

Google: web caché </title><style><!--

body,td,div,.p,a {font-family:arial,sans-serif
div,td {color:#000

.f,.fl:link {color:#6f6f6f

a:link,.w,a.w:link,.w a:link {color:#00c

a:visited,.fl:visited {color:#551a8b

a:active,.fl:active {color:#f00

.t a:link,.t a:active,.t a:visited,.t {color:#000

.t {background-color:#e5ecf9

Le Web Invisible

.j {width:34em

.h {color:#36c

.i,.i:link {color:#a90a08

.a,.a:link {color:#008000

.z {display:none

div.n {margin-top: 1ex

.n a {font-size:10pt; color:#000

.n .i {font-size:10pt; font-weight:bold

.q a:visited,.q a:link,.q a:active,.q {color: #00c;

.b {font-size: 12pt; color:#00c; font-weight:bold

.ch {cursor:pointer;cursor:hand

.e {margin-top: .75em; margin-bottom: .75em

.g {margin-top: 1em; margin-bottom: 1em

//--;

</style>

<script

<!--

function ss(w){window.status=w;return true;

function cs(){window.status="";

function clk(el,ct,cd) {if(document.images){(new

image()).src="/url?sa=T&ct="+escape(ct)+"&cd="+escape(cd)+

&url="+escape(el.href);}return true;

//--;

</script

>

sommaire

| | |
|---|---------|
| Introduction | page 1 |
| Les bases de données gratuites | page 2 |
| Les répertoirs sélectifs | page 3 |
| Outils et moteurs de recherches | page 6 |
| Bibliothèques Online | page 9 |
| Les services de Push | page 10 |
| Moteurs de recherches de news | page 11 |
| Etudes de marché | page 13 |
| Données économiques et sociales, statistiques | page 16 |
| Les portails | page 19 |

Web invisible , web caché, web profond

Le "web invisible" (deep web, hidden web) désigne la partie du web non accessible aux moteurs de recherche classiques. Le web invisible comprend des bases, banques de données et bibliothèques en ligne gratuites ou payantes...

Voici une sélection de ressources permettant d'accéder à ce web invisible ou au "not pure invisible", le web "mal" indexé par les moteurs.

Une partie du web est non accessible aux moteurs car :

- les documents sont trop volumineux pour être entièrement indexés,
- des pages sont protégées par l'auteur (balise meta qui stoppe le robot),
- des pages sont générées seulement dynamiquement, lors d'une requête par exemple,
- le format des documents n'est pas reconnu par les moteurs (de moins en moins vrai avec un moteur comme Google...).

Chris Sherman et Gary Price, "search engines' US experts", proposent dans leur ouvrage "The Invisible Web" de distinguer 4 types de web.

Le web invisible est vaste puisque selon les estimations, très variables et difficiles, les meilleurs moteurs n'indexent que 3 à 10% du web. Voir à ce sujet les différentes études sur la taille du web

Les 4 types de web distingués par Chris Sherman et Gary Price :

- **Le Web Opaque :**
les pages qui pourraient être indexées par les moteurs mais qui ne le sont pas (limitation d'indexation du nombre de pages d'un site, fréquence d'indexation, liens absents vers des pages ne permettant donc pas un crawling)
- **Le Web Privé :**
les pages webs disponibles mais volontairement exclues par les webmasters (mot de passe, metatags ou fichiers dans la page pour que le robot du moteur ne l'indexe pas).
- **Le Web Propriétaire:**
pages seulement accessibles pour les personnes qui s'identifient. Le robot ne peut donc pas y accéder.
- **Le Web réellement invisible :**
contenu qui ne peut être indexé pour des raisons techniques. Ex : format inconnu par le moteur (Google est l'un des rares moteurs à reconnaître autant de formats), pages générées dynamiquement (incluent des caractères comme ? et &).

Plutôt que le web visible et invisible, l'étude de BrightPlanet préfère évoquer, le *surface web* et *deep web* (web profond). En effet, pour cette société, le problème n'est pas tant la visibilité que l'accessibilité par les moteurs. Il y a un web de surface que les moteurs parviennent à indexer et un web profond que leur technologie ne parvient pas à encore à explorer mais qui est visible à partir d'autres types d'outils (répertoires sélectifs, bases de données..).

On pourrait donc comparer le web à un gigantesque iceberg (en perpétuelle expansion) avec un volume de ressources immergées beaucoup plus important que les ressources de surface. En se basant sur les études Bright Planet et Cyveillance, on pourrait imaginer que le web visible dépassait fin décembre 2002 les 64 milliards de pages (dans l'hypothèse d'un doublement de taille tous les 6 mois) ou "seulement" les 15 milliards de pages (dans l'hypothèse d'une stabilisation de sa croissance). Ces mêmes études permettent également de dire que le web invisible serait 260 fois plus vaste que le web visible. Ces chiffres ne sont évidemment que des hypothèses. Mais l'étude Bright Planet apporte une information intéressante plus précise : les 60 sites les plus importants représentent à eux seuls plus de 40 fois le volume du web visible. Ce sont des sites scientifiques (NASA), des bases de données (Lexis Nexis, Dialog), des sites universitaires (Berkeley), de médias et presse (USAToday), de commerce en ligne (e-bay), des sites internes de grosses sociétés (UPS, Fedex)...

