

Eric Pichon

# LA RECHERCHE D'INFORMATION SUR INTERNET

2003 

[\[Télécharger le support de cours en pdf\]](#)

[Introduction](#)



## Stratégie de recherche :

[Pour qui cherche-t-on ?](#)

[Où chercher ?](#)

- [sur internet ou ailleurs ?](#)
- [sur le web ou ailleurs ?](#)
- [la recherche sur le web : quels outils ?](#)

[Comment chercher ?](#)

[Que cherche-t-on ?](#)

[Vérifier l'information](#)

[Se tenir au courant : la veille documentaire](#)



## Panorama des outils :

[Moteurs, annuaires, portails](#)

- [les moteurs : pertinence , syntaxe](#)
- [les métamoteurs](#)
- [les annuaires](#)

[Les outils spécialisé par domaine : exemples , les trouver](#)

[Le web "invisible" et les bases de données](#)

[Rechercher par type d'information \(mél, fichiers, ...\)](#)

[La recherche sur les forums et les listes de diffusion](#)

[L'évolution des outils](#)

# LA RECHERCHE SUR INTERNET

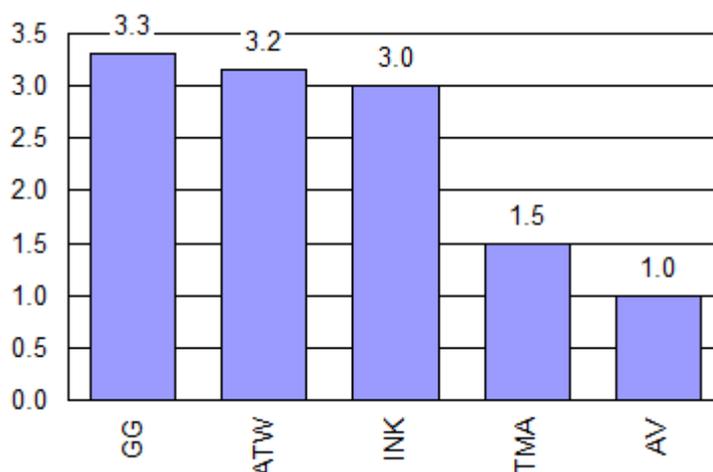
à jour le  
7 sept. 2003

**Il est à la fois simple et difficile de retrouver de l'information sur internet.**

**Simple**, parce qu'il existe des outils qui permettent de le faire et qui sont assez faciles à utiliser (moteurs indexant des sites ou annuaires mis à jour par des êtres humains)

**Difficile**, parce qu'aucun outil ne peut recenser toute l'information qui circule sur le réseau :

- il n'y aucune centralisation, pas d'organisme chargé du "dépôt légal" (cf le cours [introduction à internet](#))
- le nombre de sites est en progression constante et aucun moteur ne peut toutes les indexer, bien que leurs capacités s'accroissent de façon considérable



( <http://www.searchenginewatch.com/reports/> )

- le contenu de certaines pages recensées change, ce qui rend l'indexation rapidement obsolète  
(cf le contenu de la page du conseil des ministres <http://www.premier-ministre.gouv.fr/fr/p.cfm?ref=381> renouvelé tous les mercredis)
- de plus, il existe un web "invisible" qui ne peut pas, du fait de sa structure, être consulté par les moteurs de recherche "classiques")

Et si un recensement exhaustif était possible, les moteurs devraient évoluer significativement afin d'améliorer de façon profonde les critères de pertinence...

typologie des outils de recherche

En outre, un très grand nombre d'informations diffusées sur internet ne le sont pas par le web mais par le biais de la messagerie ou des serveurs de forums.

---

<< Vous estimez-vous satisfait des résultats proposés par les outils de recherche du Web lors de vos requêtes ? ", les internautes ayant participé au sondage semblent manifester un certain mécontentement : 81,83% ont répondu "Non", jugeant donc que leurs quêtes ne sont pas satisfaites, tandis

Eric  
Pichon

que 13% ne sont que moyennement comblés. Seuls 5% des internautes estiment qu'il trouvent l'info recherchée sans problèmes. Il y a donc encore du boulot dans le domaine des outils de recherche. La question que l'on peut se poser, à la lecture de ces chiffres est : "est-ce que ce sont les outils de recherche qui ne sont pas assez performants ou sont-ce les internautes qui ne les utilisent pas toujours au maximum de leurs capacités ?" >>

[abondance.com](http://abondance.com)

# POUR QUI CHERCHE-T-ON ?

à jour le  
7 sept. 2003

Bien évidemment, la première question à se poser avant d'entamer une recherche est de savoir

- à qui seront destinés les résultats ;
- ce que le demandeur détient déjà comme information ;
- ce qu'il espère obtenir de plus ;
- et sous quelle forme (simplement des adresses, un dossier, ...)

Ce n'est que muni de ces éléments que l'on sera en mesure de savoir quoi, où et comment chercher.



---

Eric  
Pichon

# QUE CHERCHE-T-ON ?

à jour le  
7 sept. 2003

Quelques exemples de ce que qu'il faut déterminer dans la question posée. Les pistes de réponses sont ici uniquement orientées sur internet, mais dans la pratique, il ne faut pas oublier qu'on a à sa disposition d'autres outils (usuels, dossiers de presse) cf [où chercher ?](#)

---

[Typologie des outils de recherche sur internet](#)  >>>

---

## Une synthèse ? Des chiffres ?

Pour un sujet d'actualité, tenter les fils d'actualités des annuaires (Yahoo!,...).

Dans un annuaire, chercher s'il existe une rubrique "statistiques" en rapport avec le sujet.

## Des documents originaux ?

Essayer les bases de données dans le domaine.

