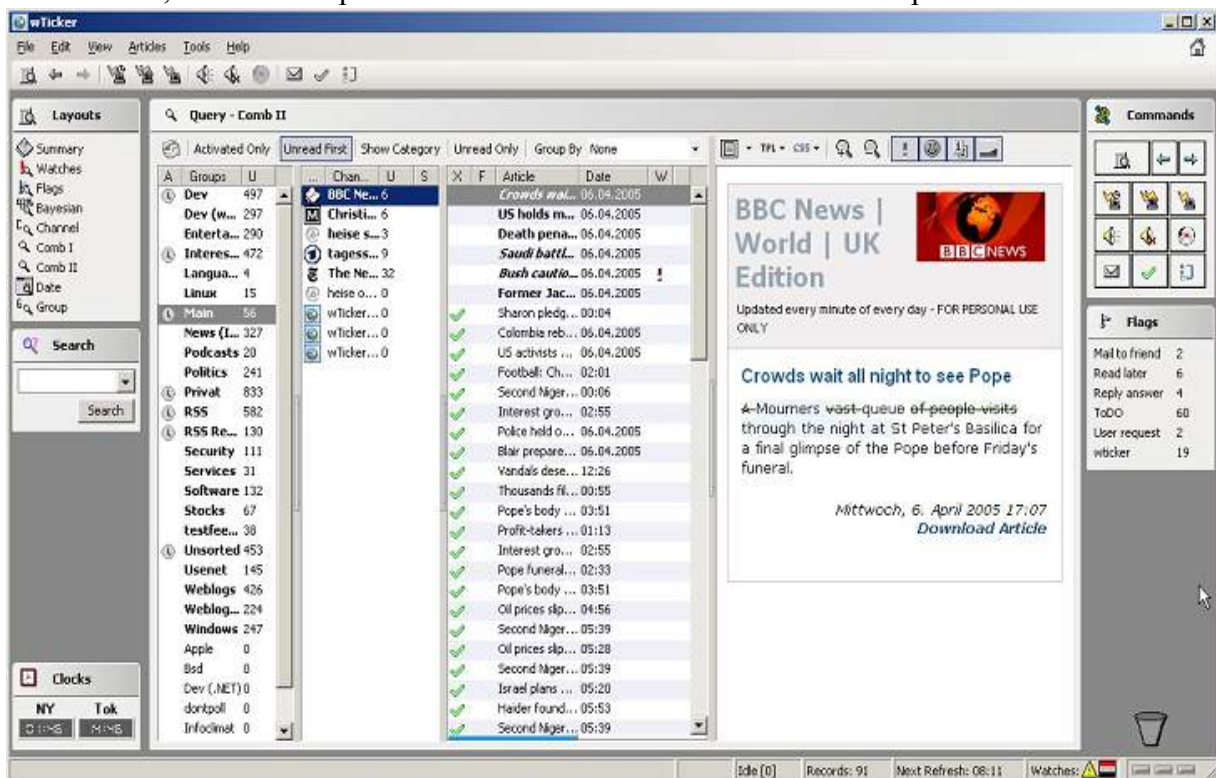


wTicker
a win32 feed aggregator
Introduction

Malgré son nom « wTicker », ce logiciel peut aussi avoir le rôle d'un simple agrégateur. Voici une reproduction de la page principale.



wTicker détermine le format de la fenêtre principale et la série d'articles affichés en exécutant une requête SQL. Cette commande SQL (cette « Vue ») est modifiable par un fichier texte et on peut en créer plusieurs. Cependant, la liste de vues par défaut devrait offrir l'ensemble des possibilités raisonnables, il ne devrait pas être nécessaire de créer d'autres « vues » personnalisées.



La série d'articles et leur format sont montrés ici dans la fenêtre de lecture (la plus à gauche). Ces formats sont modifiables par l'intermédiaire de feuilles de style en CSS et en écrivant des formats (templates) en PHP ou bien encore en XSLT. Tout ceci n'est pas nécessaire à un usage normal du logiciel et est évidemment à réserver aux utilisateurs technophiles.

Petits gadgets indispensables

Je vais décrire ci-après quelques fonctionnalités que je trouve utiles et qui ne sont pas tellement communes.

