



Au début des années 1990, la croissance du nombre de réseaux et leur interconnexion ouvraient de nouvelles perspectives de collaboration, d'échange d'informations et de diffusion à un niveau mondial.

# INTERNET et l'information scientifique

## Internet qu'est-ce que c'est ?

Par son taux de croissance, son extension géographique, le déploiement de ses utilisateurs et la diversité des ressources disponibles, Internet présente des potentialités qui semblent illimitées en termes d'échange, de collaboration, de diffusion et d'accès à des ressources informationnelles...

■ **Du point de vue technique**, Internet est un terme générique qui désigne la multitude des réseaux d'ordinateurs utilisant les protocoles TCP/IP (Transmission Control Protocol/Internet Protocol) interconnectés au niveau mondial.

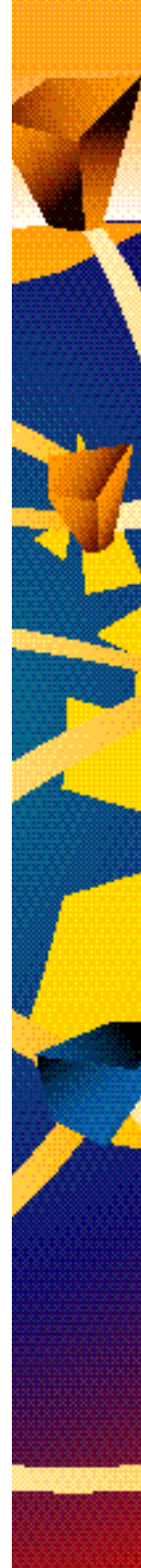
En France, le réseau **Renater** (Réseau national de télécommunication pour la technologie, l'enseignement et la recherche), interconnecte la majorité des universités, des laboratoires de recherche publics ou privés

et des autres établissements d'enseignement supérieur.

Son architecture modulaire et répartie comporte plusieurs niveaux (cf schéma dispositif technique p 6 et 7) :

- au niveau de chaque région française, un réseau régional relie les réseaux des universités et autres sites.
- le réseau national d'interconnexion (RNI) fédère à son tour les réseaux régionaux et leur donne accès aux liaisons vers les autres réseaux de la recherche et l'Internet par l'entremise du Noeud de Transit International (NTI). Renater a sa propre liaison vers les Etats-Unis. Il est également interconnecté avec les réseaux de recherche européens par le réseau TEN 34 (<http://www.dante.net>).

Renater a été lancé conjointement en 1991 par les ministères chargés de l'Enseignement



supérieur, de la Recherche et des Télécommunications. Le Groupement d'intérêt public (GIP) Renater qui regroupe le CEA, le CNES, le CNRS, EDF, l'INRIA et le ministère de l'éducation nationale, de la recherche et de la technologie assure la maîtrise d'ouvrage du RNI et du NTI tandis que les réseaux régionaux sont bien souvent gérés par les associations des principaux utilisateurs. En septembre 1997, il comptait 25 réseaux régionaux y compris les DOM et plus de 500 sites connectés (sources: <http://www.urec.fr/Renater> et GIP Renater).

D'autres réseaux de télécommunications permettent également une connexion à Internet : Transpac, Numéris ainsi que des réseaux étrangers gérés par des opérateurs tels que Pipex, EUNET.

Les offres et les possibilités se multiplient dans le cadre du développement de services pour le grand public (cf notamment celle des acteurs traditionnels de la diffusion télévisuelle, offre multiservice des câblo-opérateurs, acteurs du satellite ...).

## Quels en sont les principaux acteurs ?

### ■ Les opérateurs de réseaux

Fournisseurs des infrastructures utilisées pour véhiculer l'information, ils constituent le premier maillon du développement des réseaux de communication.

En France, le principal opérateur est jusqu'à présent France Télécom mais d'autres acteurs des télécommunications, du câble ou du satellite se positionnent dans la perspective du développement du marché du multimédia.

### ■ Les fournisseurs d'accès

De nombreux fournisseurs d'accès à Internet sont présents sur le marché français (pour en savoir plus : <http://www.nic.fr/Prestataires>) mais on commence à assister à d'importants mouvements de concentration dans ce secteur. Certains sont spécialisés en connexions haut débit, d'autres n'offrent que des accès grand public à faible débit.

Les opérateurs sont également souvent des fournisseurs d'accès : par exemple, France Télécom offre depuis la fin de l'année 1996, son service Wanadoo. Connaître l'architecture du raccordement proposé par le fournisseur est un critère de choix important pour ne pas avoir de surprises sur les temps de réponse.

Pour le monde académique, Renater reste jusqu'à ce jour le réseau privilégié. Opérateurs et fournisseurs d'accès sont interconnectés via un point d'échange mis en place par le GIP Renater appelé SFINX.

■ **Les acteurs informatiques** : éditeurs de logiciels, constructeurs, sociétés de services. La concurrence, vu les enjeux, est très forte et le rythme des développements très rapide : nouveaux terminaux (Network Computer...), prestations de services pour la réalisation de serveurs Web, moteurs de recherche et ingénierie linguistique, nouveaux langages (Java, Active X).

### ■ Les producteurs

Ce sont les acteurs qui produisent et mettent à disposition l'information. Elle est très hétérogène sur Internet : articles, lettres d'information, plaquettes commerciales, banques de données, fichiers informatiques, images, animations en 3 dimensions... Les producteurs peuvent être des chercheurs, des étudiants, des laboratoires, des universités mais également les entreprises les plus diverses, les états et les organisations qui en dépendent, ou bien encore des particuliers... En résumé, on trouve ce que toute personne veut diffuser.

### ■ Les serveurs

Par serveur, on désigne généralement sur Internet les machines qui hébergent les services.

De nombreuses sociétés proposent le développement, la maintenance et l'hébergement des applications Internet telles qu'un serveur d'information WWW. Certains fournisseurs sont aussi des prestataires de services.

Sur un campus, le centre de ressources informatiques peut jouer ce rôle mais de nombreux chercheurs développent eux-mêmes leurs services grâce à des environnements de plus en plus conviviaux.