Existe-t-il un site "officiel" sur le sujet. Offre-t-il des documents téléchargeables ? (centrer votre recherche sur le format [pdf](#), par ex. sur Google ou HotBot)

## À faire partager un point de vue ?

Trouver un forum ou une liste de discussion approprié.

## Un spécialiste à interviewer ?

Essayer de repérer à partir d'un portail les auteurs de sites sur le sujet ; des enseignants du domaine sur des sites universitaires ; des journalistes spécialisés ; un annuaire sur le site d'une association professionnelle ; ...

## Une analyse "objective" ?

Sites de journaux ; répertoires de thèses universitaires ; ...

## Un point de vue "partisan" ?

### **officiel ou au contraire dissident ?**

Sites officiels des gouvernements, des sociétés ([Google](#))

vs. discussion des forums, ou sites alternatif ([L'autre portail](#))

[Les annuaires et les moteurs](#)

[Les portails](#)

[Les bases de données](#)

[Les forums et les](#)

[listes de diffusion](#)

## Une anecdote à placer dans un discours ?

[Quid](#), [Citations du monde](#), ...

## Un dossier complet ?

Les outils spécialisés

- [par type d'information](#)
- [par domaine](#)

[Plusieurs outils combinés](#) ;  
[portails du domaine](#)

## Toute la chronologie d'un événement ?

Archives des journaux (par [World News](#), par ex.) ;

## Ses derniers rebondissements ?

Fils d'actualité (par [news.yahoo.fr](#), par ex.)

[Sélection de sites](#)

## Recherches particulières

Adresses de personnes ; logiciels ; images ; cartes : consulter [les outils spécialisés par types de documents](#)



Eric  
Pichon

---

Se demander à quoi devrait ressembler la réponse idéale

# OÙ CHERCHER ?

à jour le  
7 sept. 2003

## Sur Internet ou ailleurs ?

Chercher sur internet n'est pas une fin en soi...La réponse peut se trouver en bibliothèque, ou auprès d'autres personnes où organismes

*...que l'on pourra d'ailleurs repérer à l'aide d'internet...*

## Gratuit ou payant ?

"Imaginez une chambre d'enfant. Vous poussez la porte. Le lit est fait. Tous les jouets sont bien rangés. Vous ouvrez le tiroir de la commode. Toutes les chaussettes de votre gamin s'y trouvent, appariées et rangées par couleurs.

Imaginez maintenant une deuxième chambre. Un jouet bloque la porte. Après quelques efforts vous arrivez à l'ouvrir. Vous vous frayez un chemin en écrasant des poupées Barbie et ouvrez le tiroir. Il est plein de chaussettes dépareillées. Vous prenez une chaussette bleue. Après un quart d'heure vous trouvez enfin la deuxième sous le lit. Votre gamin peut enfin partir à l'école !

La première chambre, c'est le monde des bases de données en ligne. Tout y est rangé par des professionnels : des documentalistes. Certes, on paye, mais on gagne du temps.

La deuxième chambre est le monde d'Internet. C'est gratuit, mais personne ne se soucie d'y apporter de l'ordre."

( extrait de l' [analyse de Y.M. Marti dans Qwam.com](#) )

Désormais, la plupart des bases de données **payantes** sont accessibles via le web.

cf <http://bdd.formist.enssib.fr/index.html> 

On peut également trouver des informations dans des bases de données **gratuites**.

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

Il est intéressant de connaître les bases de données disponibles dans son domaine, car elles recensent des informations mal repérées par les outils de recherche généralistes sur internet (on parle de [web invisible](#)). L'important est de connaître comment sont constituées les unes et les autres (couverture, mise à jour). 

# OÙ CHERCHER ? .2 L'INFORMATION DISPONIBLE GRATUITEMENT SUR INTERNET

à jour le  
7 sept. 2003

## Chercher dans des sites web où ailleurs ?

L'interface "web" tend à unifier les différentes composantes de l'internet (cf le cours [introduction à internet](#)) on pourra cependant distinguer différentes sources d'information :

- Forums (usenet)
- Listes de diffusion et lettres d'information
- Sites web (à contenu "éditorial")

 [Forums et](#)  
[Listes de](#)  
[diffusion](#)

Eric  
Pichon

# OÙ CHERCHER ? .3 LA RECHERCHE SUR LES SITES WEB

à jour le  
7 sept. 2003

Pour rechercher une information dans les sites web, il existe 2 grands types d'outils : les annuaires et les moteurs.

## Chercher dans un annuaire ou un moteur ?

*"En règle générale, on peut dire qu'on utilisera un annuaire comme Yahoo! ou Nomade pour rechercher un site général sur un sujet donné. On utilisera, en revanche, un moteur de recherche comme AltaVista ou Voila pour effectuer une recherche sur un sujet beaucoup plus précis, plus fin"* (Andrieu)

Se rappeler que :

- les annuaires identifient des **sites** organisés par **catégories**
- Les moteurs conduisent à des **pages** triées en fonction de votre requête

---

[En savoir plus sur les annuaires et les moteurs>>>](#)

---

## Dans le domaine francophone ou ailleurs ?

Pour une recherche plus vaste pensez à traduire vos termes de recherche en anglais.

Si les termes de votre recherche sont en français, vous aurez de fortes chances de trouver des pages en français. Si les termes employés sont des noms propres ou multilingues, vous pouvez restreindre votre domaine de recherche.

Dans les annuaires français (.fr) la plupart des sites répertoriés sont francophones.

Dans les moteurs, voir les options : recherche sur tout le web, francophone, français

Penser aussi aux outils de traduction : [tr.voila.fr](http://tr.voila.fr) par ex.

