


Compte rendu du Club IES

Panorama des outils gratuits de veille et de recherche d'information



Conférence de **Renaud Finaz de Villaine**
Directeur marketing-communication de Micropole Univers
au Club IES - AAE IAE
lundi 9 Mars 2009, à l'IAE de Paris

- 
-
- Ce compte rendu reprend une partie des diapositives de la présentation de M. Finaz de Villaine, et sont diffusés avec son accord.

 - Sommaire
 1. Les caractéristiques du web
 2. Les technologies du Web 2.0
 3. Quelques annuaires & Moteurs de recherche
 4. Quelques outils complémentaires
 5. Quelques moteurs de recherche innovants à tester

1- Les caractéristiques du web



La masse d'informations disponible est en expansion exponentielle

L'Explosion de l'Information 2015: 280Md

- Croissance de l'Information = 1% par semaine
- Mais pas la capacité à la digérer!!

2010: 130Md

2008: 98Md

2006: 68Md

2004: 48Md

2003: 24Md

2002: 12Md

2001: 6Md

2001: 3Md

1999

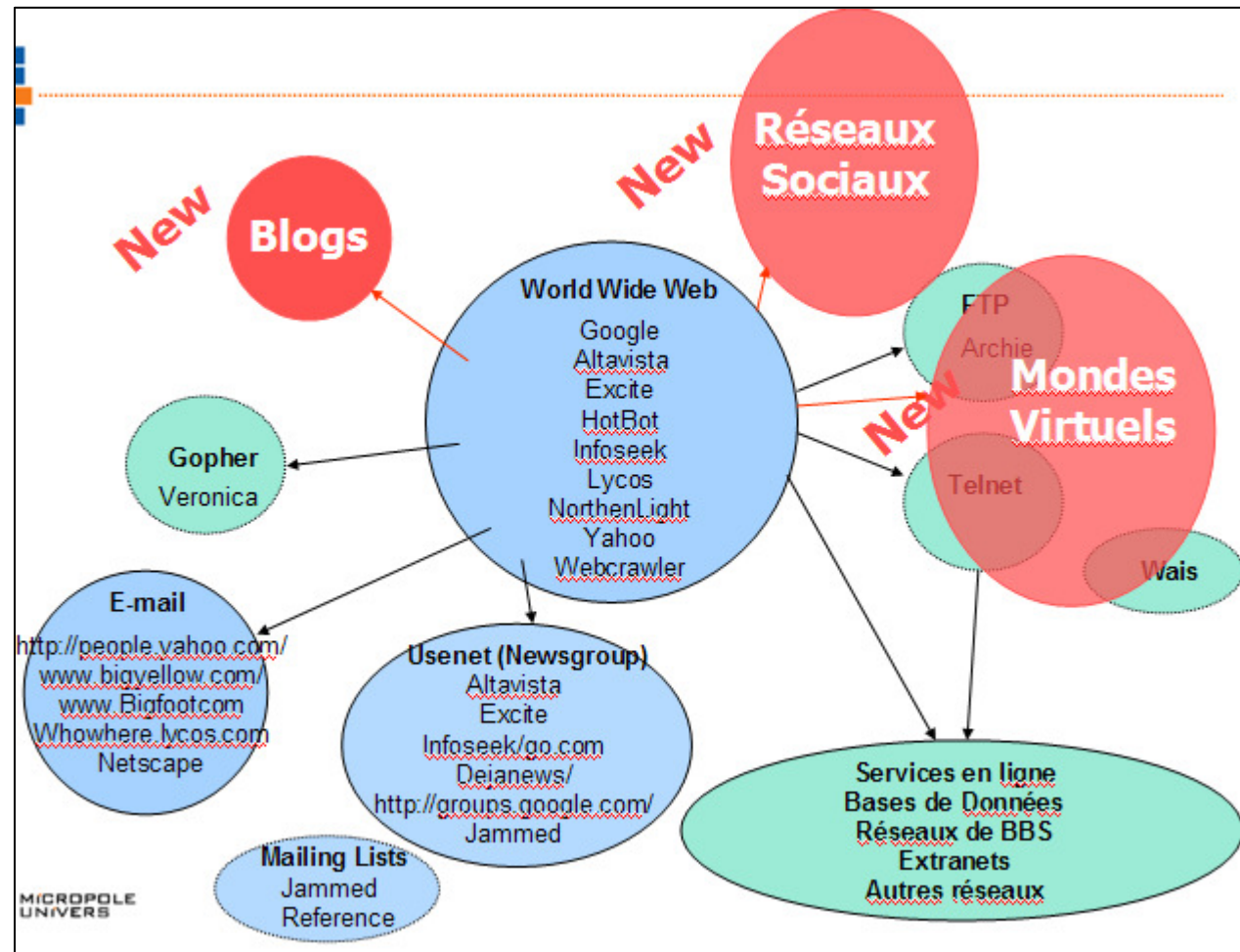
105
Milliard de
Gigabytes




GIGABYTES
Source: Dr. Henkely, School of Information Management and Systems

Le web génère constamment de nouveaux outils :

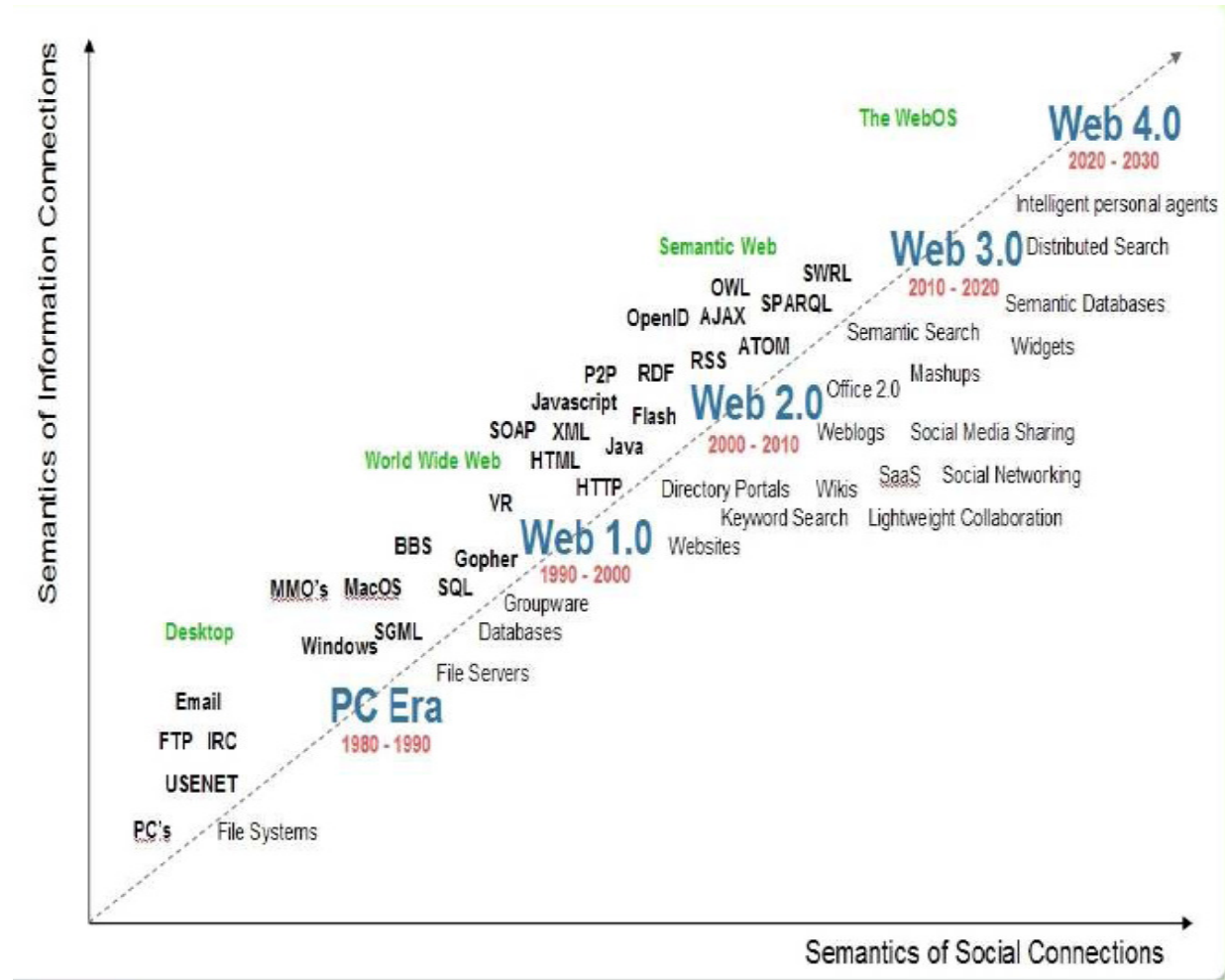
- Les blogs
- Les réseaux sociaux
- Les mondes virtuels
- Les outils de syndication