### ■ Les formateurs

Les interfaces évoluant rapidement vers plus de simplicité, de nombreux utilisateurs se forment eux-mêmes. Mais des formations permettent d'acquérir la maîtrise des différents outils logiciels ainsi que des contenus. Les possibilités de formation sont multiples. Les URFIST, par exemple, organisent ce type d'action (cf p. 62). On trouve également de nombreuses ressources de formation sur le réseau lui-même : cours envoyés sur votre messagerie, guides, supports de cours (cf exemple p. 63 et bibliographie p. 68).

## Les Coûts

■ **Pour les établissements** qui se connectent via Renater, les coûts sont forfaitaires en fonction du débit de la liaison souscrite soit de 64 kbits/s à 34 Mbits/s (volume transmis et temps d'utilisation ne sont pas limités).

• au niveau régional : les tarifs varient selon les régions et comportent des frais de raccordement (coût uniquement au moment du raccordement) et une redevance mensuelle d'abonnement. Par exemple, en Ile de France, les frais de raccordement pour une prise de 64 Kbit/s est de 12 000 francs HT et la redevance mensuelle de 4 700 francs HT (cf, <http://www.aurif.fr>).

• coût du RNI et du NTI : ce coût est souvent pris en charge par le ministère chargé de l'enseignement supérieur et de la recherche. Le prix d'une prise de 64 Kbit/s est de 36 KF HT/an (tarif de novembre 1997).

■ **Pour le grand public**, les fournisseurs de connexion Internet affichent des tarifications très diverses. La majorité des tarifs sont fondés sur un abonnement mensuel qui peut donner droit à un nombre de connexions limité ou illimité, parfois complété par un droit d'entrée annuel. La fourchette des prix se situe entre 45 et plus de 200 francs TTC par mois selon les services associés (tarifs fin 1997).

## Les services Internet

Le réseau offre différents types de services :

■ **Le courrier électronique (mél ou e-mail)** : il suffit de connaître l'adresse électronique de son correspondant pour lui faire parvenir un message.

L'adresse électronique se décompose généralement de la manière suivante :

mon nom@nom de la machine qui m'héberge.  
le sous-domaine auquel elle appartient.le domaine

ex : dupont@moka.ccr.jussieu.fr (la machine : moka, le sous-domaine : ccr.jussieu et le domaine fr pour la France)

Des logiciels avec des interfaces conviviales permettent de gérer facilement son courrier : par exemple, un logiciel dédié tel qu'Eudora ou un navigateur tel que Netscape qui intègre des fonctionnalités de gestion du courrier électronique.

■ **Le transfert de fichier (FTP, File Transfer Protocol)** : il suffit de connaître l'adresse du site où se trouve le fichier ainsi que son nom.

La commande élémentaire est donc :

FTP adresse du site

ex : ftp.ftn.cnam.fr avec comme login "anonymous" pour les sites FTP publics et sa propre adresse électronique comme mot de passe.

Un ensemble de sous-commandes permet de localiser le fichier et de le rapatrier. Il est plus simple d'utiliser une interface graphique du type Fetch (pour les Mac) ou Winftp (pour les PC) ou encore un logiciel tel que Netscape pour le téléchargement d'un fichier en local.

■ **L'ouverture d'une session de travail sur une machine distante (TELNET)** : encore une fois, il suffit de connaître l'adresse du site. La commande de base est :

telnet adresse du site

ex: telnet esrin.esa.it pour se connecter au serveur de l'Agence Spatiale Européenne ESA/IRS.

Cette commande permet notamment de consulter les banques de données et les catalogues en ligne des bibliothèques. Les accès Telnet sont progressivement remplacés par des accès plus conviviaux de type "formulaire" offerts sous WWW.

■ **L'accès aux serveurs d'informations WWW** (World Wide Web, ou encore, Web, W3, la Toile...) :

WWW est un système hypermédia de gestion de l'information : à partir de modules clients tels que les navigateurs Netscape (Netscape) ou Internet Explorer (Microsoft), on accède à des serveurs offrant des contenus multimédia reliés par des liens hypertextes. Une ressource est référencée par son URL (Uniform Resource Locator).

HTML (HyperText Markup Language) est le langage qui permet de développer les pages d'un serveur WWW.

Syntaxe d'une URL :

Protocole://adresse du serveur/chemin du fichier (répertoire/sous-répertoire...)/nom du fichier

On indique en premier le nom du protocole du service auquel on veut accéder : http est le nom du protocole pour les serveurs WWW, comme ftp est le nom du protocole pour les serveurs FTP (par exemple : ftp://ftp.cnam.fr) ensuite l'adresse du serveur puis le chemin pour atteindre le document.

ex : Adresse de la liste des universités françaises sur le serveur du ministère de l'éducation nationale : <http://www.education.gouv.fr/sup/univ.htm>

## Des ressources diversifiées

Les services cités précédemment permettent de gérer et de rendre accessible une très grande diversité d'informations et de ressources. Cette hétérogénéité est l'une des caractéristiques majeures d'Internet. En se limitant au domaine de l'information scientifique et technique, voici un aperçu des produits et services que l'on peut trouver :

### ■ Communication :

• **les listes de discussion** : des groupes de discussion peuvent communiquer grâce au courrier électronique et à des logiciels de gestion de listes (comme Listserv...). La copie d'un message envoyé à la liste est redistribuée à tous les membres qui se sont abonnés à celle-ci. Des milliers de listes sont disponibles, sur toutes sortes de sujets : médecine, informatique, éducation, histoire, bibliothèques...

• **les news** constituent un autre système pour les forums de discussion. On compte actuellement plusieurs milliers de groupes de news. Le fonctionnement est différent de celui des listes : l'utilisateur consulte les messages qui l'intéressent en accédant à une machine-serveur de news. Il ne reçoit donc pas systématiquement les messages, ce qui permet d'opérer une sélection (on peut vite être submergé si l'on est abonné à des listes dont le nombre de participants est important).

Répertoire international de forums, listes, news, journaux académiques et professionnels :

<http://n2h2.com/KOVACS>

Services d'accès aux news internationales :

<http://tile.net/news> et <http://www.dejanews.com>

Répertoire des listes de discussion françaises :

<http://www.cru.fr/listes/>

Services d'accès aux news françaises :

<http://www.fr.net/news-fr/>

• **les bulletins d'informations** : ils constituent un vecteur de communication important entre chercheurs pour connaître le calendrier des futures manifestations scientifiques de leur domaine, les appels à communication, les résumés et commentaires de certains travaux. On les trouvera, en particulier, sur les serveurs WWW des sociétés savantes et des instituts de recherche.