## Avec un outil généraliste ou spécialisé ?

Pour une recherche dans un domaine spécialisé (droit, économie, littérature) ou un type d'information particulière, il est souvent plus rapide d'aller chercher l'information directement dans un annuaire ou un portail

spécialisé. Cela permet une réponse plus riche et évite les confusions de termes.

---

[En savoir plus sur les outils spécialisés par domaine >>>](#)

[En savoir plus sur les outils spécialisés par type d'information >>>](#)

[Outils](#)

---

Eric   
Pichon

Smartborg vous oriente vers différents outils de recherche en fonction de la question posée)

# COMMENT CHERCHER ?

à jour le  
7 sept. 2003

Savoir pour qui on cherche et ce que l'on cherche.

## 1. Déterminer le contexte du sujet

Généralement, on peut dire que l'on cherchera dans un annuaire pour un sujet général, dans un moteur pour un sujet plus spécifique, MAIS :

- La distinction se fera en fonction de votre propre connaissance du sujet.

Ex. : si l'on cherche des documents sur la végétation de l'edelweiss, on peut avoir intérêt à chercher dans un annuaire s'il existe des sites spécialisés sur la flore alpine (cela suppose que l'on sait que l'edelweiss est une fleur alpine)

- Si vous connaissez mal le sujet, procédez par exploration : essayez d'interroger un **moteur pour identifier le contexte** puis utilisez un **annuaire pour chercher des sites spécialisés** sur le sujet.

## Outils intéressants pour ce défrichage

- un métamoteur
- des outils qui reformulent votre question (en anglais : [Ask Jeeves](#) 🌐)
- ou qui vous proposent des termes pour préciser votre recherche (en anglais : [Alltheweb](#) , [Teoma](#), en français : [Netscape.fr](#) 🌐 , [Altavista.fr](#) 🌐)

## 2. Définir les outils à utiliser

En fonction de la question posée et de la **réponse attendue**.

cf que cherche-t-on ? >

## 3. Déterminer les termes de sa requête

- Faire ressortir les termes les plus significatifs de la question.
- Penser aux synonymes, aux équivalents anglais

Il peut être utile de jeter un coup d'oeil sur les premières réponses trouvées : **regarder les termes employés** dans le texte et dans l'en-tête (balises META : cf [structure d'une page html](#)) et **les réutiliser pour reformuler sa question**.

# VÉRIFIER L'INFORMATION

à jour le  
7 sept. 2003

- Dans un [annuaire](#), les sites ont été sélectionnés ;
- dans un [répertoire spécialisé](#), ils ont été en principe retenus par un spécialiste du domaine ;
- La sélection proposée par un répertoire spécialisé est d'autant plus fiable qu'il lui-même bien évalué dans un "[répertoire de répertoires](#)"
- En revanche, un [moteur](#) n'évalue pas les pages qu'il vous propose en réponse à une question (même si le fait qu'elle soit très souvent citée peut être un critère de pertinence, par ex dans Google)

Avant d'utiliser ou de répertorier un site, vous devrez identifier :

## Le domaine

Le nom de domaine peut (faiblement...) vous aider.

[liste et significations des noms de domaine](#) 🌐

## La fraîcheur de la mise à jour

- vérifier sur la page qui vous intéresse si le webmestre a indiqué la date de mise à jour.
- avec Netscape, on peut également utiliser le menu : "Afficher" "Infos page"
- dans leurs résultats, certains moteurs ([Voila.fr](#), par ex.) mentionnent également la date de mise à jour.

## La source

Retrouvez la **page d'accueil** pour en savoir plus sur le site.

### S'il s'agit d'une page personnelle :

(--.free.fr, www.chez.com/--, etc)

- quelles sont les compétences de l'auteur de cette page ? (cv) ,
- est-il possible de le joindre (adresse mail)?
- participe-t-il à des [forums](#) ? lesquels ?
- sa page est-elle souvent citée par ailleurs ? (recherche de type **link**: sur un moteur)

[Le projet Google WebQuotes](#) annote les résultats de votre recherche avec des commentaires issus d'autres sites.

### S'il s'agit d'une page institutionnelle :

L'utilitaire [WhoIs \[par ex. sur Nic France\]](#) 🌐 peut vous permettre d'identifier l'origine d'un site. 🇫🇷

Quelle est la vocation de cette institution ? Quel est son "intérêt" à publier ces informations sur le web:

- rechercher sur son nom
- rechercher les sites qui pointent vers le sien (recherche de type **link**: sur un moteur)
- rechercher (prudemment...) les informations sur cette société dans les [forums](#).

**Exemple de démarche analytique d'un site :**

<http://users.skynet.be/ameurant/francinfo/validite/critique1.html>

Un site contre (?)  
les rumeurs du net  
[Hoaxbuster](#)

---

Eric  
Pichon

>>> voir aussi : <http://system.ensmp.fr/cours/evalweb>

# SE TENIR AU COURANT : LA VEILLE DOCUMENTAIRE

à jour le  
7 sept. 2003

L'activité de documentaliste ou de bibliothécaire exige d'être constamment informé(e) des évolutions d'un site ou d'un sujet donné.

Il existe des outils qui permettent assez facilement de faire de la veille :

## Etre informé de l'évolution d'un site

Vous pouvez "surveiller" des pages précises de sites pour être alerté de leur mise à jour. Des outils téléchargeables peuvent vous informer par mail sur la mise à jour des pages sélectionnées :  
ex. Check&get [www.activeurls.com](http://www.activeurls.com) 🌐

Les [aspérateurs de sites](#), qui permettent de naviguer hors connexion, ont aussi des fonctions de mise à jour des pages modifiées,  
par ex. eCatch [www.ecatch.com](http://www.ecatch.com) 🌐 " surligne" les modifications.