Les bases de données gratuites sur Internet

URFIST de Lyon

L'URFIST (Unité Régionale de Formation et de Promotion pour l'Information Scientifique et Technique) de l'Université Lyon I propose une sélection de plus de 800 bases de données gratuites. Parmi les catégories accessibles, l'agriculture bien sûr mais également les brevets, les marques, la chimie, l'environnement, l'économie, la génétique, l'histoire, l'informatique, la linguistique, les mathématiques, la médecine, les sciences et puis le cinéma, l'art, la photographie.

<http://urfist.univ-lyon1.fr/gratuits/index.html>

URFIST de Nice

L'URFIST de l'Université de Nice-Sophia Antipolis propose elle aussi une sélection de bases de données gratuites.

<http://www.unice.fr/urfist/URFIST-DEH/pages/database.html>

The Internet Archive

The Internet Archive est une bibliothèque digitale destinée à conserver tous les documents numériques issus de l'internet pour les préserver d'une disparition complète. The IA fournit des documents créés à partir de 1996 (10 milliards de pages web mais aussi usenet, films et l'ancêtre Arpanet). The Internet Wayback Machine (développée notamment avec Alexa Internet) permet à l'utilisateur de trouver des sites web archivés en tapant simplement son URL et la date désirée. Étonnant et très utile compte-tenu de la disparition fréquente de site

internet. Accessible au public depuis le 24 octobre 2001.

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

Les répertoires sélectifs

Les répertoires sélectifs recensent des ressources internet professionnelles sélectionnées sur des critères qualitatifs...les sites sont choisis par des professionnels de l'information afin de couvrir les domaines de l'université, de la recherche et donc plus globalement de l'enseignement. Les ressources recensées sont de plusieurs types : sites web de référence, bases de données, ouvrages, journaux électroniques, listes de diffusion, articles et études d'universitaires, chercheurs ou journalistes....Notion développée par B.Foenix-Riou.

The InvisibleWeb

Réalisé par Intelliseek, l'éditeur de l'agent intelligent BullsEye et du métamoteur Profusion. Bon point de départ pour rechercher des bases de données, bibliothèques en lignes et moteurs spécialisés par secteurs d'activité (plus de 10000). Ce portail permet une recherche directe par mots-clés ou à l'aide d'une liste thématique très vaste. Il propose ensuite une liste de résultats (ou de sous-catégories) classés par pertinence ou ordre alphabétique. Le site sélectionné est ensuite brièvement décrit et son interface de recherche souvent accessible sur le site The InvisibleWeb. Résultats évidemment très américains...(ne pas confondre avec Invisible-web.net).

CompletePlanet

Portails de recherche par mots-clés ou annuaire thématique sur plus de 90 000 bases de données ou moteurs de recherche spécialisés (de Agriculture à Weather). Interface et graphisme très agréable, possibilité de sauvegarder ses recherches, moteur de recherche puissant. Réalisé par la société BrightPlanet, auteur d'études de référence sur le web invisible.

Direct Search

Gary Price a compilé des centaines de liens vers des ressources du web invisible, ou du moins, difficilement accessibles. Ce site de la G.Washington University propose un accès par recherche directe ou par catégories thématiques (géographie, histoire, droit, cinéma...). Le bas de la page indique les récents ajouts. Ce n'est donc pas un répertoire au sens stricte mais un liste très utile.

Par ailleurs, Gary Price administre RessourceShelf, très bon weblog relatif aux ressources et actualités destinées aux professionnels de l'information et de la recherche.

RDN (Resource Discovery Network)

RDN est une porte d'entrée britannique d'un réseau de portails consacrées à une douzaine de grands thèmes représentant plus de 35000 ressources : *Affaires (Business), Informatique, Ingénierie, Mathématiques, Sciences Physiques, Sciences Sociales, Sciences Humaines, Droit et à des Sujets de référence (bibliothèques, périodiques) et dans un second temps l'Education, Géographie et Sports.*

Le RDN est organisé comme une coopérative avec une organisation centrale (RDN) et des fournisseurs indépendants appelés les hubs ("pivots répartisseurs") comprenant notamment :

BIOME : Health and Life Sciences

EEVL : Engineering, Mathematics and Computing

Humbul : Humanities

PSigate : Physical Sciences

SOSIG : Social Sciences, Business and Law

On peut accéder à chacun de ces hubs de manière indépendante ou en utilisant l'interface de recherche du RDN.

Autres fonctions :

RDN's "Virtual Training Suite" : tutoriaux pour aider les étudiants et chercheurs

RDN "Behind the Headlines" : liens vers des ressources très spécialisées "délaissées" par les grands médias traditionnels.