### ■ Publications scientifiques :

• **Les rapports de laboratoires** : pour diffuser leurs rapports internes, de nombreux organismes les ont mis en ligne sur leur serveur.

Rapports de l'Institut National de Recherche en Informatique et Automatique :

<http://www.inria.fr/RRRT/publications-fra.html>

• **les serveurs de pré-publications** : la revue est parfois court-circuitée par ces bases de pré-publications qui, dans certains domaines comme la physique théorique, sont très utilisées.

Serveur spécialisé en physique, lancé par le chercheur américain Paul Ginsparg : <http://xxx.lanl.gov>

Serveur miroir en France : <http://xxx.ljthe.jussieu.fr>

Serveur du Cern : <http://preprints.cern.ch>

• **les revues électroniques** : l'offre s'est largement développée en 97 avec notamment la mise en ligne sur Internet de nombreuses revues scientifiques existantes. Mais après une période de test pendant laquelle on pouvait accéder au texte intégral gratuitement, de plus en plus souvent les accès sont payants, tandis que la consultation des sommaires ou des résumés des articles demeurent des produits d'appel gratuits que la plupart des éditeurs scientifiques propo-

sent. Des services complémentaires d'alerte sont très souvent offerts.

Mais on assiste aussi à la création de nouvelles revues, purement électroniques, émanant du secteur commercial ou se posant comme une alternative à celui-ci.

#### Exemples de revues existantes :

Science Magazine : <http://www.sciencemag.org>

Nature : <http://www.nature.com>

La Recherche : <http://www.larecherche.fr>

#### Exemples de serveurs d'éditeurs offrant les sommaires et/ou des résumés, le full text (pour certains seulement) :

Daloz-Sirey : <http://www.daloz.fr/>

Masson : <http://www.masson.fr/> et <http://www.e2med.com>

Presses universitaires de France :

<http://www.puf.com/edition/revues/index.htm>

Academic Press :

<http://www.europe.idealibrary.com/glogin.htm>

Springer : <http://link.springer.de>

#### Revues électroniques uniquement :

Journaux électroniques de mathématiques :

<http://www.emath.fr/welcome.htm>

Psychology, revue en sciences cognitives, lancée par

Stevan Harnad, de l'université de Southampton :

<http://www.cognisci.soton.ac.uk/psychology>

Solaris, revue française en sciences de l'information :

<http://www.info.unicaen.fr/bnum/Solaris>

#### Serveur sur les projets des sociétés savantes au niveau mondial :

<http://www.lib.uwaterloo.ca/society/>

#### • Des actes de Congrès, de journées d'études, des mémoires d'étudiants :

Actes de congrès de l'European Series in Applied and Industrial Mathematics :

<http://www.emath.fr/Maths/Proc/proc.html>

Actes de la journée d'étude sur les revues scientifiques organisée par l'ENSSIB (Ecole Nationale supérieure des Sciences de l'information et des Bibliothèques) et la SFSIC (Société Française des Sciences de l'Information et de la communication) :

<http://www.enssib.fr/eco-doc/journees.html>

#### ■ Outils documentaires :

• **les catalogues de Bibliothèques** : de nombreux catalogues sont interrogables dont ceux de la Bibliothèque nationale de France, de la bibliothèque du Congrès, de la National Library of Medicine, du CERN et de nombreuses bibliothèques françaises (cf p 56 et 67).

Quelques serveurs recensent les catalogues :

- francophones

[http://www.enssib.fr/Enssib/f\\_biblio.fr.htm](http://www.enssib.fr/Enssib/f_biblio.fr.htm)

- des bibliothèques universitaires françaises

<http://sdbib.mesr.fr/sommaire/sbu.htm>

- des bibliothèques nationales dans le monde

<http://ifla.inist.fr/II/natlib.htm>

#### • des librairies virtuelles :

Alapage (ouvrages, disques et vidéos disponibles sur le marché français) :

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

Amazon (couverture internationale) :

<http://www.amazon.com>

• **des banques de données bibliographiques et en texte intégral** : les serveurs commerciaux ont tous désormais une adresse Internet et rendent leurs banques accessibles sur ce réseau. Ces produits restent payants, à part quelques exceptions comme la banque Medline. Les interfaces ont évolué et l'on peut de plus en plus souvent interroger ces banques de données à partir de son navigateur Web. **Les liens entre** les banques bibliographiques et le texte intégral des articles font l'objet de nombreux développements. A noter qu'à côté de ces produits, il existe aussi de petites banques de laboratoires accessibles souvent gratuitement.

Accès libre à Medline sur le site de la National Library of Medicine (Pubmed et Gratefulmed) :

<http://www.nlm.nih.gov/databases/freemed.html>

Œuvres littéraires numérisées :

<http://cedric.cnam.fr/ABU/>

Banque de données Santé publique (Ecole nationale de la Santé publique et un réseau de partenaires) :

<http://www.bdsp.tm.fr/>

#### ■ Banques d'images :

Iconocerf, banque d'images radiographiques de cas cliniques développée par le Collège des Enseignants de Radiologie de France :

<http://www.med.univ-rennes1.fr/cerf/>

Enluminures à la Bibliothèque Nationale de France et 7 000 images sur son serveur expérimental Gallica qui offre des images et textes du XIXe siècle francophone :

<http://www.bnf.fr> et <http://gallica.bnf.fr/>

#### ■ Données expérimentales partagées :

Dans certaines disciplines, l'échange scientifique majeur est la donnée factuelle, résultat d'une expérimentation. C'est, par exemple, le cas pour l'astronomie, la biologie moléculaire ou encore la sismologie.

Banque de données SIMBAD du Centre de Données Astronomiques de Strasbourg :

<http://simbad.u-strasbg.fr/Simbad>

Banques de données internationales en biologie moléculaire : <http://www.infobiogen.fr>

Données en sismologie de l'Institut de Physique du Globe : <http://geoscope.ipgp.jussieu.fr>

#### ■ Sites disciplinaires de référence :

Ce sont des sites où sont référencées de nombreuses ressources intéressantes des communautés d'étudiants et de chercheurs par disciplines.

Cellule de coordination nationale pour les mathématiques :

<http://www.mathdoc.ujf-grenoble.fr/>

Site de ressources en Médecine sur le serveur du CHU de Rouen :

<http://www.chu-rouen.fr/>

Serveur de sociétés savantes, par exemple, celui de l'Association for Computer Machinery :

<http://www.acm.org>

#### ■ Ressources pédagogiques :

Cours, ressources multimédias pour l'enseignement, expérimentations pédagogiques ...