Pensez aussi à vous abonner aux [lettres d'information](#) des sites qui vous intéressent.

## Etre informé des nouvelles réponses à une question

Il est possible de mettre régulièrement à jour une requête, automatiquement ou à la demande :

### Outils en ligne

Certains outils en ligne offrent désormais des fonctions de veille.

**Google alert** (<http://www.googlealert.com> , indépendant de Google) vous avertit par courriel des nouvelles réponses trouvées par Google à une question enregistrée par vos soins.

**Yahoo** (<http://alerts.yahoo.com>) : alerte sur l'actualité, l'horoscope, ...

**Google** (<http://www.google.com/newsalerts>) propose une alerte sur l'actualité

### Outils téléchargeables

D'autres outils, téléchargeables, combinent tout ou partie des fonctionnalités de métamoteurs, de recherche sur le web invisible, d'aspérateur de sites et d'alerteur : on parle d'"agent intelligents"

Exemples :

**Copernic** <http://www.copernic.com/fr/> -

**Strategic finder** <http://www.digimind.fr/sf.htm>

- métamoteurs,
- trient les informations trouvées,
- suppriment les doublons et les réponses obsolètes,
- peuvent mettre à jour les requêtes , automatiquement ou à la demande
- recherche spécialisées par domaines et par types d'information.

En savoir plus :

L'intelligence artificielle n'est pas pour demain et l'on en est encore très souvent à "chercher une aiguille dans une botte de liens" (titre d'un article de la Recherche, 2/00) mais dans ce domaine les progrès vont vite... surveillez de près l'[évolution des outils de recherche](#) !

---

[Télécharger des logiciels](#)

Eric  
Pichon

Sur les outils de veille et les agents intelligents : [agentland.fr](http://agentland.fr) 🌐

# MOTEURS, ANNUAIRES, PORTAILS

à jour le  
7 sept. 2003

## Définition

"[Moteurs](#)" et "[annuaires](#)" sont des sites permettant d'accéder à d'autres sites

Les différences :

- Le mode d'indexation
- Le nombre de documents référencés
- Le mode d'accès à l'information

Mais ces caractéristiques tendent à cohabiter au sein des mêmes sites

La plupart de ses sites tendent à se constituer en "**portails**" : outre un outil de recherche, ils proposent des liens directs vers des sites partenaires (cybercommerce...) et offrent différents services (adresses mail gratuites, ...) afin de fidéliser l'internaute.

Ceci peut entraîner une confusion entre les différents liens proposés : site "sponsor", page indexée automatiquement, ou fruit d'une sélection réfléchie et impartiale ? Il semble néanmoins que l'on s'achemine vers une plus grande distinction entre les "espaces" proposés.

On trouve aussi des "[méta-moteurs](#)" et des "[méta-portails](#)".

## A savoir ...

Les annuaires sont constitués par des indexeurs qui sélectionnent un certain nombre de sites : ils référencent donc beaucoup moins de sites que les moteurs qui indexent automatiquement, mais ces derniers le font de façon évidemment moins avisée...

Le mode de recherche, la syntaxe, les champs couverts varient suivants les outils : aussi il est plus important de bien maîtriser quelques-uns d'entre eux, plutôt que de les utiliser tous, mais incorrectement.

Pensez à consulter les pages d'aide ou les FAQ que la plupart de ces sites mettent à votre disposition.

Certains sites sont inaccessibles aux moteurs ([web invisible](#) : pages dynamiques, etc.)

**Pour en savoir plus ⓘ :**

[les moteurs](#)

[les annuaires](#)

[les outils spécialisés](#)

annuaire d'outils  
de recherche :  
[www.enfin.com](http://www.enfin.com)

# LES MOTEURS DE RECHERCHE

à jour le  
7 sept. 2003

## Principe des moteurs de recherche

Les moteurs de recherche sont des sites qui lancent des requêtes sur une base de données constituée de façon automatique d'après le contenu des pages recensées par les "spiders", outils qui parcourent régulièrement le web en se basant sur les liens hypertextes.

Les données contenues dans les pages rencontrées sont indexées et pondérées automatiquement, selon différents critères qui varient selon les outils :

- structuration du [code "source"](#) de la page (les termes reçoivent une pondération plus ou moins importantes selon les "zones" de la page où ils sont placés.
- "notoriété" de la page (déterminée par le nombre de liens hypertexte pointant vers celle-ci)
- fréquentation de la page.

Lorsque l'on interroge un moteur de recherche, c'est dans cette base préconstituée qu'il va puiser pour proposer une liste de réponses, triées en fonction de leur adéquation supposée (c'est à dire calculée) avec la question. Chaque moteur a sa propre [méthode de tri](#) - jalousement gardée...

## Limites

Les robots ne peuvent pas parcourir toutes les pages existantes : ils faut qu'au moins un lien les conduisent sur une page donnée (cf [la théorie du noeud papillon](#))

Délais de mise à jour (entre 2 visites d'un robot à une page web ) : de 1 jour à quelques semaines. Ainsi des bases très importantes peuvent contenir des informations périmées.

Certaines informations accessibles par requête leur échappent : cas de certaines pages "dynamiques" et du contenu des bases de données ([web invisible](#))

La structuration des pages HTML ne permet pas une indexation très poussée des documents.

---

[Des évolutions devraient permettre d'améliorer les performances des moteurs >>>](#)

---

# LES MOTEURS. 2, LA PERTINENCE DES RÉSULTATS

à jour le  
7 sept. 2003

Les moteurs se distinguent entre eux non seulement par les résultats trouvés, mais aussi par la façon dont ils évaluent la pertinence de ceux-ci. Les méthodes de tri des résultats combinent à des degrés divers selon les moteurs, les éléments suivants(\*):

**place du mot dans le document** : plus important s'il est dans le titre ou le début du texte.

**densité** : fréquence du mot dans la page (un mot qui figure 2 fois sur une page de 200 mots est plus "dense" qu'un mot qui figure 3 fois sur une page de 500 mots).