Lycos Directory Searchable Databases

Répertoire de plus de 8000 ressources construit à partir de sections de l'Open Directory Project

Invisible-web.net

Chris Sherman (SearchDay) et Gary Price (DirectSearch), les 2 experts américains des moteurs de recherche (www.searchenginewatch.com) ont réalisé ce répertoire qui propose plus de 1000 sources d'accès au web invisible. C'est en fait la version on-line et la mise à jour du répertoire "papier" de la seconde partie de leur ouvrage "The Invisible Web Finding Hidden Internet Resources Search Engines Can't See". (ne pas confondre avec The InvisibleWeb)

WebData.com

Répertoire de plus de 20000 ressources. Recherche directe ou parmi des thématiques. Les bases de données sont catégorisées, organisées et notées selon leur qualité. Catégories : Arts, Commerce, Divertissement, Annuaire, Enseignement, Gouvernement, Droit, Médecine & Santé, Actualité, Société, Immobilier, Ouvrages de référence, Sciences et Technologies,

Sports, Telecom, Internet, Voyage et Hébergement... Très complet.

Bubl Link

Un autre répertoire de ressources internet (plus de 12000) à vocations scolaires et universitaires : sites web de référence, mailing lists, newsgroups ,bases de données..Arts, Sciences Humaines, Littérature, Sociale, Mathématiques, Santé, Ingénierie, Physique... Chaque grand domaine est subdivisé en centaines de sous-catégories.

Librarians' Index to the Internet (Lii.org)

Fondée par la Bibliothèque de Californie : Library of California

Répertoire de près de 9000 ressources internet sélectionnées par les documentalistes. Le LII a été fondée en 1990 et a migré en 1993 vers le serveur de la bibliothèque publique de Berkeley (CA). Le moteur de recherche interne a été développé à partir de 1996. Ce moteur permet de rechercher par sujet, titre, contenu ou URL. La recherche avancée permet d'utiliser les booléens AND, OR et NOT et restreindre le champ des recherches aux bases de données, répertoires, best of ou ressources spécifiques. La rubrique Browse All Subjects permet d'afficher tous les thèmes classés par ordre alphabétique. On peut également effectuer la recherche via les thématiques (de Arts à Women).

The Internet Public Library (ipl.org)

Créée à l'initiative de l'Université du Michigan fin 1995. University of Michigan

Plus de 40000 ressources internet organisées par les bibliothécaires et étudiants de l'université. Répertoire d'ouvrages de références, journaux, critiques littéraires, magazines, journaux et newsletters on-line, associations (essentiellement à but non-lucratif). Recherche par moteur ou thématiques.

The Scout Report Archives

Editeurs de Scout Report. University of Wisconsin-Madison

Les archives de Scout Report : Plus de 10000 sites web et mailings lists sélectionnés pour leur contenu éducatif. Moteur de recherche par sujet, titre, résumé ou URL. Recherche avancée par booléens. The Scout Report est publié tous les vendredis depuis 1994 et propose une sélection de ressources internet dans tous les sujets à valeur éducative. 3 lettres sont spécialisées en Sciences Physiques et Ingénierie, Commerce et Economie et Sciences Sociales

All Academic search

Répertoire de ressources scolaires et universitaires : plus de 200 publications, journaux et magazines sur plus de 50 domaines spécialisés. Recherche directe par mots-clés (auteur, titre, sujet). Liste de plus de 300 publications par disciplines : Agriculture, Anthropologie, Géographie, Art, Architecture, Astronomie, Biologie, Commerce, Chimie, Communications, Informatique, Ethnologie, Sciences Economiques, Enseignement, Ingénierie, Anglais, Environnement et Ecologie, Cinéma, Folklore, Géologie, Geophysique, Gouvernement, Santé, Sciences Humaines, Histoire, Journalisme, Linguistique, Droit, Littérature, Mathématiques, Histoire médiévale, Musique, Philosophie, Physique, Sciences Politiques, Population et Démographie, Psychologie, Services Publiques, Religion, Sciences (général), Sciences Sociales(general), Travail, Sociologie, Théâtre, Urbanisme,

Internets.com

L'interface est dépouillée mais le site est efficace. Utiliser les menus déroulants de gauche : Un moteur spécialisé et un répertoire de base de données et ressources très complet : Arts, animaux, architecture, aéronautique, automobile, biologie, cinéma, enfants, villes, pays, communication, enseignement, informatique, logiciels, web, environnement, divertissement, finance, famille, alimentation, santé, histoire, droit, littérature, politique, sociologie, transports, voyage ,sports..etc...

Outils et moteurs de recherche sur le web invisible

Turbo10

Turbo10, le métamoteur britannique, utilise des moteurs de recherches spécialisés permettant de rechercher dans des bases de données ou des documents du "web profond" dans des domaines spécialisés. Turbo10 offre en effet la possibilité de se connecter à plus de 1000 moteurs spécialisés ou généralistes. Turbo10 interroge par défaut altavista.com, dogpile.com, google.com, hotbot.com, lycos.com, metacrawler.com , search.msn.com et yahoo.com, donc des moteurs plutôt anglophones.

Vous pouvez choisir de rajouter, grâce au module "My Collection", les moteurs figurant dans une liste assez impressionnante (1170 à ce jour). Vous pouvez choisir un moteur généraliste comme voila.fr ou des moteurs spécialisés. Ce qui permet à Turbo10 d'explorer (un peu) le "web invisible" ou "web profond" (l'équipe a même déposé la marque Deep Net !). Car cette liste propose des moteurs internes de portails, de vastes bases de données universitaires ou sites de e-commerce : zdnet.com, amazon.com, europages.net (annuaire de sociétés), imdb.com (cinéma), dictionary.com (dicos et thésauri), eea (european environment agency), encyclopedia.com, findarticles.com (archives d'articles depuis 1998), etc... Une fois constituée votre collection de moteurs, celle-ci apparaîtra dans un menu déroulant à droite de votre boîte de requête.