CNDP (Centre national de la documentation pédagogique) :

<http://www.cndp.fr>

CNED (Centre national d'enseignement à distance) :

<http://www.cned.fr>

Serveur sur l'enseignement à distance et le multimédia éducatif : <http://www.cru.fr/EAD>

GEMME (Groupement d'intérêt scientifique "Enseignement supérieur sur Mesure Médiatisée") :

<http://www.unice.fr/cdiiec/gemme.htm>

#### ■ Informations administratives :

Informations publiques des organismes gouvernementaux, les offres de bourses, les appels d'offres, les renseignements divers...

Ministère de l'éducation nationale :

<http://www.education.gouv.fr/>

Association Bernard Gregory :

<http://www-com.grenet.fr/abg/>

Commission européenne :

<http://europa.eu.int/>

#### ■ Logiciels :

les logiciels les plus divers peuvent être téléchargés comme les visualiseurs (viewers) et leurs modules d'extension (plug-ins) nécessaires, par exemple, pour afficher différents formats de données :

Adobe Acrobat Reader pour les fichiers en pdf (format majoritaire, pour l'instant, des journaux électroniques commerciaux) :

<http://www.adobe.com>

Serveurs ftp anonymes français sur lesquels on peut trouver de nombreux logiciels :

[http://web.urec.fr/docs/ftp\\_fr.htm](http://web.urec.fr/docs/ftp_fr.htm)

# La recherche d'informations sur Internet

Comme système d'information et de communication, Internet présente les caractéristiques suivantes, il est :

- **distribué**,
- **hétérogène** : des ressources très différentes coexistent ce qui soulève le problème du degré de fiabilité de l'information à laquelle on accède.
- **public** en grande partie actuellement : de nombreux espaces d'information et de communication sont accessibles librement par tous,
- **en perpétuelle évolution** et donc instable : les sites d'information naissent, évoluent, disparaissent quotidiennement.

L'ampleur du réseau et son architecture distribuée rendent donc impossible le recensement exhaustif de ses ressources. Contraire-

ment à des dispositifs tels que ceux offerts par les serveurs commerciaux ASCII, il n'existe pas de stratégie unique pour retrouver l'information. De nombreux outils sont disponibles, caractérisés par un niveau de spécialisation plus ou moins important, des méthodes de recueil de l'information variées et à des degrés de finesse très divers : pour une recherche optimale, il est nécessaire de **multiplier les stratégies** et de **recouper les résultats**.

Les pages d'accueil des outils présentés ci-dessous peuvent offrir de nombreux autres fonctionnalités (aide pour restreindre les résultats de la requête, traduction), services et points d'entrée complémentaires (serveurs les plus utilisés, sélections de sites nouveaux, dépêches de presse...) qui ne seront pas détaillés ici.

## Les moteurs de recherche

Les moteurs de recherche sont des programmes informatiques qui se chargent d'une recherche sur tout ou partie de l'Internet. Ils répondent à une requête exprimée généralement sous forme de mots, reliés éventuellement par des opérateurs booléens.

Les plus connus sont AltaVista, Hotbot, Excite, Lycos, InfoSeek, Lokace, Ecila... Certains sont accessibles par le bouton "Net Search" de l'interface du navigateur Netscape ou Search d'Internet Explorer.

AltaVista : <http://www.altavista.digital.com/>  
Hotbot : <http://www.hotbot.com>  
Excite : <http://www.excite.com>  
Lycos : <http://www.lycos.com>

Les adresses des moteurs francophones sont présentées p 49.

### ■ Ces outils se décomposent en 3 modules principaux :

- le module de collecte automatique des ressources (c'est un logiciel qui demande en permanence des copies de pages HTML)
- le module d'indexation
- le module de recherche de l'information qui gère l'interface avec l'utilisateur.

### ■ Choix des sites explorés :

Les stratégies sont variées mais le plus souvent, sont utilisés les pages "What's new", des classements de sites très demandés, les déclarations spontanées des auteurs mais également des explorations aléatoires.

### ■ Contenus indexés :

On constate également une grande variété dans les pratiques d'indexation :  
- pages entières (AltaVista)  
- parties seulement : titres HTML, premiers paragraphes, liens (Lycos)...

Cette indexation est associée à une pondération des mots fondée sur différents critères dont la fréquence d'apparition du terme ou sa position dans certaines zones de codage des pages HTML.

### ■ Utilisation :

Pour une utilisation efficace de ces moteurs, il faut connaître la syntaxe du langage d'interrogation. Un guide en ligne est toujours disponible.

Les interfaces présentent généralement 2 modes d'interrogation :

- **un mode simplifié** : on donne une liste de mots et le système retrouve des documents répondant "au mieux" à celle-ci selon des algorithmes peu explicites mais qui tiennent compte de la localisation, de la proximité et de l'occurrence des termes. La troncature à droite est souvent implicite (cf p 18).

- **un mode "avancé"** (advanced) ou "expert" : l'utilisateur peut utiliser des opérateurs pour lier les termes de sa question et, éventuellement, orienter la pondération des documents retrouvés.

### ■ Avantages et inconvénients :

Compte tenu de l'automatisation du processus, les moteurs sont les outils qui offrent la couverture la plus exhaustive de l'espace public Internet avec une mise à jour rapide. Ils sont bien adaptés à des **recherches précises ou pointues**. L'interaction est simple. Mais les réponses sont hétérogènes, les doublons ne sont pas toujours détectés, le bruit peut être considérable. De plus, leurs recensements connaissent certaines limites : délais de plus en plus importants pour parcourir la masse d'information, augmentation des refus d'accès pour des raisons de sécurité et de confidentialité...

### Quelques règles de base tendent à se généraliser :

- l'utilisation des doubles quotes (") pour spécifier une expression ou une phrase et éviter que chaque mot ne soit pris en compte séparément,  
*ex : "revue scientifique".*

- l'utilisation d'opérateurs de présence absolue (+)  
*ex : cnrs +observatoire , requête pour retrouver des pages HTML concernant les sites d'observatoires du cnrs, le + signifie que le mot observatoire doit absolument être présent dans les pages retrouvées.*

- l'utilisation de l'exclusion absolue de termes (-)

- Il faut tester la prise en compte ou non des caractères accentués et de la casse (majuscule/minuscule) des caractères.