**rareté** : un mot rare dans la base de données du moteur sera mieux "considéré".

**expression et proximité** : plus les termes de la requête sont proches dans un document, mieux cette page sera classée ; les pages contenant l'expression exacte arrivent en tête.

**notoriété** : plus une page est citée (i.e. référencée par un lien hypertexte) par d'autres, plus elle est importante ; cette importance est accrue si les pages qui pointent vers elle sont elles-mêmes très citées.

**popularité** : le moteur augmente la pondération des pages qui ont été jugées pertinentes par l'internaute pour une question donnée (quels liens de la liste ont été visités et pendant combien de temps avant que l'internaute ne reviennent à la liste des résultats).

Le classement des moteurs de recherche dépend largement de la façon dont on formule sa question.
---

Eric  
Pichon

\*d'après : Recherche d'information sur internet / Jean-Pierre Lardy. -  
ADBS, 2001.

# LES MOTEURS. 3, SYNTAXE

à jour le  
7 sept. 2003

## Important

Chaque moteur de recherche a sa propre "base de données"

mais aussi ses propres règles de recherche, concernant notamment:

- la prise en compte des majuscules et des lettres accentuées
- le traitement des mots vides
- les opérateurs utilisables

**Pensez à consulter l'aide (guide, help, tips,...) de chaque moteur.**

>>> *tableaux comparatifs des fonctionnalités des principaux moteurs :*

<http://www.abondance.com/outils/comparatif.html>

Pensez à utiliser plus d'un terme dans votre requête, sinon il y aurait trop de "bruit".

La plupart des moteurs offrent une [recherche simple](#) et une [recherche avancée](#). ↗

## Recherche simple

La recherche simple permet généralement :

### une succession de terme séparés par des espaces

Les pages trouvées seront celles contenant ces mots, ensemble (les pages seront les premières dans le tri par pertinence) ou séparément (les pages ne contenant pas tous les termes seront moins bien classées - Google ne les présentera pas).

Les mots vides ne seront en principe pas pris en compte.

### une expression exacte

En plaçant les termes entre guillemets.

"chaîne de caractères"

### de forcer ou d'exclure la recherche sur certains termes

Généralement le + placé devant un terme permet de forcer la recherche sur ce terme (utile pour les mots considérés a priori comme "vides" par le moteur)

paris +10 +du

Le - placé devant un terme permet d'exclure ce terme de la recherche.



## Recherche avancée

La recherche avancée accepte généralement :

### Les opérateurs booléens

- ET (AND, +, "tout les mots", ...),  
SAUF (NOT, -, AND NOT, "termes exclus", ...)  
pour restreindre votre recherche ;
- OU (OR, "n'importe quel mot", "any word")  
pour élargir votre recherche (aux synonymes, par ex.);
- Les troncatures : `petrol*` trouvera pétrolière, petrol, pétrolifères  
(mais aussi pétrolette et pétroleuses...)

>>> sur le principe des opérateurs booléens et des troncatures

<http://www.bibl.ulaval.ca/vitrine/giri/mod3/booleen.htm>

### L'opérateur de proximité

NEAR suivi éventuellement d'un chiffre pour préciser la distance maximale entre deux termes.

### La recherche sur une partie précise du document

ex sur voila :

`keywords:internet` : le terme "internet" doit être trouvé dans les mots-clés ( `<META name="keywords">`) de la page.

`title:actualites` : le terme "actualites" devra figurer dans le titre (`<TITLE>`) de la page

`url:voila` le terme "voila" devra figurer sur l'adresse internet de la page

### Des restrictions dans le temps et l'espace

recherche des pages mises à jour depuis telle date, appartenant à tel domaine (.gouv.fr, par ex), etc.

# LES METAMOTEURS

à jour le  
7 sept. 2003

Les **métamoteurs** sont des outils qui peuvent lancer la même requête simultanément dans plusieurs moteurs de recherche mais dans ce cas, la syntaxe est réduite au plus petit commun dénominateur.

Les métamoteurs sont surtout intéressants par la façon dont ils présentent les résultats, par exemple :

- cartographie : [Kartoo](#)
- organisation en dossiers ("clusterisation"): [Vivisimo](#)

D'autres outils, basés sur un métamoteur, ont également des fonctions de [veille et d'alerte](#) (Copernic).

---

Eric  
Pichon

[< méthode](#)

# LES ANNUAIRES

à jour le  
7 sept. 2003

---

[Moteurs, annuaires, portails : vue d'ensemble](#) 

---

## Principe des annuaires

"Les annuaires se présentent tous sous forme de grandes rubriques (Art et Culture, Sciences ...), elles-mêmes déclinées en sous-rubriques ( ...) proposant ainsi une organisation en arborescence de plus en plus fine.

Le travail de classement et d'indexation est fait par des humains.

La plupart des annuaires offrent 2 possibilités de recherche : recherche par thème et recherche libre (...) sur tout l'annuaire ou sur l'une de ses catégories.

Les annuaires ont tendance à devenir des outils hybrides : il est possible de rechercher sur l'annuaire mais également de déclencher une recherche sur le web via un moteur intégré." Il faut donc faire très attention à la partie du site que l'on interroge (l'annuaire, le moteur) et aux réponses trouvées (sites sélectionnés ou pages web de la base de données du moteur, lien sponsorisés...)