- 
- Le web génère de nouvelles habitudes
 - On a pu notamment constater les faits suivants. Au cours du mois d'août 2008 :
 - 26 millions d'internautes français ont effectué un total de 2,5 milliards de recherches.
 - Ils ont effectué une moyenne de 96 recherches par personne au cours du mois, soit 3,1 recherches par jour
 - Ils ont effectué un total combiné de 620 milliards de visites au cours du mois, avec une moyenne de 4,1 recherches par visite.
 - 2.2 millions d'abonnés Français en téléphonie mobile ont utilisé leur téléphone portable pour effectuer au moins une recherche au moyen de leur navigateur, sur la base d'une moyenne de trois mois se terminant en août 2008.



Nous sommes passés du web 1.0, au web 2.0, mais qui sera lui-même bientôt remplacé ...



GO2WEB20.net, Répertoire du Web 2.0

Le site
go2web20.net
recense les outils
du web 2.0



2- Les technologies du Web 2.0

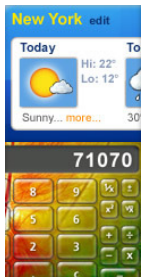


Mashups, Widgets, RSS & Vidéo

■ Mashups

- Un **mashup** en informatique est un site internet ou une application dont le contenu provient de la **combinaison de plusieurs sources** d'information.
- De plus en plus d'éditeurs de contenu proposent gratuitement des API, afin d'encourager la communauté des développeurs à créer des mashup utilisant leur contenu. Les applications composites (mashup en anglais) se basent sur la programmation événementiel

■ Widgets



- Recouvre deux notions distinctes en relation avec les interfaces graphiques, contraction des termes window et gadget.
 - un composant d'interface graphique, un élément de base d'une interface graphique (bouton, ascenseur, liste déroulante, etc...)
 - un widget de bureau, un petit outil qui permet d'obtenir des informations (tableaux de bord, météo, actualité, carte, pense-bête etc...)

■ RSS

- Objectif : informer l'internaute de la survenue d'une nouveauté
- Permet à l'internaute de suivre simultanément l'actualité de plusieurs sites web ou applications web et d'être alerté à chaque fois qu'une nouvelle information y est publiée

■ Vidéo

- Web TV pour communication interne ou externe, e-learning, marketing

3- Quelques annuaires & moteurs de recherche

- *Solution personnelle*
- *Solution d'entreprise*



INTERNET SEARCH ENGINES ROADMAP

PERSONAL TOOLS

CORPORATE SOLUTIONS

CHARGEABLE OFF LINE SEARCH ENGINE

Off line engines to install on your PC/Mac

BULLSEYE	http://www.intelliseek.com/be/bullseye.htm/
COPERNIC 2000	http://www.copernic.com/
ECHOSEARCH	http://www.iconovex.com/
INFORIAN QUEST 99	http://www.lead.com/webutilities/quest99
MACROBOT	http://www.ipgroup.com/Macrobot/
QUERY METASEARCH	http://www.freeflowsoftware.com/
RETRIEVE IT	http://www.mpsolutions.com/
WEB WOLF	http://www.msw.com.au/
WEB FERRET PRO and INFO FERRET	http://www.bluesquirrel.com/
WEBSEEKER	http://www.bluesquirrel.com/
WEBSLEUTH 98	http://www.prontosoft.com/