- wTicker ne récupère pas d'informations seulement des fils rss. Il est aussi possible de gérer des courriers e-mails et des nouvelles de l'Usenet. Il existe aussi un type particulier de canaux pour les cours de bourse, et sur les versions récentes de Windows, on peut aussi tracer les événements système. Une possibilité supplémentaire consiste à « scrapper » (monitorer et surveiller les modifications) des pages web.
- Une fenêtre de notification avec quelques fonctions utiles peut être activée.
- Parfois, certains éditeurs de fils modifient leurs articles. wTicker est capable de montrer seulement les différences et de les surligner.



- Si la fonctionnalité « Microsoft's speech-library » est installée (c'est le cas par défaut, mais en version anglaise), il est possible de se faire lire à haute voix tout article.
- wTicker peut tout aussi bien être utilisé dans un réseau local ou un intranet en tant que "Feed cache server".
- Les utilisateurs avancés peuvent non seulement ajouter ou modifier les commandes SQL mais aussi modifier le programme à leur guise grâce à un interpréteur de langage Pascal fourni avec le logiciel. Par exemple, la fonctionnalité qui permet de voir l'horloge est un petit script écrit en Pascal.
- Les émoticônes sont détectées et peuvent être transformées en images. Il est possible de modifier ces images et d'en ajouter.

Ainsi, à la place de :

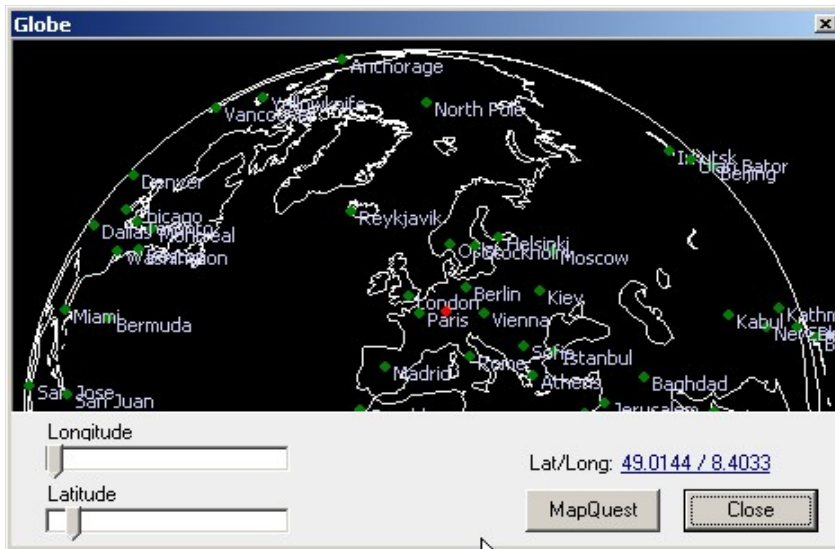
```
These functions to protect users from even  
more actions than I usually present. ;-)
```

on peut avoir :

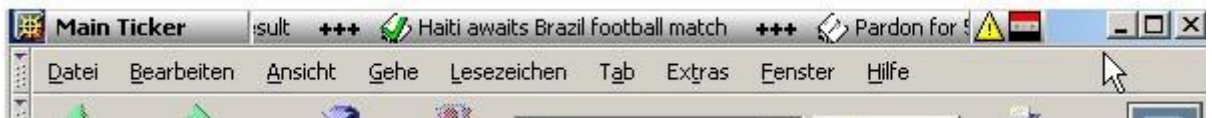
```
These functions to protect users from even  
more actions than I usually present. 😊
```

Tous les articles récupérés peuvent être exportés en CHM-Helpfile (fichier d'aide windows), ce qui est très pratique pour les lire même dans un environnement sans wTicker ni de navigateur.

- Certainement complètement inutile mais tellement beau : la possibilité, si le producteur de l'information le permet, de localiser le fil rss.



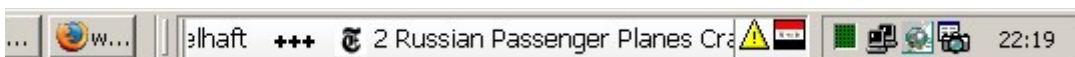
- Ce logiciel est surtout fait pour afficher des déroulants (tickers). Leurs forme et contenu sont totalement modifiables. Ci-dessous un déroulant simple et flottant. Il est positionné en haut de l'écran, laissant un espace à droite et à gauche.