## Exemple de recherche avec le moteur AltaVista.

Question: "Retrouver des images de Mars prises par la NASA".  
 Pour une question précise comme celle-ci, l'utilisation d'un moteur de recherche est bien adaptée.

**Possibilité de choisir de rechercher sur le Web ou les Newsgroups**

**le + rend obligatoire la présence de ces deux mots**

**résultats : nombre de pages retrouvées**

**Documents 1-10 d'environ 1000 correspondant à la requête, les meilleures correspondances d'abord.**

**1. Mars Today**  
 Mars Today, Mars Today, created by Howard Houben of t...  
<http://humbake.ars.nasa.gov/> - size 5K - 29-Feb-97 - English - Traduction

**2. Mars - Viking 1 Orbiter**  
 Mars - Viking 1 Orbiter. Cirrus clouds north of Valles Marineris, Mars. This Viking 1 Orbiter image show clouds to the north of Vallis Marineris. These...  
[http://imgcat.html/object\\_page/va\\_L\\_056a02.html](http://imgcat.html/object_page/va_L_056a02.html) - size 2K - 21-May-97 - English - Traduction

**première page retrouvée: affichage du titre, de l'URL, de la taille de la page et de sa langue**

**4 Site 009**  
 Site 009 - Sinus Sabæus. Site Name: Sinus Sabæus Type of Site: Rover/Sample Return (A link to the appropriate page of Part 1 or 2) Latitude: 6.5 deg. S...  
[http://www.ars.nasa.gov/MarsTools/Mars\\_Cat/Part\\_3/Sites/site009.html](http://www.ars.nasa.gov/MarsTools/Mars_Cat/Part_3/Sites/site009.html) - size 3K - Traduction

**la connexion sur le 4<sup>ème</sup> résultat conduit au site présenté page de droite**

Adresse de la version française d'AltaVista :  
<http://altavista.telia.com>

le site américain d'AltaVista :  
<http://www.altavista.digital.com>

**Options supplémentaires sur le serveur américain :**

- différents annuaires thématiques
- possibilité de préciser les résultats de la recherche à partir de termes proposés par le logiciel

## Sommaire du site et exemples d'images proposées par le "Center for Mars Exploration"

**CMEX**  
 NASA Ames Space Science Division

Welcome to the Center for Mars Exploration

Table of Contents

- Mars News and Information: Mars Global Surveyor
- Mars Atlas/Map/Images
- Educational Resources
- Frequently Asked Questions

**VIKING ORBITER VIEWS OF MARS**

BY THE VIKING ORBITER IMAGING TEAM

- H. H. Larr
- W. A. Pearen
- K. R. Blasius
- G. A. Briggs
- J. A. Cutler
- T. C. Duxbury
- H. Grotzinger
- J. Guest
- H. Masursky
- B. A. Smith
- L. A. Soderblom
- J. Veverka
- J. E. Wellman

Greg P. Sullivan, Editor

Scientific and Technical Information Branch  
 National Aeronautics and Space Administration  
 Washington, DC

**MARS TODAY** 3/11/98 12:00:00Z

Diameter: 3.99 Phase Angle: 10.23  
 Ls: 290.46 De: -23.60  
 Longitude: 287.89 De: -25.43

Temperature (K) 140 165 190 215 240

30 m/s

# Les annuaires thématiques

(ou encore répertoires ou catalogues)

Contrairement aux moteurs, ce ne sont pas des outils automatiques. Ils sont l'aboutissement de recensements thématiques réalisés selon des modalités très diverses : sondage, déclaration des serveurs, "veilleurs nationaux". Les annuaires Yahoo!, Argus Clearinghouse, Magellan, Nomade en sont des exemples.

Yahoo! : <http://www.yahoo.com>  
Magellan : <http://www.mckinley.com>  
Un annuaire intéressant pour les ressources académiques :  
Argus Clearinghouse : <http://www.clearinghouse.net>

Les adresses des annuaires francophones sont présentées p. 49.

Alors que les moteurs fournissent des adresses de pages HTML, les annuaires fournissent des adresses de sites (pages d'accueil des serveurs).

**Les ressources sont organisées hiérarchiquement en rubriques, sous-rubriques...** selon une classification donnée : classification de la Library of Congress, Mesh, Dewey ou autre classification propre à l'organisme.

## ■ Choix des sites explorés :

La majorité des ressources recensées sont déclarées par leurs auteurs et validées par les responsables du répertoire. Des collectes automatiques peuvent constituer un appui éventuel.

## ■ Contenus indexés :

Contrairement à l'approche retenue par les moteurs, seules certaines zones sont indexées : URL, titres, résumés éventuels fournis par l'auteur, commentaires des administrateurs, titres des rubriques thématiques de la classification.

## ■ Utilisation :

L'utilisateur peut :

- soit naviguer à partir des rubriques de la classification offerte,
- soit formuler des équations de recherche qui porteront sur toute la base de données ou une partie de celle-ci seulement. Il faut aussi comprendre la syntaxe des requêtes pour contrôler sa recherche.

## ■ Avantages et inconvénients :

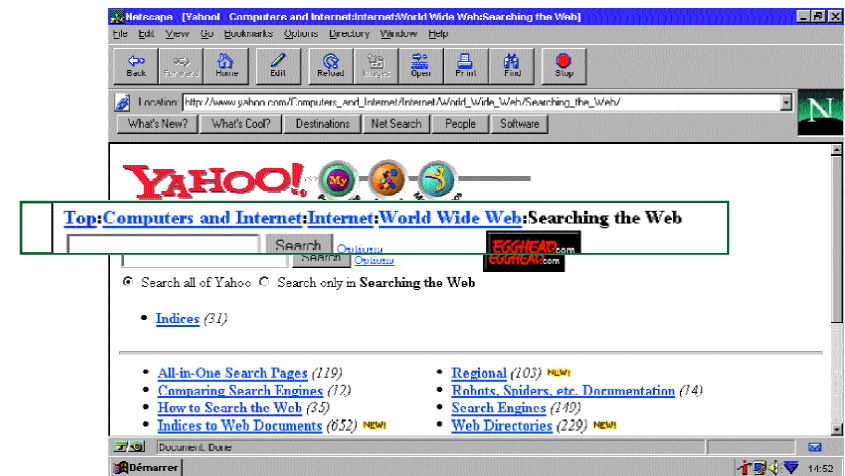
L'intérêt majeur de cette approche est la valeur ajoutée humaine (contrôle, validation, organisation) qui permet de limiter le bruit dans les réponses. La classification des ressources favorise des **recherches de type exploratoire sur une thématique générale**. De plus, ces outils pointent en majorité sur des pages d'accueil de sites et non sur n'importe quelle page.