Google

Encore lui ? Google, moteur étonnant, à part, indexe des formats de documents de plus en plus nombreux et même certaines pages "confidentielles" oubliées par les webmasters sur le net.

Parmi les formats de documents indexés par Google : le PDF, les documents Words, Excel.

Pour une requête sur un format précis, tapez **filetype:** suivi de l'extension du type de document recherché. Exemple : pour rechercher des documents au format Excel contenant le mot "bilan", tapez : filetype:xls bilan. **Plus d'infos dans Astuces et Fonctions Google.**

Xrefer

Moteur de recherche britannique spécialisé dans les ouvrages de référence : encyclopédies, dictionnaires, thesauri & recueils de citations. Thèmes : art, santé, langues, philosophie, musique, sciences, technologies, géographie et littérature anglaise...

Profusion

Le métamoteur Profusion (d'Intelliseek) propose notamment de rechercher des résultats dans le "web invisible" c'est à dire des bases de données de texte, sons, images, articles, magazines, livres...). Profusion propose de sélectionner les sources d'information et moteurs parmi des thématiques verticales (gouvernement, emploi, science, santé, juridique, enseignement, technologies, voyages, finance, art, business...). Ainsi, par exemple, vous pouvez choisir d'effectuer une recherche au sein d'encyclopédies (Encarta, Encyclopedia.com, InfoPlease, Internet Public Library, References) ou de bases de données et publications dédiées à la biologie (BioLinks, Discover Magazine, NBII ...).

Adobe PDF Search

Permet de rechercher parmi plus d'1 million de documents au format Adobe PDF (Portable Document Format)

Looksmart

Dans sa nouvelle version de novembre 2003, Looksmart intègre FindArticles. FindArticles, en collaboration avec Gale, permet de rechercher sur une base données de 3,5 millions d'articles issus de plus de 700 sources (magazines, revues spécialisées [technologie, santé, actualités, art...], journaux). Looksmart permet aussi de rechercher sur le web (Wisnut) et sur un répertoire (Zeal).

Wondir

Wondir associe les possibilités d'un métamoteur et d'un service de recherche humain. Wondir est différent des autres outils de recherche. D'abord parce qu'il est géré par une organisation à but non lucratif. Le but de cette fondation est de fournir de l'information de haute qualité à tous.

Ensuite parce que Wondir combine la technologie d'un métamoteur à des technologies propriétaires qui permettent d'utiliser les ressources du web invisible.

Lorsque l'on tape une requête, la page de résultats se divise en plusieurs parties :

- les résultats web
- les résultats issus des newsgroups (forums de discussion) et mailing lists (listes de diffusion)
- les propositions de service d'experts de la communauté Wondir pouvant potentiellement répondre à votre question
- des questions et leurs réponses en relation avec la requête
- des dépêches d'actualités liées au domaine de la requête.

Wondir dispose en effet d'une communauté de volontaires qui répondent aux questions trop complexe pour le moteur de recherche.

C'est pour cela que Wondir a besoin de votre aide. Vous pouvez mettre votre savoir et vos connaissances à disposition en répondant aux différentes questions qui défilent sur le bandeau "question ticker". Ce bandeau déroulant permet de montrer à tous les visiteurs quelles sont les questions posées par les internautes. On peut y répondre par messagerie instantanée, par mail ou sur un forum. En répondant à une question, vous pouvez gagner des prix.

Les possibilités et la puissance de Wondir augmenteront donc avec le nombre d'internautes utilisateurs.

Thunderstone web site catalog

Thunderstone permet de rechercher des sites via son moteur ou les catégories de son annuaire thématique. Thunderstone privilégie la qualité à la quantité : l'essentiel y est : Politique, Sciences, Economie, Commerce, Santé, Loisirs, Société, Culture, Art, Informatique, Internet, Actualités et Médias... Pour chacune de ces rubriques s'affichent des sous-catégories à la droite de l'écran.

IncyWincy, The Invisible Web Search Engine

Incy Wincy de la société LOOP improvements utilise l'Open Directory Project. En crawlant à l'intérieur des sites de cet annuaire, le moteur Incy Wincy construit un index de milliers de moteurs de recherche internes rencontrés. Exemple : Une requête sur la "CIA" permettra de détecter, sur les sites proposés par l'ODP, le moteur de l'United States Intelligence Community. Une recherche sur "Chimie" affichera, parmi les résultats les moteurs internes du Bottin de la Chimie et de l'Ecole Nationale Supérieure de Chimie de Rennes. On peut ensuite réutiliser ces moteurs par un simple

Bibliothèques on-line

Les catalogues de la Bibliothèque Nationale de France (BnF)

Page recensant les catalogues de la BnF. Ces catalogues décrivent les documents et objets conservés à la BnF (documents imprimés, documents audiovisuels, cartes et plans, monnaies et médailles, manuscrits).

Certains sont numérisés et/ou microfilmés.