Une autre option est d'avoir les déroulants affichés dans une fenêtre à part (un bandeau), placé soit en haut soit en bas de l'écran. C'est utile quand il y a plusieurs déroulants à la fois.



Il est aussi possible d'afficher un déroulant dans la barre d'outil du bureau à côté de la barre des tâches



De nombreux attributs peuvent être affectés aux déroulants : couleur, police, sens et vitesse de déroulement, contenu, etc.

Problèmes connus

Il y a bien d'autres fonctionnalités qui méritent d'être mentionnées, mais au lieu de continuer les louanges, je vais plutôt lister ici les problèmes reconnus afin d'éviter toute déception.

- Il n'y a pas d'autres langues disponibles que l'anglais pour l'instant. Le multilinguisme est déjà codé, mais cette fonctionnalité n'est pas encore ouverte.
- wTicker est en perpétuel travaux. Beaucoup de choses changent rapidement. Parfois, après une mise à jour, certaines fonctionnalités peuvent être défectueuses. Bien sûr, Christian fait tout pour l'éviter, mais si cela arrive. Soyez tolérants (ou n'essayez pas wTicker).
- Le problème le plus délicat selon le programmeur est la gestion défectueuse du multitâche. C'est un domaine qui ne peut que s'améliorer dans le futur, mais il reconnaît que cela peut être dérangent pour les personnes pressées.
- Ce n'est pas vraiment un bug, mais plutôt un choix technique = wTicker ne contient pas de navigateur Internet.

- Cet agrégateur est plutôt puissant, mais un peu complexe
(Le programmeur reconnaît ici explicitement qu'il a des « problèmes » avec la convivialité de son interface ! Mais il y travaille.)

Télécharger wTicker : <http://www.wticker.org/index.php?section=history>

Ne pas rater la révolution Folksonomy / social bookmarking

(voir NR p. 57 et suiv.)

(Tag Technorati et Del.icio.us)

<http://www.fredcavazza.net/index.php?2005/10/25/925-quelle-est-la-difference-entre-un-lien-et-un-tag>

<http://www.internetactu.net/?p=6166> : " Les services basés sur des tags (balises, ou plus précisément ici, descripteurs) censés correspondre à une indexation thématique du contenu ont fleuri sur le web. Fausse bonne idée ou évolution majeure ?"

<http://www.homo-numericus.net/blog/Les-mysteres-de-la-folksonomy.html>

<http://www.servicedoc.info/Que-mutualiser-sur-le-Web-L.html>

Que mutualiser sur le web ? L'utilisation des logiciels sociaux (Journée ADBS 11 mai 2006, FIAP, Paris)

Visualisation de la présentation powerpoint sur l'utilisation des logiciels sociaux dans un contexte professionnel de recherche documentaire

Exercice : création d'espace del.icio.us

Des méthodes et des stratégies (3)

(voir NR p. 19)

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. Choisir les bons mots-clés
4. 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.
5. Toujours analyser l'information : recouper l'information, faire preuve d'esprit critique, évaluer rapidement
6. 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.
7. 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).
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

| <i>(voir NR p. 65)</i> | |
|---|--|
| 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/adviceengine.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://reper.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.

(voir NR p. 69 et suiv.)

- 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. Choisir les bons mots-clés

(voir NR p. 73)

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://www.researchbuzz.org/2003/09/the_domain_purity_test.shtml (ma google key est 3wJB0/JQFHIjLag9z1eBxH2WVqKVTqcZ)

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

(voir NR p. 79)

- 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>)

5. Toujours analyser l'information :

(voir NR p. 91)

recouper l'information, faire preuve d'esprit critique, évaluer rapidement (connaître les principes de bases des construction 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 [nétiquette](#). (voir aussi www.arobase.org) Faire des synthèses à partager avec les autres.

6. 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)

7. 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 : garder 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.

(voir NR p. 76)

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".