### Exemples

[Yahoo](#)  
(partie annuaire),

[Nomade](#),

[OpenDirectory](#)

(d'après : centre de documentation de l'ENA)

---

Eric  
Pichon

[<méthode](#)

# LES OUTILS SPÉCIALISÉS PAR DOMAINE .1

à jour le  
7 sept. 2003

Pour rechercher de l'information sur un domaine précis, on peut avoir besoin de recourir à des sélections plus pointues que celles proposées par les outils généralistes.

Il existe beaucoup de répertoires qui recensent uniquement des sites spécialisés sur un domaine de connaissance particulier (droit, éducation, ...) Ces sites offrent en outre l'avantage de vous orienter vers des ressources "invisibles" aux moteurs généralistes.

*Pour les repérer, il existe des...*

## "répertoires de répertoires "

### En anglais

Le site de référence [Argus Clearinghouse](#) n'est plus mis à jour, ses données ont été prises dans l'Internet public library [IPL.org](#)  
[Beaucoup.com](#) ,  
[Bubl Link 5:15](#)  offre plusieurs classifications : alphabétique, par rubriques, Dewey, etc.

### En français

[www.Enfin.com/dir.php](http://www.Enfin.com/dir.php) 

*sans oublier que ... :*

## les annuaires généralistes

donnent très souvent, dans chacune de leurs rubriques, une liste d'annuaires spécialisés.

## les méta-portails

permettent d'interroger simultanément le contenu de plusieurs sites du même domaine.

ex : pour les professionnels de l'information : [drapo.com](#)

les [portails institutionnels](#) offrent généralement cette possibilité de requête simultanée.

## l'intuition

# LES OUTILS SPÉCIALISÉS PAR DOMAINE. 2, QUELQUES PISTES...

à jour le  
7 sept. 2003

... pour repérer des répertoires de sites

## La plupart des sites à vocation d'information ou de formation

contiennent une "webographie" offrant des liens vers des sites apparentés.

Par exemple :

- [BNF](#) ---> liste de sites utiles aux bibliothécaires, accès aux catalogues d'autres bibliothèques ("[les signets de la BNF](#)")
- [ENA](#) --> liens vers d'autres grandes écoles, d'autres formations de fonctionnaires, des sites administratifs ("[Documentation et ressources](#)")
- [Fac de médecine \(CHU de Rouen\)](#) --> liens vers des sites médicaux ([CISMeF](#))
- [Mediadix](#) ---> liens vers d'autres instituts de formation de professionnels de l'information, vers des outils utiles à ces formations ("[Ressources](#)").
- [Courrier international](#) --> liens vers de nombreux journaux du monde entier ("[Kiosque en ligne](#)")

## Sites d'associations de professionnels

(généralement avec extension '.asso.fr')

Ayant vocation à informer leurs adhérents, leurs sites peuvent offrir des liens très intéressants.

- Associations de bibliothécaires : ABF [www.abf.asso.fr](http://www.abf.asso.fr) , ADBS [www.adbs.fr](http://www.adbs.fr) , ADBDP [www.adbdp.asso.fr](http://www.adbdp.asso.fr)
- Syndicats (cf le cours "[l'économie sur internet](#)")
- etc.

## Votre propre sélection

Pour vous, le répertoire spécialisé de sites le mieux adapté est votre fichier de "[signets](#)" ou "[favoris](#)" !

Une recherche par un moteur ou un annuaire généraliste peut conduire à un répertoire qui permettra d'élargir la recherche. Après avoir **vérifié** l'information qu'il rassemble, pensez à mettre ce répertoire en pour vous y reporter ultérieurement.

Selon votre domaine, vous trouverez ainsi quelques incontournables qui vous serviront de porte d'entrée sur le web.

**Par exemple :**

- Information gouvernementale : [les portails institutionnels](#) 📄
- Presse : [une sélection](#) 📄
- Informations économiques : [voir le cours "économie"](#)
- Droit : [voir le cours "internet juridique"](#)
- Culture : [www.portail.culture.fr](http://www.portail.culture.fr) 🌐: (ministère de la culture) et le guide des sites de [telerama.fr](http://telerama.fr) 🌐



Eric  
Pichon

Consulter aussi [les signets](#) en annexe de ce ce cours

# LE WEB INVISIBLE

à jour le  
7 sept. 2003

## Sur internet, un très grand nombre de pages ne sont pas ou sont mal repérées par les moteurs de recherche

- sites exploitant des bases de données (les pages d'interrogation sont repérées mais pas le contenu des bases de données)

### Exemples :

- les librairies en ligne, les catalogues de bibliothèques (par ex. [uncaps](#)), ...
- pour le cinéma : [IMDB](#),
- pour les CD : [CDDB](#) sur Gracenote.com

- sites protégés par mot de passe
- contenu des documents téléchargeables inclus dans les pages (rtf, pdf, etc. )
- et tous les sites mal référencés... (cf la théorie du "noeud papillon")

Mais les progrès des moteurs permettent de diminuer la part de ce "web invisible", par exemple, [Google](#) et [HotBot](#) indexent désormais le contenu des documents pdf et Microsoft Office (doc, xls, ppt) et Alltheweb les pages en "Flash"©

D'autres sites s'efforcent de recenser et de mieux exploiter les bases de données >>>

---

Eric  
Pichon

# LE WEB INVISIBLE, 2. LES BASES DE DONNÉES

à jour le  
7 sept. 2003

## Recensement de bases de données

### Sites en anglais

**The invisible web** <http://www.invisibleweb.com/>

**Direct search** <http://www.freepint.com/gary/direct.htm>

### Sites en français

**Formist** <http://bdd.formist.enssib.fr/index.html> 

**ADBS** <http://www.adbs.fr/adbs/adherent/bdd/html/index.htm>  
(malheureusement réservé aux adhérents)

**DADI Les bases de données gratuites sur internet**  
<http://urfist.univ-lyon1.fr/gratuits/> 

Les "[répertoires de répertoires](#)" recensent (mais pas exclusivement) des bases de données spécialisées.