Par contre, la couverture ne peut évidemment pas être aussi importante que celle des outils automatiques.

## Répertoire thématique yahoo! versions anglaise et française :

Les rubriques ou sous-rubriques à parcourir pour aboutir aux ressources consacrées à la recherche d'information sur le Web.

Version américaine : [http://www.yahoo.com/Computers\\_and\\_Internet/Internet/World\\_Wide\\_Web/Searching\\_the\\_Web/](http://www.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Searching_the_Web/)



Version française : [http://www.yahoo.fr/informatique\\_et\\_multimedia/Internet/World\\_Wide\\_Web/Recherche\\_sur\\_le\\_Web](http://www.yahoo.fr/informatique_et_multimedia/Internet/World_Wide_Web/Recherche_sur_le_Web)



# Exemple de recherche avec l'annuaire thématique français NOMADE

Recherche du site du **Ministère de l'éducation nationale**  
**1<sup>ère</sup> option** : Parcours à partir des rubriques et sous-rubriques

**services et points d'entrée complémentaires, presse, météo, sélections de sites, annuaires...**

**exploration à partir de la rubrique ETAT et SERVICE PUBLIC**

**sélection du mode avancé**

**rubriques thématiques proposées**

**première sous-rubrique: précision du pays FRANCE**

**choix de la rubrique suivante: MINISTÈRES ET ADMINISTRATIONS**

**lancement de la requête**

**zone de saisie des mots-clés**

**Partenaire**  
Pages Jaunes

**Recherche**  
E-mail  
Web (options)

**L'info en direct**  
Actu 24/24 (AFP)  
Météo  
Bourse

**Nomade's Land**  
Nouveautés  
Sélections  
Dossiers

**Ajouter votre site**

**Actualité, presse**  
Echos, Journaux...

**Arts et culture**  
Cinéma, Musique, Musées, Mode...

**Enseignement et emploi**  
Formation...

**Informal**  
Jeux, Loisirs, Internet...

**Loisirs et tourisme**  
Spectacles, Sports, Voyages perso...

**Science et technologie**  
Physique, Biologie, Ingénierie...

**Sciences humaines et sociales**  
Religions, Histoire, Droit...

**Vie politique et sociale**  
Francophonie, Ecologie...

**Web pratique**  
Annonces, Rencontres, Achats...

**Etat et service public**  
Administrations, Régions, Europe...

**Famille et santé**  
Enfants...

Découvrez les guides européens partenaires de Nomade: [Virgilio](#) (Italie) et [Web.de](#) (Allemagne)

[Aide](#) | [Nomade?](#) | [Nous contacter](#) | [Ajouter votre site](#) | [Publicité](#) | [Nos partenaires](#)

© 1997 Obicsoft Net  
Avertissement

Nomade tourne sous Microsoft Book Office, et est accessible par [Orbicon](#)

http://www.nomade.com

**sous-rubrique suivante : EDUCATION NATIONALE**

**affichage des résultats**

**Recherche**  
E-mail  
Web (options)

**L'info en direct**  
Actu 24/24 (AFP)  
Météo  
Bourse

**Nomade's Land**  
Nouveautés  
Sélections  
Dossiers

**Ajouter votre site**

**Etat et service public : FRANCE : Ministères et administrations**

- Ambassades, Consulaires (69)
- Département officielle (112)
- Préfectures (9)
- Etablissements publics (58)
- Fonction publique (6)

- Gouvernement (9)
- Ministères et administrations (118)
- Parlements (9)
- Enseignement (2)
- Régions, départements, communes (206)

**Etat et service public : Ministères et administrations : Education Nationale**

- Recherche (2)

**Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche**

MENESR - Public - Paris, France  
 Le serveur officiel du Ministère de l'Education Nationale, de l'enseignement supérieur et de la recherche. Les Etats généraux de l'Université, la violence à l'école, le cabinet du ministre, une présentation de certains dossiers, etc.  
<http://www.men.fr>

**Direction de l'évaluation et de la Prospective**

Ministère de l'Education Nationale - Créteil - Paris, France  
 actualité des publications de la DEP, données issues de ses services de production statistique.  
<http://www.men.fr>

**2<sup>ème</sup> option** : à partir des mots-clés en mode avancé (bouton options)

**zone de saisie**

**critère supplémentaire: FRANCE**

**opérateur ET sélectionné**

**affichage des résultats**

**Recherche**  
E-mail  
Web (options)

**L'info en direct**  
Actu 24/24 (AFP)  
Météo  
Bourse

**Nomade's Land**  
Nouveautés  
Sélections  
Dossiers

**Ajouter votre site**

**Web (options)**

Tapez le(s) mot(s) clé(s):  
 ministère éducation

**Cherchez les sites avec:**

- tous les mots (et)
- au moins un des mots (ou)
- la racine d'un mot
- le mot exact

**Alignez votre recherche en définissant un ou plusieurs critères:**

Région (France):

Pays:

Nature:

Ville:

Public

**Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche**

MENESR - Public - Paris, France  
 Le serveur officiel du Ministère de l'Education Nationale, de l'enseignement supérieur et de la recherche. Les Etats généraux de l'Université, la violence à l'école, le cabinet du ministre, une présentation de certains dossiers, etc.  
<http://www.men.fr>

**Ministère de l'Education nationale, de l'Enseignement supérieur et de la Recherche**

MENESR - Public - Paris, France  
 Ce site, adressé à EduNet, est la base de données sur l'éducation et l'enseignement. Le calendrier scolaire, les examens, informations sur l'administration, les établissements sous tutelle, les organismes de recherche...  
<http://www.men.fr>