Curiosité : une méthode originale de recherche = la recherche orientée "résultat"

<http://www.servicedoc.info/+Methode-la-recherche-orientee+.html>

<http://supercoach.over-blog.com/article-964137.html>

"...Les informaticiens, consciemment ou pas, ont compris cette contrainte 'logique' du questionnement : ils savent que pour trouver une information, ils ne doivent pas exprimer « ce qu'ils cherchent » mais plutôt « ce que le résultat doit être. » Ne me demandez d'où cette faculté leur vient, ce serait effrayant!
Je propose simplement de **dialoguer** avec le moteur de recherche comme s'il était un moteur de recherche. C'est **le principe de la recherche orientée résultat : apprendre à exprimer non pas ce que l'on cherche mais plutôt ce que l'on souhaite obtenir...**"

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 : Wysiogt

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 fenêtres 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)

(voir NR p. 95)

- 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

- 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/>

INFOMINDER

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

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

... ou peu onéreux :

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

<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)

(voir NR p. 71 et suiv.)

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/textesintegaux/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/jargonf/>

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 \$)

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/>
- [Search Engine Guide](http://www.searchengineguide.com) - <http://www.searchengineguide.com> -
- [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/>

Guides de recherches (généralistes)

CERISE 'Conseils aux étudiants pour une recherche d'information spécialisée efficace) : <http://www.ext.upmc.fr/urfist/cerise/index.htm> (voir en général le site de l'Urfist de Paris, école des Chartes où il est hébergé : <http://www.ext.upmc.fr/urfist/>)

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://csidoc.insa-lyon.fr/sapristi/>

- 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.

Exercices proposés :

En plus des applications pratiques le long du cours (installation de wticker ou d'autres logiciels), on pourra, au choix des stagiaires sélectionner parmi des exemples d'applications qui sont proposées dans la [boîte à outils \(partie 5\)](#) ou **NR pp. 71 à 136**

On privilégiera aussi des exercices proposés par les stagiaires (approfondissement sur un domaine en particulier...), mais en cas de carence en initiatives, voici quelques exercices qui devraient donner des idées.

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 :

- 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

5) **Exercice dirigé : création d'une liste de discussion sur Yahoo-Groupes ou Google-groups.**

Gestion de la liste et apprentissage des principales fonctionnalités

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

http://www.servicedoc.info/article.php3?id_article=188

Du bon usage du Whois : comment trouver le responsable technique d'un site web ?

mars 2004, par [Stéphane Cottin](#)

Tout nom de domaine fait l'objet d'un dépôt auprès d'un 'registrar' [1] = un prestataire assurant la gestion administrative et technique du nom, caractérisé notamment par ses dernières lettres après le dernier point.

Plus d'information se trouve auprès notamment sur le site du prestataire en charge des noms de domaine se terminant par ".fr" = l'AFNIC, sur cette page <http://www.afnic.fr/outils/whois/>

Il y est notamment permis de faire une recherche sur l'outil "Whois" de la zone ".fr", c'est-à-dire de rechercher les éléments techniques (dates de création, noms et lieux des serveurs...) et surtout administratifs (noms des responsables techniques et administratifs) du nom de domaine, pour autant qu'il se termine par ".fr" (comme "top level domain" [2])

WHOIS [3] = *Base de données, autrefois gérée par l'Internic et désormais maintenue par Network Solutions, aussi connue sous le nom de « NICname ». Elle stocke pas mal d'informations sur le réseau lui-même (adresses des sites, des entreprises, noms de domaines, classes attribuées, gestionnaires locaux...). C'est un annuaire.*

Pour les autres "top level domains", et notamment les plus courants que sont les gTLD (generic Top level domains) comme les **.com**, **.net**, **.org**, et, plus récemment les **.biz** et les **.info**, le travail de recherche est plus difficile car les bases de données ne sont pas (ou plus) unifiées. Un outil (Betterwhois <http://www.betterwhois.com>) permet de retrouver facilement les références parmi les gTld sus-cités.