Web site recorders

WEB VCR	http://www.netresults.com/
WEBSNAKE	http://www.arawave.com/
FLASHSITE	http://www.incontext.com/
MEMOWEB	http://www.memoweb.com/
NETATTACHE PRO	http://www.tympari.com/
TELEPORT PRO	http://www.termax.com/
WEBWHACKER	http://www.bluesquirrel.com/
GRABASITE	http://www.bluesquirrel.com/

COMPETITIVE INTELLIGENCE

Research software

INFORMATION MINER 4U	http://www.arisem.com/
RETRIEVAL WARE	http://www.excalibur.com/
APERTO LIBRO	http://www.inforama.com/
FULCRUM	http://www.pdocs.com/
INFORMATION SERVER	http://www.verity.com/

Concept mapping and visualization tools

SAMPLER	http://www.cia.fr/
THEME SCAPE	http://www.caria.com/
KARTOGRAPH	http://www.loks.fr/
MAPINFO	http://www.mapinfo.com/
SEMIO MAP	http://www.semio.com/
UMAP	http://www.umap.com/
HYPERBOIC TREE	http://www.insight.com/

Analysis softwares

TROPES	http://www.aoclic.fr/
DATAVIEW	http://www.crm.univ-mrs.fr/
PERICLES	http://www.datops.com/
V-STRAT	http://www.digimind.fr/
TETRALOGIE	http://www.atlas.vit.fr/
NTK.FOCUS	http://www.nemesia.com/
WINCITE	http://www.wincite.com/

International search engines

ALTAVISTA	http://www.altavista.com/
AOL NETFIND	http://www.aol.com/efnd/
ASK JEEVES	http://www.askjeeves.com/
BEAUCOUP	http://www.beaucoup.com/
FAST	http://www.alltheweb.com/
EINET GALAXY	http://www.einet.net/
EUROSEEK	http://www.euroseek.com/
EXCITE	http://www.excite.com/
GOOGLE	http://www.google.com/
GOTO	http://www.goto.com/
HOTBOT	http://www.hotbot.com/
INFOSEEK	http://www.infoseek.com/
LYCOS	http://www.lycos.com/
NORTHERN LIGHT	http://www.northernlight.com/
VOILA	http://www.voila.com/
WEBCRAWLER	http://webcrawler.com/

FREE TOOLS ON LINE

Directories

ABOUT	http://www.about.com/
BEAUCOUP	http://www.beaucoup.com/
INDEXA	http://www.indexa.fr/
LOOKSMART	http://www.looksmart.com/
MAGELLAN	http://magellan.excite.com/
NOMADE	http://www.nomade.fr/
OPENTEXT	http://www.pinstripe.opentext.com/
SNAP	http://www.snap.com/
TROUVEZ	http://www.trouvez.com/
WANADOO - QUIQUOUCU	http://www.wanadoo.fr/
YAHOO	http://www.yahoo.com/
YAHOO France	http://www.yahoo.fr/

Metasearch engines

A1 DIGISEARCH	http://www.digisway.com/digisearch/
ALL 4 ONE	http://www.all4one.com/
CYBER 411	http://www.cyber411.com/
DEBRIEFING	http://www.debriefing.com/
DOGPILE	http://www.dogpile.com/
HIGHWAY 61	http://www.highway61.com/
MEGAWEB	http://www.come.to/megaweb/
METACRAWLER	http://www.metacrawler.com/
METASEARCH	http://www.metasearch.com/
MOTHER OF ALL SEARCH ENGINES	http://www.momma.com/
ONESEEK.COM	http://www.oneseek.com/
PROFUSION	http://www.profusion.com/
SAVVYSEARCH	http://www.savvysearch.com/