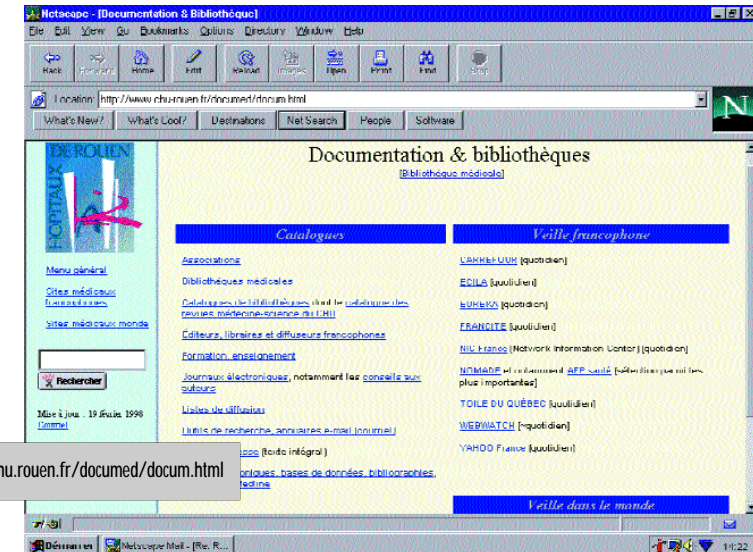
## Les sites et outils spécialisés

Dans la même logique de recensement que celle des annuaires, on trouve des sites

spécialisés sur un certain type de ressources, par exemple :

Annuaire géographique	ex : <a href="http://www.tue.nl/europe/">http://www.tue.nl/europe/</a>
Répertoires de forums électroniques	ex : <a href="http://n2h2.com/KOVACS">http://n2h2.com/KOVACS</a>
Répertoires de revues électroniques	ex : <a href="http://www.edoc.com/ejournal">http://www.edoc.com/ejournal</a>
Annuaire et robots pour une discipline	ex : Histoire médiévale <a href="http://argos.evansville.edu">http://argos.evansville.edu</a>
Pages Web spécialisées (académiques ou professionnelles)	ex : Médecine : <a href="http://www.chu-rouen.fr">http://www.chu-rouen.fr</a> Droit : <a href="http://www.rabenou.org">http://www.rabenou.org</a>

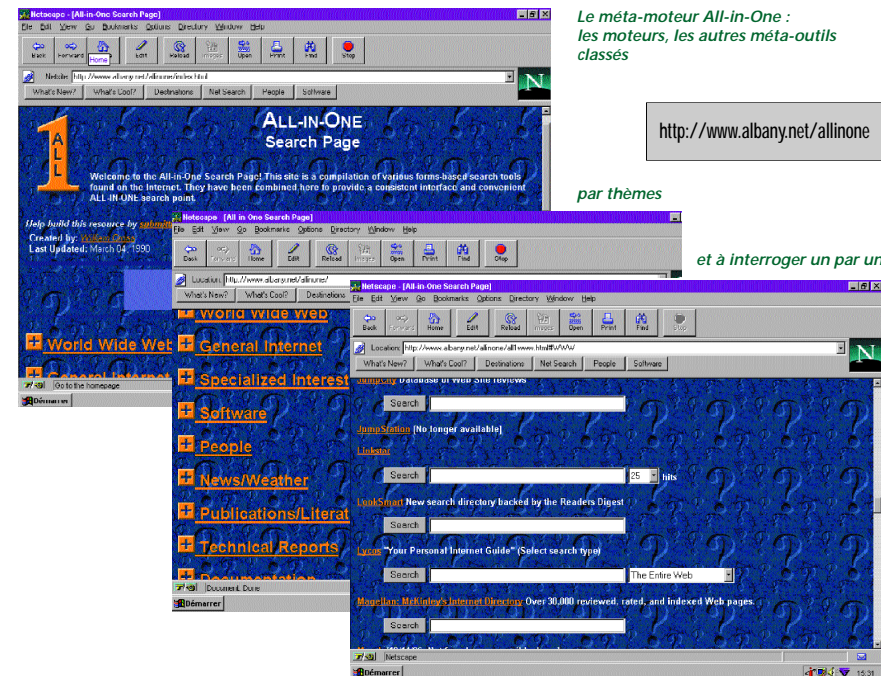
## Exemple de pages web spécialisées



Le serveur WEB du Centre Hospitalier Universitaire (CHU) de Rouen pour les ressources en médecine.

<http://www.chu.rouen.fr/documed/docum.html>

## Exemple de méta-moteur



Le méta-moteur All-in-One : les moteurs, les autres méta-outils classés

<http://www.albany.net/allinone>

par thèmes

et à interroger un par un

## Les méta-outils

Comme cela a déjà été dit plus haut, il est souvent nécessaire de recouper les résultats fournis par plusieurs outils de recherche. Les méta-outils simplifient cette démarche en regroupant plusieurs moteurs et/ou annuaires qu'ils permettent d'interroger :

### ■ Simultanément :

Une requête est envoyée à plusieurs moteurs et/ou annuaires simultanément. Le méta-outil peut se charger d'écartier les doublons, de trier globalement les réponses, de vérifier la validité des liens et enfin, de fusionner en une liste synthétique les résultats obtenus.

### ■ Un par un uniquement :

L'utilisateur bénéficie d'un accès à un ensemble d'outils classés par thèmes, types de services ... et de la possibilité de comparer les résultats d'interrogation qui peuvent se révéler sensiblement différents.

### Exemples :

SavvySearch développé à l'Université du Colorado : <http://guaraldi.cs.colostate.edu:2000/>

MetaCrawler développé à l'origine par l'université de Washington : <http://metacrawler.com>

### Exemples :

- All-in-one développé par AlbanyNet : <http://www.albany.net/allinone>

- Internet Sleuth développé par Internet Business Connection : <http://www.isleuth.com>

# Les différents outils et services de filtrage de l'information

Les outils et services de filtrage de l'information ont pour objectif d'aider l'utilisateur à effectuer une veille permanente sur ses différents centres d'intérêt. Ils permettent, en effet, d'extraire automatiquement parmi des ensembles d'informations (pages web, dépêches de presse, messages, banques de données...) celles qui correspondent à un profil défini, selon une périodicité donnée. Certains de ces outils d'assistance à la recherche d'information ou de diffusion automatique appartiennent à la famille de technologies que l'on désigne sous le terme de « agents intelligents ».