Quant aux autres noms de domaines, et notamment les ccTLD (Country code Top level domains = domaines par pays), il faut chercher tld par tld. L'Afnic propose de passer par un service du 'NIC générique' assez efficace : <http://www.generic-nic.net/dyn/whois/>, mais pour vraiment tout avoir, on peut aussi passer par la liste Yahoo ! [Domain Registration > International TLD Registry Operators](#)

Exercice pratique :

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

ex : conseil-constitutionnel.fr [4]

ex : www.adbs.fr [5]

ex : servicedoc.info [6]

ex : whitehouse.gov [7]/ whitehouse.com [attention, contenu "adulte"] [8]

ex : legifrance.gouv.fr [9] / legifrance.com [10]

[1] définition © Le Jargon Français [registrar](#) / en anglais [registrar](#)

[2] [Top level domain selon le 'Jargon Français'](#)

[3] [définition Jargon Français](#)

[4] Solution : rechercher 'conseil-constitutionnel.fr' dans <http://www.afnic.fr/outils/whois>

[5] Solution (idem) : rechercher 'adbs.fr' dans <http://www.afnic.fr/outils/whois>

[6] Plusieurs solutions : via [Betterwhois](#), via [Gandi](#) ...

[7] pour les 'dot Gov', l'officiel <http://www.nic.gov/whois.html> ne vous donnera pas plus que le fait que le site existe ou a été réservé

[8] Whitehouse.com via [Betterwhois](#)

[9] legifrance.gouv.fr à rechercher via <http://www.afnic.fr/outils/whois>

[10] legifrance.com via [Gandi](#)

Complément

Attention : un Whois est dépendant des informations qui lui sont fournies par le responsable technique.

Or :

- ce responsable technique est souvent un intermédiaire professionnel et non le webmestre réel du site, c'est-à-dire la personne, organisme ou société qui le met à jour. En général, il y a deux catégories d'intermédiaires de ce genre :
 - un technicien du fournisseur de nom de domaine
 - carrément, une société dont le métier est d'enregistrer des noms de domaine pour d'autres. Exemple sur le [Whois de Tucows](#), le nom de domaine [ballade.com](#) donne comme nom de responsable administratif et technique Net City Ltd, adresse à Hong-Kong. Peu parlant. Et cette adresse est-elle exacte ? Nous allons voir ci-dessous comment la vérifier
- les informations fournies au registrant sont de fait rarement mises à jour, malgré le risque de perte de contrôle du nom de domaine. Par exemple, si l'e-mail donné n'est plus en service, il peut être très difficile de reprendre le contrôle en cas de piratage ("hack") du nom de domaine par une personne non autorisée), certains changements nécessitant de répondre à un e-mail de confirmation
- surtout, pour se protéger des spammeurs et indiscrets, les informations données, jusqu'au nom, sont souvent très maigres, partiellement fausses ou encore renvoient sur une personne qui n'est pas le vrai responsable. Par exemple, mettre son vrai numéro de téléphone ou son e-mail principal est une garantie d'ennuis à n'en plus finir. En effet, les données à fournir comprennent nom, adresse, téléphone et e-mail... (cf Internet Privacy Advocate > Protect Your Personal Information > [What Information Is Made Public ?](#))
Les registrants protègent maintenant beaucoup mieux leurs bases de données Whois des téléchargements massifs, qui sont le plus souvent le fait de pirates ("hackers"). Ce risque a donc nettement diminué, mais il existe toujours et la plupart des gens qui procèdent aux enregistrements de noms de domaine sont maintenant devenus très prudents.

Conclusion : les informations des Whois sont à prendre avec des pincettes et donc à vérifier et recouper par d'autres moyens, si possible.

Quelques parades aux limites des Whois :

- cherchez, sur le plan du site, la page Contact, celle où se trouve l'adresse e-mail du webmestre et éventuellement l'adresse physique de l'organisme auteur du contenu du site. Pour reprendre l'exemple de [ballade.com](#), la page d'accueil de ce site donne une adresse e-mail pour contacter le "propriétaire du nom de domaine". Or cette adresse est une adresse qui se termine, après le @, par [allthenet.com](#). En allant sur le site web correspondant <http://www.allthenet.com>, on se rend compte qu'il s'agit d'un annuaire de sites web très commercial/publicitaire
- les moteurs de recherche ([Google](#) ou [Yahoo Search](#), surtout) permettent également de trouver des commentaires sur le site, incluant les noms "en clair" de ses auteurs réels ou bien au moins quelques indications (pays, etc.) ou pistes. Ainsi, une [interrogation de Google](#) par "Net City Limited" permet de savoir qu'il s'agit d'une société de Hong-Kong, grâce à la base de la Securities and Futures Commission (SFC) de Hong-Kong (voir <http://www.hksfc.org.hk/eng/takeove...>)
- les bases de données d'informations sur les sociétés (RCS, organismes de régulation boursiers) sont également un bon moyen de vérifier les identités des responsables de site et d'en apprendre plus sur eux. A condition que la société ait pignon sur rue. C'est très efficace si le pays possède un registre du commerce interrogeable sur le Web, gratuitement (exemple : en France, par [Societe.com](#)) ou bien si la société est cotée en bourse ou bien a été l'objet d'une acquisition ou vente (cf notre exemple supra avec la SFC de Hong-Kong).