- BN-OPALE PLUS, le catalogue des collections patrimoniales
- Catalogue des imprimés en libre-accès
- BN-OPALE. Accessible sous Telnet, principalement tous les livres entrés à la BnF depuis 1970
- Catalogue des documents audiovisuels
- BN-OPALINE : le catalogue en ligne des collections spécialisées
- Catalogue des documents numérisés

Catalogue Collectif de France (CCFR)

Ce site permet de :

- **Trouver des informations détaillées sur les bibliothèques françaises**, leurs collections et leurs fonds et interroger leur catalogue en ligne. Moteur de recherche par type de documents et langues (dont l'Achinese et le Bambara), période de publication ou bien par nom, ville, département et type de l'établissement.
- **Localiser des ouvrages** (documents imprimés, audio, video, multimédia) dans les principales bibliothèques et demander le prêt ou la reproduction des documents. Moteur de recherche par type de documents et langues, ou recherche avancée par auteur, titre, année, éditeur...).

Revue.org

Revue.org est un portail fédérateur de revues francophones en sciences humaines et sociales. Sont notamment proposés : Les Annales historiques de la Révolution française, Cahiers d'histoire , Chroniques yéménites, Corps et culture, Cultures et conflits, Espaces Temps, Etudes photographiques, REMMM (Monde musulman et méditerranéen), RH19 (Histoire du XIXe), Rives, Ruralia...

En outre, Revue.org développe Virtual Library Histoire (annuaire sélectif de liens en Histoire), Calenda (calendrier de l'actualité scientifique des sciences sociales) ainsi que des moteurs de recherche spécialisés.

LibDex

Créé par Peter Scott: University of Saskatchewan (Canada)

Répertoire de plus de 17000 bibliothèques publiques mais aussi privées à travers le monde (133 pays de l'Albanie au Zimbabwe). La recherche peut s'effectuer par pays mais également par Open Access Catalogs (OPAC) c'est-à-dire les catalogues informatisés signalant les ouvrages et les périodiques présents dans la bibliothèque. Egalement des liens vers les sites de e-commerce proposant des ouvrages en ligne et une liste des magazines, journaux et newsletters consacrés aux bibliothèques. Très complet.

OCLC

Réseau informatisé de bibliothèques fondé en 1967 par les présidents des collèges et universités de l'état d'Ohio (Ohio College Library Center).

Près de 100 bases de données utilisées par 41 000 bibliothèques dans 82 pays (en France dans plus de 90 bibliothèques universitaires ou publiques).

Les bases :

ECO (Electronic Collection Online) : accès aux revues électroniques

Worldcat : 45 millions de notices bibliographiques depuis 30 ans en 400 langues

NetFirst : base de sites web indexés et mis à jour

Fastdoc, ContentsFirst : bases de sommaires de périodiques

Sitesearch : Bibliothèque virtuelle.

Outils de recherche :

Firstsearch : accès à des bases de données internationales

Firstsearch eco : accès à plus de 2500 titres de revues électroniques.
click pour préciser la recherche. Une démarche originale.

Les services de push

Net2one actualité française

Net2one permet de rechercher des articles d'actualité récente parmi 2600 sources d'informations, grâce à un moteur de recherche.

Net2one propose désormais 3 grands produits :

- **un Data Center** qui permet d'accéder 24/24h à l'ensemble des informations paraissant sur les sujets de votre choix (actualité entreprise, concurrents, conflits internationaux, informations juridiques etc...). Présenté sous la forme de dossiers, l'utilisateur peut créer un profil d'information personnalisé

- **Une Newsletter** qui permet de recevoir par email jusqu'à 3 fois par jour l'ensemble de l'actualité selon les paramètres définis au sein d'un profil. Un mode "Alerte " vous averti en

temps réel dès la parution d'une information sur le sujet de votre choix (disponible pour les abonnés payants du service).

• un Moteur de recherche sur l'actualité

Le moteur de recherche sur l'actualité permet, en tapant une requête simple, de consulter les dernières informations parues. Il intègre de nouvelles fonctionnalités (recherche sémantique, extraction de concepts les plus cités au sein des articles, liste des sources traitant d'un sujet de façon privilégiée, recherche multilingue etc.. . Aussi vous pouvez taper des phrases ou des questions et non seulement des mots clés exacts. Exemple : Fusion Vivendi Universal, Raison de la grève des professeurs en France... (Accès payant).

NetZone continue à proposer un mode gratuit qui permet de surveiller des sources d'infos (dont les titres apparaîtront au sein de la newsletter ou DataCenter) et un accès aux 8 pages thématiques NetZone. Les sélection de mots clés deviennent payantes.

Pertinence Information Network actualité française et internationale

Pertinence Information Network (PIN) est une plate-forme de veille.

Elle permet la collecte, le traitement et la diffusion d'informations (communiqués, articles de presse, blogs, mailing lists...). PIN traite les données de plusieurs milliers de sources d'informations (Grand site de Presse, portails comme Google News, sites technologiques spécialisés... et est partenaire de groupes tel que PRNewsWire).. L'utilisateur choisit ses sources en fonction de ses centres d'intérêt et de la langue (français, anglais, espagnol, allemand...). Vous choisissez ensuite d'être alerté par mail de la parution d'une information en fonction de vos mots clés. . Détails.

