



## **RECHERCHER L'INFORMATION SUR INTERNET**

**Support rédigé pour les stages  
organisés par l'ADBS en 2001**

**Document réalisé par Françoise Quaire pour le compte de l'ADBS**

### **PREAMBULE**

Si, à première vue, il semble très facile de naviguer sur l'internet (pour certains, il suffit de quelques clics de souris !), on s'aperçoit vite que cette facilité n'est que bien relative dès qu'il s'agit de rechercher de l'information à bon escient.

L'impression est fréquente de perdre son temps, voire de se perdre dans ce labyrinthe en perpétuelle évolution. Les repères spatiaux sont différents des documents papier, la construction hypertextuelle incite à des cheminements difficiles à mémoriser, et la quantité d'informations

aussi hétérogènes rebute parfois les souhaits de pertinence et de fiabilité du documentaliste.

Il existe également un grand nombre d'outils différents, spécialisés dans la recherche d'informations sur l'internet ; il n'est pas question ici de tous les passer en revue. Mais plutôt de comprendre leur fonctionnement et leurs limites afin de les utiliser au mieux.

C'est donc une méthodologie de l'usage de ces outils qui est présentée ici, ainsi qu'un balisage général de la recherche des différents types d'information sur le net.

En fin de document, un glossaire de termes spécifiques à la recherche d'information sur l'internet et une bibliographie sélective complètent ce support de cours.

Dans la mesure du possible, les adresses URL citées ont été vérifiées lors de l'écriture du document, mais elles peuvent malheureusement se trouver modifiées par la suite et devenir alors obsolètes.

# INTRODUCTION

## ou

### des spécificités de la recherche sur Internet

#### ❖ L'information est multiple et hétérogène

- Sur Internet, on trouve divers espaces d'information : Web, Newsgroups, listes de diffusion, bases de données gratuites ou payantes, etc. correspondant à différents types d'informations (pages multimedia - conversations entre personnes - ensemble d'informations structurées et organisées pour l'interrogation par exemple).
- L'information est multimédia et peut se présenter sous forme de texte, d'image fixe ou animée (séquence vidéo ou animation de schéma), de son, de logiciel.
- La convivialité et l'ouverture d'Internet impliquent que tout internaute peut être producteur d'information et diffuser ses propres données ; on y trouve donc aussi bien des documents d'initiative privée (sites personnels) que des documents à caractère officiel émanant de centres de recherche et d'universités, de bibliothèques et centres de documentation, d'entreprises, d'organismes publics, parapublics, associatifs, d'organes de presse, etc.

#### ❖ L'information est abondante et en croissance exponentielle

Bien que difficilement quantifiable, la taille du web, qui est estimée à environ 28 millions de serveurs web (<http://www.netcraft.com/survey/>), dépasse les 2 milliards de pages uniques accessibles. On sait aussi que la toile (surnom que les Québécois donne au web) s'accroît de façon exponentielle d'environ 50% par an.

Si on considère également la structure réticulaire de l'information, où chaque page web peut être liée à plein d'autres pages stockées aux quatre coins de la planète, on imagine aisément l'ampleur et la complexité de l'outil qu'on interroge.

#### ❖ L'information est mobile et évolutive

- Les informations présentes sur Internet ne sont pas archivées dans un but patrimonial. Des sites naissent, disparaissent ou changent de localisation donc d'adresse (l'instabilité des URL explique les fréquentes apparitions du message "404 not found").
- Certaines pages modifient leur contenu constamment, en fonction de l'actualité par exemple. Sur le web, la réactivité prime sur la permanence.

Toutes ces spécificités expliquent la différence profonde entre l'information diffusée sur Internet et l'information que les documentalistes pratiquent habituellement, qu'elle soit sous forme papier ou électronique (banques de données, système de GED par exemple).

En documentation, on estime qu'une recherche est performante lorsqu'elle nous procure **tous** les documents recherchés et **uniquement** les documents recherchés. Selon ces critères, la recherche sur Internet comporte un tel taux de **silence** et de **bruit** qu'on ne peut la qualifier de performante. Il faut constamment se rappeler que sur le net, on ne peut pas avoir d'idée claire :

- de la **couverture** du gisement d'informations où la recherche est lancée – il faut savoir abandonner le rêve d'exhaustivité car il n'existe pas ici ;
- ni de son **indexation** puisque jusqu'à présent, si les documents en format HTML sont organisés en vue de leur présentation à l'écran, ils ne sont pas structurés logiquement comme dans une base de données. Et si on parle d'indexation sur le web (balise « keyword »), elle reste en langage naturel.
- ni du **mode de fonctionnement précis** de la recherche puisque chaque outil garde jalousement ses algorithmes, qui de plus évoluent avec le temps.