## Recherche dans plusieurs bases de données

Certains [meta-portails](#) offrent la possibilité de recherche simultanée dans plusieurs bases de données.

Des outils téléchargeables, comme **Copernic** ([www.copernic.com](http://www.copernic.com)) ou **Strategic finder** ([www.digimind.fr](http://www.digimind.fr)), peuvent lancer une requête simultanément dans plusieurs bases de données (notamment dans celles des librairies en ligne).

C'est également le principe des comparateurs de prix.

### Ouvrages en ligne

**Internet public library (reading room)** <http://www.ipl.org/div/reading>

### Articles

voir "[la presse sur internet](#)  "

## Recherche sur des types d'information ou de documents particuliers

# LES OUTILS SPÉCIALISÉS PAR TYPE D'INFORMATION OU DE DOCUMENTS

à jour le  
7 sept. 2003

Ils permettent de trouver facilement, par exemple : des [logiciels](#), des [adresses](#), des [forums](#), des [listes de diffusion](#), ...

## Des logiciels

voir le cours sur [la récupération et le traitement de l'information](#).

## Des adresses

### Mél

La plupart des outils de recherche généralistes ont des formulaires permettant de retrouver des adresses e-mail, mais cela nécessite souvent que la personne recherchée se soit inscrite auprès d'eux.

Il faudra donc parfois recourir à plusieurs annuaires (NB : Copernic offre aussi une option de recherche simultanée sur plusieurs annuaires d'adresses électroniques).

D'autres sites permettent de retrouver les e-mails de personnes à partir des contributions aux forums de discussion (<http://usenet-addresses.mit.edu>)

Il est parfois plus aisé de retrouver l'e-mail de quelqu'un par le biais du site de l'entreprise où il travaille ou bien en lui téléphonant...

### Téléphone

En France, les [pages jaunes](#) et [blanches](#) sont disponibles sur internet ; ainsi que les annuaires inversés (cf <http://www.quidonc.fr>)

Pour les téléphones mobiles, il n'existe pas encore d'annuaire inter-réseau.

Des sites recensent les annuaires des autres pays, par ex. "Annuaire du monde" sur [pagesjaunes.fr](http://pagesjaunes.fr) ; [www.phonenumbers.net](http://www.phonenumbers.net)

## Forums et listes de diffusion

Des sites spécialisés permettent de rechercher [des forums et des listes de diffusion](#) sur un thème donné, ainsi que d'interroger directement dans les contributions à ces débats >>>>

# FORUMS, LISTES DE DIFFUSION, LETTRES D'INFORMATION

à jour le  
7 sept. 2003

## Les forums

Ce sont des "lieux publics virtuels" de discussion entre internautes. Il en existe plusieurs dizaine de milliers, sur les sujets les plus divers. Certains ont un "modérateur" qui filtre les contributions (pour éviter les injures mais aussi les redites et les hors-sujet).

Pour consulter les messages d'un forum et participer à la discussion, il faut se connecter à un serveur de forum - généralement celui de votre fournisseur d'accès, qui sélectionne les forums qu'il diffuse.

On peut trouver des contributions sur un sujet donné par le biais d'un moteur de recherche général, en précisant que l'on recherche dans "usenet" ou dans "groups"

## Les listes de diffusion

Ce sont également des lieux d'échange, mais vous recevez les contributions directement dans votre boîte aux lettres. Pour cela, il est nécessaire de vous abonner auprès du gestionnaire de la liste en envoyant un message du type :

*subscribe nom-de-la-liste votre@dresse*

(*unsubscribe nom-de-la-liste votre@dresse* pour vous désabonner)

Une fois abonné, vous pourrez lire les messages et envoyer vos propres contributions. Les listes de diffusion sont généralement "modérées", c'est à dire que le responsable de la liste ne laisse pas passer les messages "hors-sujet".

Des répertoires recensent les listes de diffusion. Pour les listes francophones, vous pouvez consulter le site **Francopholistes** <http://www.francopholistes.com>, qui offre un descriptif de chaque liste, un accès à ses archives et vous indique les procédures d'abonnement. Des recherches par thèmes, publics, ou contenu des messages publiés sont disponibles.

## Les lettres d'information

Il s'agit souvent de "produits dérivés" d'un site. Elles permettent d'être informé régulièrement sur l'actualité du domaine couvert par ce site.

Par ex.

[www.abondance.com](http://www.abondance.com) propose une lettre d'information sur

l'actualité des moteurs de recherche :

Abonnement à la lettre d'information gratuite "[Actu Moteurs](#)"

[www.adbdp.asso.fr](http://www.adbdp.asso.fr) une lettre concernant les bibliothèques départementales de prêt.

<http://www.crdp-toulouse.fr:8000/netclair/netclair.htm>  
un bulletin destiné au documentalistes de CDI.

[où chercher ?](#)



---

Eric  
Pichon

# L'EVOLUTION DES OUTILS DE RECHERCHE

à jour le  
7 sept. 2003

Les moteurs de recherche évoluent de façon quantitative mais aussi qualitative :

voir par exemple, les projets de Google :  
<http://labs.google.com/>

## Amélioration de la pertinence

### du fait de la structuration des pages

La structuration des pages HTML ne permet une indexation très poussée des documents. De nouveaux formats (XML) ou recommandations ([Dublin core](#)) tentent de pallier cet inconvénient (pas encore vraiment pris en compte dans les moteurs généralistes).