Les moteurs de recherche sur les news

Google Actualités : actualité française, francophone, anglophone et germanophone

Google Actualités (version française lancée en juillet 2003), fournit des liens vers des articles en français ainsi que des photos provenant de sources d'informations du monde entier et notamment de France, de pays francophones comme la Belgique, le Canada, la Suisse et l'Afrique.

Les sources d'informations utilisées sont issues de la presse écrite (nationale et régionale), presse web, TV et radio. Parmi les sources :

Libération - Le Monde - L'Humanité- Le Parisien -Le Figaro - Le Télégramme de Brest - Les DNA - La Voix du Nord - Ouest France - Le Nouvel Obs - L'Express - Actu Live - TV5 - France 2 - DH Net - Radio France...

Le moteur de recherche d'actualités Google News a été lancé en mars 2002 pour être ensuite perfectionné en septembre. En mai 2003, Google News a lancé des versions pour le Canada, l'Inde, la Nouvelle-

Zélande, le Royaume-Uni et l'Australie. En juillet, c'est le lancement de Google Deutschland, le premier Google News non anglophone.

DeepIndex News : actualité française et francophone

Le moteur français DeepIndex propose depuis novembre 2002 un moteur de recherche d'actualités. DeepIndex News, balaie 100 000 urls quotidiennement (chaque heure pour certains sites) pour rapporter environ 200 000 articles de la presse francophone. Intéressant car pour l'instant, l'essentiel des moteurs de recherches de news privilégiaient les actualités anglophones. DeepIndex News recherche en effet parmi les dépêches de la presse française, belge, suisse et canadienne. Côté interface utilisateur, c'est très simple : une boîte de requête par mots clés et un choix parmi 4 durées d'ancienneté: inférieure à 24h, à 7 jours (par défaut), à 1 mois ou supérieure à 1 mois. Rapide, simple et efficace. Liste de sites de presse indexés (en constante évolution)

RocketNews : actualité américaine et internationale

Ce métamoteur lancé fin octobre 2001 permet de rechercher parmi l'actualité récente (news et communiqués de presse de moins de 5 jours). Dans la partie Hot News, des menus déroulant permettent de rechercher 300 sujets répartis en 6 grandes thématiques (Business, Technologie, Santé, Une, Sports, Culture). Interface sympathique et très claire.

Daypop : actualité américaine, française et internationale

Moteur de recherche très puissant d'actualités récentes. Géré par un seul webmaster, Daypop crawle le web au moins une fois par jour et indexe une base de plus de 6300 sites de news et weblogs (sites web où des informations sont publiées fréquemment, souvent présentées par ordre chronologique et par catégories et issues d'internautes non professionnels). On peut choisir de chercher parmi le texte des articles, seulement le titre ou les weblogs. Les résultats affichés peuvent être ensuite affinés par langue (anglais , français, italien, espagnol, allemand, portugais, suédois...). Difficile à prendre en défaut.

AlltheWeb (Fast) : actualité française et internationale

Depuis novembre 2001, le moteur norvégien AlltheWeb qui n'en finit pas d'innover, propose une recherche sur l'actualité. AlltheWeb indexe les sites d'actualités de façon journalière voire plusieurs fois par jour. Cette indexation très fréquente garantie une réelle fraîcheur de l'information. Alltheweb indique dans ses pages de résultats quand le site à été visité par ses robots (ex : Indexed: 6 hours ago ou Indexed: 2001-11-13, 10:20 GMT).

Kooha :
actualité nouvelles technologies-nouvelle économie (France et monde)

Kooha.com est un moteur de recherche spécialisé dans les NTIC et la nouvelle économie

Etudes de marché

Ressources spécialisées par secteur

Une sélection de ressources proposant des études de marché gratuites ou payantes : portails, sites web, organismes publics et cabinets d'études spécialisés dans un secteur : Agriculture-Agroalimentaire, Biotechnologies, Bâtiment (BTP)-Construction-Ingénierie, Luxe-Parfumerie-Cosmétiques-Parfums.

Voir également :

- Ressources multisectorielles
- Données économiques et sociales
- Portails verticaux spécialisés par secteurs d'activité

Quelques sites

• **Agriculture/Agroalimentaire : Tescult**

Recherche et application de technologies nouvelles, études de marché dans le domaine du développement agricole, génie agricole et infrastructures rurales, développement rural et aménagement du territoire.

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

• **Biotechnologies : Biostrategy**

Société spécialisée dans le recueil et l'analyse marketing sur les marchés des biotechnologies et agroalimentaires.

<http://www.biostrategy.fr/>

• **Bâtiment (BTP), Construction, Ingénierie : ABC Marketing**

Cabinet d'études de marché et conseil spécialisé notamment dans les secteurs du bâtiment/construction et des fluides (électricité, climatisation, chauffage, éclairage...)

<http://www.abcmarketing.fr/>

SES (Ministère de l'Equipement)

Statistiques sur les permis de construire, les logements. Enquête Annuelle sur le secteur de la construction : entreprises du secteur (activité, structure et analyse du marché).

http://www.equipement.gouv.fr/statistiques/produits/produit_.htm

•Bébé, Maternité, Femmes enceintes, accouchement : L'Institut des Mamans

Institut d'études spécialisé sur le comportement économique, marketing et social de la maman. Devenir mère modifie les priorités. L'Institut à pour objectif de permettre aux différents acteurs économiques, sociaux et politiques de mieux connaître les besoins, points de vue et opinions des mamans pour comprendre et anticiper leurs comportements.