Emmanuel Barthe

2) Utilitaires pratiques :

- Installer et utiliser une toolbar (googlebar, yahoo, mozbotbar..., ou NeedleSearch pour les utilisateurs de Mozilla/Firefox)
- Installer un utilitaire "anti-pop-up"

PB = la plupart des manipulations nécessitent ici des droits d'administrateurs sur son poste

Différences entre les toolbars

- Une toolbar universelle = NeedleSearch (spécial Mozilla/Firefox) = <http://needlesearch.mozdev.org/>

- Google Deskbar / Toolbar = <http://toolbar.google.com/> <http://deskbar.google.com/>

Plus loin que les toolbar = le "desktop searching" = mettez un moteur de recherche dans votre PC
Google Desktop Search = <http://desktop.google.com>

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

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

La république du Web

La croissance du Web se poursuit à une vitesse stupéfiante. Les moteurs de recherche automatisés ont de plus en plus de difficultés à fournir des résultats satisfaisants. Les petites équipes d'édition professionnelles travaillant sur les sites commerciaux des répertoires ne peuvent plus répondre aux requêtes, et la qualité et le contenu de leurs répertoires s'en ressentent. Les liens deviennent obsolètes et ne peuvent plus suivre le rythme de croissance de l'Internet. Au lieu de combattre la croissance explosive de l'Internet, le projet Open Directory permet à l'Internet de s'organiser. Parallèlement à la croissance de l'Internet, le nombre de citoyens du Web augmente. Ces citoyens sont habilités à organiser une petite partie du Web et à la présenter au reste de la population, en supprimant les éléments inutiles ou inintéressants, et en conservant les meilleurs éléments.

« Le » catalogue du Web

L'Open Directory marche dans les pas de certains des plus importants projets collaboratifs du 20e siècle. Tout comme l'Oxford English Dictionary a eu le dernier mot en matière de mots de la langue anglaise grâce aux efforts de bénévoles, l'Open Directory suit ses traces afin de devenir « le » catalogue du Web.

L'Open Directory a été fondé dans l'esprit du mouvement Open Source et est le seul répertoire d'importance majeure à être totalement gratuit. Il n'y a rien et il n'y aura jamais rien à déboursier pour y soumettre un site ni pour utiliser ses données. Elles sont gratuitement à la disposition de quiconque accepte de respecter notre "free use license".

Le cerveau de l'Internet

L'Open Directory est la plus diffusée des bases de données sur le contenu du Web catalogué par des êtres humains. Son groupe d'éditeurs citoyens du Net forme le cerveau collectif qui alimente la quête de ressources présentes sur le Web. L'Open Directory fournit leurs répertoires aux plus grands et aux plus populaires des moteurs de recherches et des portails.

Vous avez le moyen de changer les choses

Comme dans toute communauté, vous échangez et récoltez ce que vous avez semé. L'Open Directory permet à tout le monde de participer.

L'inscription est facile : choisissez un sujet que vous maîtrisez et inscrivez-vous. L'édition des catégories est simple comme bonjour. Nous disposons d'un ensemble d'outils complet permettant d'ajouter, de supprimer et de mettre à jour des liens en quelques secondes. En quelques minutes, vous pouvez participer à l'amélioration du Web et être reconnu en tant qu'expert du sujet choisi.

Rejoignez l'Open Directory Project

1. Recherchez une catégorie que vous aimeriez entretenir.
2. Suivez les instructions du lien **Devenir éditeur** situé en haut de la page de catégorie.

Remarquez que les pages de catégorie de niveau supérieur ne disposent pas de liens **Devenir éditeur** ; vous devez rechercher une catégorie plus précise et faire une demande pour cette catégorie.