### du fait des nouveaux critères utilisés par les moteurs

- mesure de la notoriété : [Google](#)
- mesure de la popularité : Directhit (technologie intégrée dans [Teoma](#))
- indexation de nouveaux formats de documents : Flash ([Alltheweb](#)), documents bureautiques ou pdf ([Google](#))
- recherche dans le contenu des bases de données (Copernic, Strategic finder, comparateurs de prix)

## Analyse de la question posée

- recherche de mots proches : [Google](#) (ex. si l'on tape "contrôle" propose "contrôle")
- recherche sur des synonymes (Google version anglaise- faire précéder le terme de ~)
- reformulation des questions posées en langage naturel (en anglais : [Ask Jeeves](#))
- propositions de termes pour préciser votre recherche ([Alltheweb](#), [Teoma](#), [Netscape](#), Altavista)

## Evolution de l'interface offerte à l'utilisateur

### simplification :

- [Google](#), imité par [Raging search](#) d'AltaVista, par [Voila pour les geeks](#) et bien d'autres...
- intégration dans les navigateurs :
  - sous forme de barre d'outils (quasiment tous les outils généralistes en proposent)

- ou appel automatique d'un moteur de recherche lorsque ce que l'on tape dans la barre d'adresse n'est pas reconnu comme une URL

## organisation des résultats :

- présentation graphique ( [Kartoo](#) 🌐 ),
- "clustering" regroupement en dossiers (Alltheweb, Vivisimo),
- aperçu des réponses ( [EO.st](#) ; [GoogleViewer](#) )

## Personnalisation

- possibilité de paramétrer l'interface sur le fond ; ou même sur la forme (par ex.à l'aide de feuilles de style\* : Alltheweb)
- [fonctions de veille](#) <<
- géolocalisation : l'internaute trouvera en priorité les sites lui proposant des services proches de lieu d'où il interroge ("Par exemple il cherche un CD d'un artiste, [DeepIndex](#) lui propose [ra] les sites en relation avec sa requete, et dans sa proximité géographique. Dans cet exemple, nous lui proposerons par exemple le site d'un disquaire ayant le CD de l'artiste se trouvant physiquement/géographiquement dans une commune voisine.")

---

[panorama des outils actuels](#)

---

Eric  
Pichon

Voir aussi la [sélection de sites](#) de ce cours

Vu !

# LES "PORTAILS" INSTITUTIONNELS

sept. 2003



La multiplication de sites officiels a conduit à la création de sites "portails" qui ont vocation à rassembler l'ensemble des sites dépendant d'un même gouvernement, de la même organisation internationale...

- Ces sites fournissent de plus divers types de renseignements pratiques,
- Ils permettent généralement de chercher sur les sites qu'ils fédèrent à l'aide d'un moteur interne ou d'index
- Ils peuvent donner accès à des bases de données (gratuites ou non)

Exemples :

<b>Service public</b> <a href="http://www.service-public.fr">www.service-public.fr</a>	<ul style="list-style-type: none"> <li>● liens vers les sites des institutions et des services publics de la République française, et les principaux sites étrangers</li> <li>● annuaire des services publics</li> <li>● formulaires administratifs téléchargeables</li> <li>● textes officiels</li> </ul>
<b>Europa</b> <a href="http://europa.eu.int">europa.eu.int</a>	sites des institutions de l'Union européenne + bases de données de l'UE
<b>Système des Nations Unies</b> <a href="http://www.unsystem.org">www.unsystem.org</a>	sites des différents organismes dépendant de l'ONU

voir aussi "[Internet juridique](#)"

Eric  
Pichon



Vu !

Vu !

# LA PRESSE SUR INTERNET

8 sept. 2003

[Les sites de journaux](#) - [les métamoteurs de recherche d'articles](#) - [autres bases de données d'articles](#) - [sélection d'articles et dossiers d'actualité](#) ↗



## Les sites de journaux

**Webdo** [www.webdopresse.ch](http://www.webdopresse.ch) recense des sites de journaux en ligne du monde entier.

### Courrier international

[www.courrierinternational.com/kiosk/kiosq.htm](http://www.courrierinternational.com/kiosk/kiosq.htm)

et **Worldpress** <http://worldpress.org/gateway.htm>

signalent aussi les journaux n'ayant pas de site internet.



## Les métamoteurs de recherche d'articles

**World News** <http://www.wn.com> ;

**Rocket News** [www.rocketnews.com](http://www.rocketnews.com) ;

**Worldpress Review** <http://worldpress.org/search>

interrogent simultanément les journaux en ligne du monde entier.

**Google news** <http://news.google.com> compose automatiquement sa "une" en fonction des sujets les plus traités par les sources, essentiellement en anglais, qu'il recense.

**DeepIndex actualités** [www.news.deepindex.com/](http://www.news.deepindex.com/) recherche sur la presse en ligne francophone.



## Les bases de données d'articles

**FindArticles** [www.findarticles.com](http://www.findarticles.com) possède sa propre base de données d'articles - ceux-ci n'étant pas tous accessibles en ligne par ailleurs.

**Worldpress** <http://worldpress.org/search> offre également accès à sa propre base de données de fiches pays et de documents "exclusifs"



## Les sélections d'articles et les dossiers d'actualité

**Yahoo! actualité** <http://fr.news.yahoo.com> établit des dossiers

d'actualité <http://fr.fc.yahoo.com> composés de dépêches, d'une sélection d'articles en lignes, ainsi que de pages et de sites en rapport avec le sujet concerné.

cf aussi les animations <http://fr.news.yahoo.com/flash>

**Courrier international** [www.courrierinternational.com](http://www.courrierinternational.com) : traductions en français d'articles étrangers.

*voir aussi "[La presse économique](#)"*

Eric  
Pichon

Vu !