<http://institutdesmamans.com/>

•Jouets, Jeux : Jouets-veille.com

Portail de veille informative du secteur Jeu et Jouet: Son objectif est de repérer et analyser les informations concernant le secteur jeu & jouet et les mettre à la disposition des professionnels.

- annuaire des professionnels des jeux et jouets
- revue de presse : actualité mensuelle du secteur
- études thématiques

<http://www.jouets-veille.com/>

•Luxe, Parfumerie, Cosmétiques, Parfums. : ABC-Luxe

Portail des professionnels du luxe : Actualités, Economie, Formation, Enquêtes et Etudes. Ex : les japonaises et la beauté, l'industrie du luxe dans l'Union Européenne. Certaines analyses succinctes sont disponibles en ligne gratuitement.

<http://www.abc-luxe.com/>

Euromonitor

Portail d'études de marché : sur la branche "Cosmetics et Toiletries", de nombreuses études des marchés cosmétiques et toilette par pays et par catégorie : bébé, coloration, bain-douche, hygiène buccale, peau, solils, parfums, déodorants...

<http://www.euromonitor.com/Cosmetics>

DONNEES ECONOMIQUES ET SOCIALES, STATISTIQUES

Une sélection de sites offrant des données et statistiques économiques, sociales, démographiques, commerciales... sur les pays du monde entier. De l'Institut National de la Statistique à la CIA...

voir aussi : cartes statistiques

à voir absolument

INSEE

L'Institut National de la Statistique et des Etudes Economiques propose en ligne toutes les données de l'économie et la société française. 3 grandes rubriques : les Grands Indicateurs (pour l'industrie, l'emploi, la consommation, les prix, taux de chômage en France...), la France en Faits et Chiffres (Territoire, population, travail, revenus, conditions de vie, santé, enseignement, économie, commerce extérieur, entreprises, agriculture, industrie, commerce, services...), Les Publications et Prestations (catalogues des collections, ouvrages et périodiques sur support papier et CDROM. Achat en ligne). Un site de plus en plus riche...

Le Quid

Le Quid propose :

- une base de donnée. Pour chaque pays : Cartes, fiches descriptives (infos statistiques), liens web, photos. Navigation par planisphères interactives cliquables.
- une sélection de sites web utiles sur l'économie, l'histoire, la géographie, la démographie de plus de 210 pays.

Egalement des informations sur les 36861 communes de France.

Country Commercial Guides

Enormes guides (en format PDF ou HTML) destinés à l'origine à l'exportateur américain. On y trouve... tout (tendances économiques, environnement politique, commercial, voyages d'affaires, climat d'investissement, comment s'habille les femmes...) et en index toutes les principales statistiques pour un pays. A titre indicatif, le guide France représente 140 pages.

Egalement disponible sur Usatrade.gov pour l'année la plus récente.

Country Briefings

Economist.com fournit des études de marché, études pays et études financières. Pour de nom-

breux pays, Economist.com publie également des enquêtes et analyses. Exemple de la France. Mises à jour régulières.

Country Studies of the Library of Congress

La librairie du Congrès propose des études détaillées sur une centaine de pays avec une priorité aux pays en voie de développement. Approche économique et sociale mais également une étude de la défense nationale, des religions, des partis politiques et de l'histoire de chaque pays. Très documentée, chaque fiche est accompagnée d'une biographie détaillée.

Statistics.com

Grand portail intégrateur de données statistiques. La requête s'effectue par moteur de recherche directe ou par un menu thématique au champ très étendu (Agriculture démographie, géographie, emploi, industrie, commerce, forces militaires, drogues, crime, justice, enquêtes d'opinion, éducation, enseignement, sciences, technologies, recherche et développement, environnement, sports, finances, politique, transports, internet, téléphone, santé...).

Si vous ne trouvez pas ce que vous cherchez, vous pouvez poser votre question par formulaire. Il vous en coûtera environ 20\$

CountryData.com

Choix par menu déroulement d'indices économiques (PIB, Import/Export, dettes, inflation...), de données Risques (justice, armées, religions, commerciales (barrières tarifaires, délais de paiement) pour environ 140 pays. Le prix des données s'affiche au fur et à mesure des sélections.

Strategic Road.com

Cet excellent site d'Intelligence stratégique propose notamment un panorama très complet de données sur les pays, en géopolitique et relations internationales. Une rubrique actualité avec des réflexions et des éditoriaux sur des événements récents. La rubrique liens, classés par pays propose des données générales, et aussi les actualités, l'économie, les institutions, le juridique, opportunités. Le classement est impeccable. A voir notamment pour la rubrique actualités.

CIA : The World Factbook

Certainement les infos pays en lignes les plus célèbres (remises à jour régulièrement) : Géographie, démographie; gouvernement, transports, économie, communication, relations internationales, défense. Synthétique mais l'essentiel y est. La partie Reference Maps propose

de très belles cartes. Le choix par la rubrique Champ (field listing) permet de faire des comparaisons entre pays.