E-commerce search engines

BARGAIN FINDER	http://bl.cstar.ac.com/bl/
DEALPILOT	http://www.dealplot.com/
K-COMMERCE	http://www.infrence.com/
INSWEB	http://www.insweb.com/
PRICESCAN	http://www.pricescan.com/
PRICEMATCH	http://www.pricematch.com/
ROBOSHOPPER	http://www.roboshopper.com/
SHOPPING EXPLORER	http://www.shoppingexplorer.com/
WEBMARKET	http://www.webmarket.com/

French-speaking engines

ECILA	http://www.ecila.fr/
EXCITE FRANCE	http://www.excite.fr/
FRANCITE	http://www.francite.fr/
INFOSEEK FRANCE	http://www.infoseek.com/fr/
LOKACE	http://www.lokace.com/
LYCOS FRANCE	http://www.lycos.fr/
NETSCAN	http://www.netscan.fr/
VOILA	http://www.voila.fr/

KNOWLEDGE MANAGEMENT

Information personalisation (Push)

PERSONIFY ESSENTIALS	http://www.allnicast.com/
BACKWEB	http://www.backweb.com/
DATA CHANNEL RIO	http://www.datachannel.com/
DESKTOP NEWS	http://www.desktopnews.com/
CASTANET TUNER & TRANSMITTER	http://www.marimba.com/
TIB ACTIVE ENTERPRISE	http://www.tbco.com/
INTERMIND	http://www.intermind.com/
INTERNET EXPLORER 4	http://www.microsoft.com/
NETCASTER	http://www.netscape.com/
ENTRYPOINT	http://www.entrypoint.com/

Corporate portal

CLASS 4U and CONSULT 4U	http://www.arisem.com/
KNOWLEDGE SERVER	http://www.cubantony.com/
DATABASE	http://www.database.com/
INTASPECT	http://www.intaspect.com/
LEVELINK	http://www.opentext.com/
DOMINO	http://www.loks.fr/
NET PORTAL	http://www.medtopps.com/
TOP TIER PORTAL	http://www.topinfos.com/
KNOWLEDGE ORGANIZER	http://www.verity.com/

USEFUL TOOLS

Synthesis and translation

BABYLON TRANSLATOR	http://www.babylon.com/
POWER TRANSLATOR PRO	http://www.lhl.com/
REVERSO PRO	http://www.softissimo.com/
SUMMARY SERVER	http://www.insight.com/
SYSTRAN ENTERPRISE	http://www.systransoft.com/

Linguistic and semiotic analysis

MORE SENSE 4U	http://www.arisem.com/
INQUIZIT	http://www.inquizit.com/
INTUITION	http://www.coraso.org/
NEURO TEXT	http://www.grimmersoft.com/
INTELLIGENT MINER FOR TEXT	http://www.software.ibm.com/
LEXQUEST	http://www.lexquest.com/
ILICO	http://prologianet.univ-mrs.fr/
SPIRIT	http://www.technologies.gtd.com/
TROPES	http://www.aoclic.fr/



- Le crawling ou spidering consiste en la récupération d'un ensemble de pages Web du réseau Internet.
- L'agent de crawling part d'une adresse, enregistre la page correspondante et parcourt les URL référencées dans la page. Il recommence à partir des nouvelles pages récupérées.
- L'agent de crawling saute ainsi de sites en sites, il peut se retrouver sur une page qu'il à déjà visitée. Il conserve en mémoire la liste des pages vues, aussi, il ne récupère pas deux fois une même page.

- L'indexation consiste à extraire la liste des mots et phrases des pages récupérées par l'agent de crawling et à construire une table appelée index référençant ces mots et phrases.
- On crée ainsi une table des mots et phrases avec comme paramètre pour chacun d'entre eux une adresse de page et également un index lié à ces données.
- L'intérêt de construire un index est d'accélérer la recherche d'un mot dans la table de données correspondante.