Parmi ces outils et services de filtrage, on trouve :

■ **des logiciels à installer sur son micro-ordinateur** qui vont effectuer des tâches similaires aux méta-outils du type de Savvy-Search (cf ci-dessus) et assumer aussi des fonctions telles que la recherche de nouvelles pages en fonction de profils définis, la mise à jour automatique et le téléchargement de celles-ci en local...

Exemples :  
logiciels DigOut4U de la société ARISEM : <http://www.arisem.com>,  
Netattaché de Tympani : <http://www.tympani.com>

■ **des services distants** qui effectuent des sélections de sites Web ou d'autres types d'informations (les news par exemple) selon des profils définis. L'utilisateur se connecte sur le site pour visualiser les informations sélectionnées pour lui :

Exemple :  
The Informant : <http://informant.dartmouth.edu>

■ des services identiques au précédent mais avec **une diffusion automatique** sur le poste de l'utilisateur. On désigne cette pratique sous le terme de « **push** » : l'information est automatiquement poussée vers l'utilisateur dès qu'une mise à jour du service auquel il est abonné est effectuée. La presse et la télévision (Webcasting, chaînes d'informations) ont immédiatement investi ce créneau mais les services et les applications sont en train de se multiplier (revues de presse et journaux personnalisés, chaînes thématiques, météo, cotations boursières, mises à jour de versions logicielles...). Ce principe peut se décliner dans toutes sortes de contextes pour faciliter la diffusion sur mesure d'informations.

Exemple :  
le service Rapidus, mis en place sur le serveur Cordis de la Commission des Communautés européennes qui offre la diffusion sur profil des annonces d'appels d'offre de la Commission : <http://www.cordis.lu>

## Exemples d'outils francophones

SITE	ADRESSE	TYPE D'OUTIL	REMARQUES
Lokace (Internet Plus)	<a href="http://www.lokace.com">http://www.lokace.com</a>	Moteur de recherche Annuaire	Propose aussi un accès thématique. Recherche d'adresses e-mail francophones
Francité (Société I3d)	<a href="http://www.i3d.qc.ca/francite.htm">http://www.i3d.qc.ca/francite.htm</a>	Moteur de recherche Annuaire	Outil québécois avec un service de recherche sur plusieurs outils simultanément.
Ecila (Ecila)	<a href="http://France.ecila.com">http://France.ecila.com</a>	Moteur de recherche	Indexe le Web français
AltaVista (Digital Equipment)	<a href="http://altavista.telia.com">http://altavista.telia.com</a>	Moteur de recherche	la limitation aux documents en langue française est également paramétrable.
Yahoo! France (Yahoo Corp.)	<a href="http://www.yahoo.fr">http://www.yahoo.fr</a>	Annuaire	même interface que Yahoo! Accès aux dépêches AFP, Reuters.
UREC (CNRS)	<a href="http://www.urec.fr">http://www.urec.fr</a>	Annuaire thématique	L'un des premiers annuaires des serveurs WEB français. Désormais, orientation enseignement et recherche uniquement.
Nomade (Nomade)	<a href="http://www.nomade.com">http://www.nomade.com</a>	Annuaire	Très similaire à Yahoo ! France. Accès aux dépêches AFP.
QuiQuoiOu (France-Télécom)	<a href="http://www.wanadoo.fr">http://www.wanadoo.fr</a>	Annuaire d'adresses et Moteur de recherche	Intégration de ce service avec d'autres produits de France Télécom : accès aux pages jaunes et aux pages blanches, aux services Minitel...

*Il existe différents outils consacrés aux ressources en langue française. On trouve à la fois des moteurs et des annuaires.*

*A noter aussi, le développement de versions françaises d'outils tels qu'AltaVista et Excite.*

## Quelques conseils pour terminer

Les conseils généraux présentés dans la partie méthodologie p.12 s'appliquent bien évidemment à Internet. Cependant, quelques conseils plus spécifiques sont nécessaires :

■ Bien préparer sa recherche en définissant le type de ressources utiles et le type d'acteur qui est susceptible de produire et mettre à disposition l'information recherchée facilitera le choix du point de départ :

- moteurs ou annuaires
- autres points d'entrée qui ont été présentés dans le chapitre « des ressources diversifiées »
- ou encore serveur qui dessert les différentes communautés auxquelles on appartient : serveur de l'université, de la bibliothèque, serveurs disciplinaires... Les réalisateurs auront certainement recensé d'autres sites utiles pour leurs utilisateurs.

■ Prendre connaissance des particularités des différents moteurs, annuaires, pour bien évaluer leur mode de fonctionnement. Attention, de plus en plus souvent, des moteurs de recherche et des annuaires sont présentés sur la même page. Il ne faut pas confondre les formulaires d'interrogation des uns et des autres et il est nécessaire de prendre la mesure du champ couvert.

■ Vérifier les qualifications de l'auteur de la page, du site : sur Internet, tout un chacun pouvant publier, il est important de pouvoir cerner la compétence de l'auteur, son organisme de rattachement, ses sources, la date du document, de ses mises à jour...

- Pour évaluer un serveur Web : on peut, après avoir interrogé un moteur de

recherche, le comparer à d'autres sites qui traitent des mêmes sujets afin de comprendre ses orientations, l'étendue de la couverture du sujet...

- Avec les commandes avancées de certains moteurs, on peut également cerner l'importance du site par le nombre de pages Web qui ont établi des liens avec lui.

Par exemple : sur Alta-Vista, la commande [link:pasteur.fr](#) permettra de retrouver les pages avec [pasteur.fr](#) dans l'URL d'un de leurs liens.

- On peut encore vérifier qu'il a été sélectionné et validé dans le cadre d'un répertoire thématique.

■ Etre vigilant sur l'authenticité de la source et sur les problèmes de droit :

- sites pirates, problème de l'utilisation des données privées etc. Le site de la CNIL donne des illustrations très claires sur ces aspects : <http://www.cnil.fr>.

- la diffusion libre sur le réseau ne veut pas dire que l'on peut réutiliser les données, images et autres documents sans se préoccuper des règles de la propriété intellectuelle. Un autre site recense tout les cas de figure que peuvent rencontrer étudiants, enseignants et chercheurs : <http://www.jm.u-psud.fr/recherch/multimed/index1.html>.

■ S'organiser pour optimiser la gestion de l'information : enregistrement dans un répertoire (bookmark) des sites favoris, tri, classement, procédures de mises à jour (avec des logiciels tels que ceux décrits plus hauts), exploitation de l'information...