CountryReports.org

Données synthétiques pour plus de 100 pays : Carte, Géographie, Histoire, Gouvernement, Partis politiques, Syndicats, Défense, Economie, Transports, Communications, Habitants, Relations internationales, Climat. Le petit plus : une sélection de liens essentiels pour chaque pays

Rechstat.com

Ce nom aux consonances curieuses signifie Recherche Statistique et recense les ressources statistiques présentes sur le web. Pas très joli mais dense et complet. Un bon point de départ.

Ministère des Affaires étrangères, Commerce internationale du Canada

Le ministère présente pour 130 pays, des fiches documentaires (renseignement généraux, données économiques, politiques et commerciales, les perspectives d'investissement dans ces pays...Egalement des études de marché

Les Portails

Sites fédérant de nombreuses ressources (articles, forums ,actualités, fiches produits...) organisées autour d'un annuaire thématique. Le portail horizontal concerne une activité transversale (le droit, les moteurs de recherche). Le portail vertical (ou vortail, vortal) cible une communauté professionnelle et donc un secteur d'activité (l'agriculture, l'aéronautique...).

Portails horizontaux

Enfin.com

Enfin.com répertorie les outils de recherche francophones en les classant en grandes catégories : art et culture, commerce et économie, informatique, sexe, institutions, internet, loisirs et sports, santé & spiritualité, régional (pays). Enfin est également disponible en version internationale

Fossick Europe

Fossick est un portail de répertoires et de moteurs de recherche. Il recense plus de 3000 moteurs de recherche et annuaires spécialisés, classés par thématiques ou grandes zones géographiques, qui vous guideront ensuite vers des recherches plus "pointues". Le site propose également un métamoteur qui permet de centrer un rcherche sur un pays européen, Russie comprise.

Page d'accueil internationale : www.fossick.com

Beaucoup !

Portails de moteurs de recherches et répertoires spécialisés. Il fédère plus de 2500 outils de recherche organisés en 15 grande catégories (Geographical, Health, Media, Buisness, Society...).

Les Portails

Sites fédérant de nombreuses ressources (articles, forums ,actualités, fiches produits...) organisées autour d'un annuaire thématique. Le portail horizontal concerne une activité transversale (le droit, les moteurs de recherche). Le portail vertical (ou vortail, vortal) cible une communauté professionnelle et donc un secteur d'activité (l'agriculture, l'aéronautique...).

Portails verticaux (vortails)

Les vortails (portails spécialisés dans un secteur d'activité) sont nombreux. Ainsi pour le secteur de la chimie, il existe (entre autres...) France-Chimie, Chemindustry, Chem.com, Chemscope, Chemweb, Chemicalonline ...auxquels il faut rajouter tous les vortails concernant les secteurs dérivés de la chimie

(plasturgie, pharmacie, cosmétique...)

Pour identifier et trouver ces portails, il convient d'utiliser....d'autres portails :

- les portails de moteurs de recherche (décrits ci-dessus): ils vous permettent d'identifier des outils de recherches spécialisés par secteurs d'activités qui pourront vous amener sur des portails.

- les répertoires de portails. Attention, le terme de portails est parfois utilisé abusivement pour de simples sites spécialisés. En voici quelques-uns :

TradeWorlds

Répertoire de portails verticaux B2B.

Portails 2

Ce portail regroupe des ressources grand public ou professionnelles (portails ou simple sites spécialisés). En français.

Les 1000 meilleurs portails d'affaires sectoriels

Le site Objectif Grandes Ecoles propose une rubrique consacrée aux portails sectoriels : Agriculture et Agro-alimentaire, Bâtiment et Construction, Commerce et distribution, Energie et ressources naturelles, Biens de consommation, Informatique et Télécommunications, Industries lourdes, Médias et journalisme, Professions artistiques, loisirs et tourisme, Santé et Environnement, Secteur public, enseignement et recherche, Services financiers et consulting, Transport et logistique.

<http://www.objectifgrandesecoles.com/pro/secteurs/index.htm>

Indexa

Indexa est un annuaire de sites professionnels. En explorant la catégorie "Autres sites, portails" présente dans chaque secteur d'activité, ou en tapant "portails" sur le moteur de recherche interne, vous pourrez découvrir plus de 500 portails verticaux.

VertMarkets

VertMarkets (marchés verticaux), créé à l'origine par VerticalNet propose plus de 65 places de marché dans de multiples secteurs d'activité (communication, énergie, alimentaire, santé...). Chaque portails, outre les services d'achat-vente, proposent des actualités du secteur, des

forums, une librairie, un calendrier des évènements...

EDG.net

C'est un peu l'équivalent français de VertMarkets. EDG propose plus de 30 portails verticaux (Techniques de communication, Protection biens et personnes, Gestion des flux, Bâtiment, Equipements et services industriels, Transformation de la matière, Recherche et analyse, Dépollution, Electronique, Télécom et informatique, Agro-alimentaire industries et commerces, Santé, Aménagement des paysages, Equipements et services pour la distribution et les commerces, Emballage, Automobile...). Les portails offrent notamment des infos produits, un agenda évènementiel, l'actualité du secteur.