- L'interface de recherche par mots clés est accessible à l'utilisateur final. Le ou les mots clés qu'il saisit génèrent ensuite une requête d'interrogation sur l'index de la base de données construit préalablement.
- Le résultat de la requête est un ensemble d'URL qui sont les liens vers les documents qui correspondent à la requête de recherche. Cette liste de liens est retournée à l'utilisateur qui peut ensuite visualiser les documents associés.

Moteurs de recherche

- Pas de frame, de carte image, de mot de passe ou de page dynamique (réécriture d'URL)
- Placement des mots clés :
 - dans le titre, les Meta Tag
 - en début ou dans le de document
 - Balises dans le body, H1 ou H2
- Fréquence des mots clés
- Des liens extérieurs
- Des mises à jour fréquentes
- Référencer la page principale
- Respect des normes d'accessibilité
 - Google agit comme un mal voyant
- Les sites à fort contenu (20 / 30 000 pages)
- Création d'un fichier XML Google site map (description de ttes les pages du site : adresse, sa priorité de 1 à 0,1, date de création & fréquence de mise à jour de page)

Référencement

<HEAD> Charset : qui défini le jeu de caractères

*<TITLE>Business & Decision, leader en Data Warehouse, bien choisir ses mots ! **Jamais plus de 10 mots, maxi 2 occurrences du mot clé***

<META name="description"

*content= " Business & Decision est le leader reconnu du Data Warehouse en France">, **pas plus de 256 caractères, répéter le mot clé ciblé dans le titre !***

<META name="keywords"

*content= " Business & Decision, Data Warehouse, Data Mining, décisionnel, Internet, SSII, société de service, conseil"> **Ne sert plus à rien !***

Quelques moteurs de recherche & annuaires

- Moteurs de recherche US/UK



- Moteurs de recherche d'origine française



- Annuaires de recherche US/UK/Fr



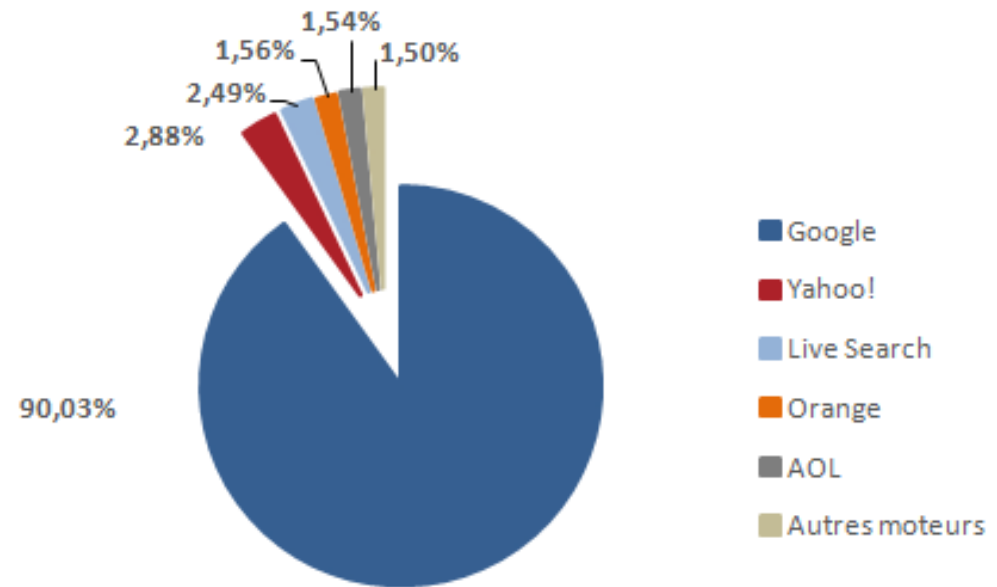
Le taux de mortalité des moteurs est assez élevé !