4) 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

http://www.servicedoc.info/article.php3?id_article=242

Trouver des thèses en ligne

juillet 2004, par [Stéphane Cottin](#)

Généralités

La super [liste de ressources générales de l'ABES](#), agence bibliographique de l'enseignement supérieur renvoie vers : <http://www.abes.fr/abes/DesktopDefault.aspx?tabindex=7&tabid=17>
[La page Thèses](#), <http://www.abes.fr/abes/DesktopDefault.aspx?tabindex=10&tabid=223>

regroupant des liens vers Textes réglementaires ; Guide du doctorant ; Dépôt papier ; Dépôt électronique ; Thèses électroniques ; [SPARTE](#) ; Rechercher une thèse ; Quelques questions ; Autres sites

et notamment vers la [page de liens "Autres sites"](#)

[Fichier central des thèses](#) ("en cours") = Nanterre <http://fct.u-paris10.fr/>

International

[google it ! \(related :edd.vt.edu\)](#)

[Site du CREDO](#), (apparemment page perso de K. Ben Kahla)

<http://iquebec.ifrance.com/credo/thesesenligne.html>

[electronic theses and dissertations \(ETDs\) in humanities](#)

<http://etext.lib.virginia.edu/ETD/ETD.html>

[Portail des ETD](#) <http://etd.vt.edu/>

[EtdGuide](#)

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

The [Networked Digital Library of Theses and Dissertations \(NDLTD\)](#) is an international organization dedicated to promoting the adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper-based theses and dissertations.

This website contains information about the initiative, how to set up Electronic Thesis and Dissertation (ETD) programmes, how to create and locate ETDs, and current research in digital libraries related to NDLTD and ETDs.

Sciences de l'information

[MemSic](#) <http://memsic.ccsd.cnrs.fr/>

Memsic est une partie de [ArchiveSic](#) équivalent de ArXiv pour les articles dans le même domaine
Extraits de "[l'A-propos](#)"->]

Le site du CCSD @[rchiveSIC](#) a pour objectif de faire découvrir l'auto-archivage en ligne de documents et de montrer la souplesse et l'efficacité de cet outil.

Le concept d'auto-archivage, dont le plus ardent défenseur est Stevan Harnad, a pour but de permettre [l'inéluctable évolution](#) des revues érudites vers la solution optimale pour les étudiants et les chercheurs. Résoudre une [anomalie post-gutenbergienne](#).

Ce site a été configuré pour archiver des articles et des working papers dans le domaine des Sciences de l'Information et de la Communication (SIC) de manière à les rendre disponibles gratuitement à tous via le web et par la même augmenter leurs visibilité. Chaque auteur (ou organisme dont il dépend) peut librement déposer ses documents sur ce site après [s'être enregistré](#). Chaque document est examiné par un [modérateur de domaine](#) qui pourra l'accepter ou non. Une [bibliographie sur @rchiveSIC](#) est disponible.

Les responsables du serveur sont :

➡ Gabriel Gallezot

➡ Ghislaine Chartron

➡ Jean-Max Noyer

Multidisciplinaire

[CCSD : thèses-EN-ligne. Serveur de Thèses Multidisciplinaire](http://tel.ccsd.cnrs.fr/) <http://tel.ccsd.cnrs.fr/>

Voir *A propos*->http://tel.ccsd.cnrs.fr/information_fr.html] et [l'aide en ligne](#)

Surtout des thèses de mathématiques, d'informatique, un peu de sciences humaines
extrait de "[l'A-propos](#)" :

Le site du [CCSD thèses-EN-ligne](#) a pour objectif de promouvoir l'auto-archivage en ligne des thèses, qui sont des documents importants pour la communication scientifique entre chercheurs ; il a été développé en collaboration avec la Cellule [MathDoc](#). Son but est de compléter le serveur [ArXiv](#), qui comprend surtout des documents destinés à être publiés comme articles, mais peu de thèses, avec un format de fichiers peu adapté à ces dernières (qui n'a pas été conservé dans thèses-EN-ligne) ; dans le même esprit qu'ArXiv, il permet de rendre aisément disponibles à tous, gratuitement, thèses de doctorat et habilitations. Il complète également [l'index national des Thèses et Habilitations de Mathématiques en France de la cellule Mathdoc](#). Le CCSD comme MathDoc n'effectue aucune évaluation scientifique sur les thèses ou habilitations déposées, car c'est le rôle du jury.