Top 5 des Moteurs de Recherche

Une domination sans partage de Google

Janvier 2008

TOP 5 des moteurs de recherche en parts de visites
Janvier 2008





Google est omniprésent

- Google lance :
 - un tableur, <http://spreadsheets.google.com>
 - un traitement de texte Writely, www.writely.com
 - un agenda, www.google.com/calendar
 - et un éditeur de pages web et gratuits, <http://pages.google.com>
- Google Pack : des programmes à télécharger, <http://pack.google.com>
 - Navigateur Firefox
 - Gestionnaire d'images Picasa
 - Antivirus Norton
 - Le lecteur de document PDF Acrobat Reader
 - La barre d'outils Google
 - Le moteur de recherche locale Googlr Desktop

4- Quelques outils complément aires



Meta Search ou Meta Crawlers

- Copernic v6.0 www.copernic.com (\$79.95)
- BullsEye (www.intelliseek.com)
- LexiBot
- WebFerret (\$29.95) www.ferretsoft.com
- Metacrawler (Google · Yahoo! · Ask · Jeeves · About · LookSmart · Overture, FindWhat)
- Dogpile
- SavvySearch
- ProFusion (Intelliseek)
- Metasearch
- Strategic Finder (www.strategicfinder.com)
- Webseeker (www.buesquirrel.com)
- SurfWax
- WebWolf ou SearchWolf www.trellian.com/webwolf
- Internet Search ActiveX www.lencom.com
- WebHuntPro www.wintops.com/webhunt
- EasySeeker (www.buesquirrel.com)
- Web Info Finder www.infogist.com/webinfo
- FirstStop WebSearch www.firststopwebsearch.com
- The Sleuthhound www.isleuthhound.com
- WebFoil www.webfoil.iwarp.com



Les aspirateurs de sites

- Veb VCR2.0 www.netresultscorp.com
- Memo Web4 (www.goto.fr)
- Teleport Pro www.tenmax.com
- WebWhacker5.0 & Grab A Site de BlueSquirrel
www.bluesquirrel.com
- eCatch (www.ecatch.com)
- HTTrack Open Source www.httrack.com
- MemoWeb de Goto Software
- Teleport Pro de Tenmax
- WebCopier Pro de Maximun Soft
- Wysigot www.wysogot.com/fr

Moteurs de recherche de News

- Altavista News, <http://news.altavista.com> (utilise Moreover), disparition des archives de plus d'un an
- AllTheWeb.com, www.alltheweb.com/?cat=news, articles conservés 1 semaine
- Yahoo News, <http://news.yahoo.com>, articles conservés 1 mois, 7 000 sources dans 35 langues, service d'alerte par e-mail
- Daypop, www.daypop.com
- Moreover, www.moreover.com/news/ (plus de 1 000 sources, e-mail quotidien ou hebdomadaire)
- MagPortal.com
- FindArticles.com
- News.Google.fr, conserve les articles sur 1 mois, surveille 4 500 sources anglophones, 500 Fr, 700 Ge, 250 It, = 6650 sources d'actualité, service d'alerte par e-mail



Outils pour constituer une revue de presse

- My Yahoo ou Biz <http://biz.yahoo.com/secteur>
- My NewsPages
- Excite news tracker, <http://news.excite.com>
- Infoseek News : www.infoseek.com/news
- Individual.com (ex NewsPage.com, puis Office.com)
- Net2One (www.net2one.fr)
- Moreover (www.moreover.com)
- Companynews.fr
- Prline.com



Différents Agents

- Agents de Veille offline
 - Webspector (www.illumix.com)
 - C4U (www.c4u.com)
- Agents d'alerte par e-mail
 - The Informant ([//informant.dartmouth.edu](http://informant.dartmouth.edu))
 - Mind-it de Net Mind
- Agents de Push offline
 - Infogate (www.infogate.com, ex PointCast)
 - WordFlash
- Pour surveiller quelques pages web
 - The Easy Bee www.theeasybee.com



Moteurs de surveillance monopostes


Logiciels permettant de mettre sous contrôle tout ou partie de sites internet, de forums, de listes de diffusion.


- Arisem* Kaliwatch Professional
- Copernic Tracker
- KBCrawl* de BEA Conseil
- KeyWatch* de Iscope
- Vigilius de Pragtec
- WebSite Watcher de Aignes
- Webspector de Illuminix

5- Quelques moteurs de recherche innovants à tester



Classés par ordre alphabétique

- 
-
- **123people** : moteur dédié à la recherche de personne (sur des données publiques)
 - **50matches.com** : indexe uniquement les sites notés sur le réseau Digg ou encore Reddit.
 - **A9** : recherche de produits comme les livres, les films ou encore les articles de presse
 - **Blinkx** : propose à ce jour 18 millions d'heures de vidéos.
 - **Blogdimension** : permet non seulement la recherche de blog, mais également de podcasts, de vidéos en ligne ou encore d'images.
 - **CiteULike** : outil destiné à sauver, retrouver et partager des articles (le plus généralement scientifiques et techniques).
 - **Clusty** : classe les résultats automatiquement
 - **Feedminer** : moteur de recherche entièrement dédié aux flux RSS.
 - **Grokker** : analyse chacun de vos résultats et les regroupe visuellement pour une recherche facilitée
 - **Hittery** : offre un tableau de Bord de Recherche
 - **Intute** : moteur pour les recherches académiques
 - **Ixquick** : Métamoteur de recherche, Ixquick interroge simultanément plusieurs moteurs de recherche dont Google, Ask, MSN, Yahoo ou encore Altavista, et anonymise les recherches.
 - **Kartoo** : cartographie les résultats de recherche.
 - **Kosmix** : Excellent pour les phases d'exploration, les résultats de Kosmix sont plus complets que Google, offrant une vue à 360° d'un sujet.

- 
- **Lexxe** : a développé un moteur de recherche internet de troisième génération avec des technologies d'interrogation en langage naturelle
 - **Like.com** : offre la possibilité de pouvoir effectuer un tri à partir des photos d'objets "semblables".
 - **Local.com** : se spécialise dans la sélection de résultats de recherche localisés, pratique pour trouver les commerces dans un quartier donné.
 - **MetaGlossary** : glossaire anglais complet. Il compte à ce jour plus de 2 millions de termes. Il recense les différentes sources proposant des définitions avec une page de résultats de recherche exhaustive et claire.
 - **Pipl** : moteur de recherche sur les personnes. Il collecte des informations sur de nombreuses sources (Infospace, Google, base de données de brevets, etc.).
 - **QueryCat** : moteur de recherche qui explore les FAQs (Frequently Asked Questions)...
 - **Quintura** : doté d'une interface graphique innovante avec une fenêtre à deux volets : à droite sont présentés les résultats d'une manière classique, à gauche les clusters d'information générée à partir des pages de résultats découvertes par le logiciel sous forme d'une cartographie
 - **Scandoo** : moteur de recherche qui affiche les résultats de Google, Yahoo! ou Live Search, en ajoutant un icône de dangerosité selon que le site proposé est "douteux" ou pas...
 - **sciseek** : recherche dans les domaines sciences & nature
 - **Ujiko** : offre présentation innovante des résultats
 - **Widepress** : moteur d'actualités proposant une interface intéressante et assez complète.
 - **Wikia Search** : permet aux utilisateurs de contribuer au mode de classement des pages de résultats.
 - **Xcavator.net** : permet de trouver des images en fonction de la couleur et de la thématique de la photo.
 - **Yoono** : facilite la recherche sur le Web grâce au partage de la sélection des favoris.



Le programme des prochaines conférences du Club IES (AAE IAE de Paris) est en ligne :

http://www.inter-ligere.net/pages/Prochains_rendezvous-51816.html

La prochaine conférence aura lieu le 7 avril 2009

M. Abdelmalek Alaoui interviendra sur

Le développement de l'Intelligence Economique au Maroc

Voir l'annonce : <http://www.inter-ligere.net/article-28897418.html>

Inscription : <http://iae-paris.org/> (rubrique agenda)