L'internet n'est donc pas cette immense « bibliothèque virtuelle » présentée partout mais plutôt un gisement évolutif d'informations hétérogènes où le pire côtoie le meilleur, où l'on peut trouver aussi bien des informations prévisibles qu'inattendues. Il faut également se rappeler qu'historiquement Internet a été conçu pour diffuser et échanger de l'information et non pour la chercher. Les outils de recherche sont apparus ultérieurement, de façon empirique et sans cohésion organisée.

Finalement, sur Internet, on peut trouver de tout, ce qui ne veut surtout pas dire que tout y est !

Pour trouver l'information souhaitée, on recense actuellement **trois schémas de recherche** :

- l'utilisation directe d'adresses de site repérées soit hors du web (presse, réseau professionnel, etc.) soit déjà enregistrées en signet ou favori,
- la recherche navigationnelle qui consiste à explorer les liens existants depuis un site pertinent sur le sujet,
- l'utilisation d'outils spécialisés comme les moteurs, métamoteurs et répertoires, créés spécifiquement pour faciliter la recherche sur Internet.

## Les différents outils généralistes

Deux grandes familles de services se partagent la tâche de chercher de l'information sur Internet : **les répertoires et les moteurs de recherche**. D'apparence parfois similaire, ils se distinguent cependant par leur façon de recenser et d'organiser d'imposantes sommes d'informations. On tend désormais vers une intégration de ces différents outils sous un portail "mixte" d'où la confusion encore plus grande qui règne !

### Les répertoires ou annuaires (directories en anglais) = outils humains

Ils sélectionnent, de façon plus ou moins stricte, les sites et les regroupent dans des catégories thématiques (finances, sports, tourisme, etc.). Souvent, un court résumé décrit le contenu des adresses référencées.

Ce sont donc des **inventaires thématiques** de ressources organisées selon une classification. Les sites sont sélectionnés en fonction de leur qualité et de leur pertinence. Ces répertoires peuvent être "par soumission" : les créateurs de sites proposent leur web à l'équipe d'indexeurs (qui le refuse ou l'intègre) ou bien être constitués par le parcours systématique (mais jamais exhaustif) des indexeurs sur le web. Dans tous les cas, un répertoire ne recense que des adresses de site (et non les pages contenues dans ce site). L'indexation se fait sur les termes présents dans le titre du site et dans la fiche descriptive communiquée par le webmaster, mais jamais sur le texte intégral du site. On pose ainsi une requête sur un index limité (d'où la confusion éventuelle avec les moteurs de recherche), mais on peut aussi naviguer dans l'arborescence de la classification qui s'affine progressivement.

**Avantages pour l'utilisateur** : l'intérêt majeur de cette approche est la valeur ajoutée humaine (contrôle, validation, organisation) qui limite ainsi le bruit des réponses.

Le répertoire propose deux modes de recherche :

- soit en parcourant les différentes rubriques (recherche de type arborescente guidée)
- soit en entrant des mots-clés qui porteront sur les titres des sites et leur présentation. C'est en quelque sorte une requête sur résumé de contenu, avec des fonctionnalités de recherche généralement limitées.

Cette classification des ressources permet de répondre à des requêtes de type exploratoire sur une thématique ou bien d'avoir une vue d'ensemble d'un thème ; elle permet également de trouver des sites ressources sur un sujet, ou encore d'autres sites comparables à celui qui sert de repère.

**Points faibles** : ce type d'outil nécessite d'entrer dans la logique de classification (ce qui n'est pas compatible avec tous les types de recherche). La sélection humaine est forcément subjective. La couverture y est moins « exhaustive » que celle des outils automatiques de type robot avec une mise à jour généralement manuelle ; et il n'y a guère de possibilité de recherche pointue et sophistiquée.

☛ **Un truc** : si vous avez trouvé une adresse intéressante avec une recherche par mots-clés, cliquez sur la **rubrique associée** pour voir s'il existe des sites du même type.

➤ **Les principaux annuaires de recherche francophones**

• **Yahoo ! France** [<http://www.yahoo.fr>]

Clair et très riche, malgré son nom américain il recense uniquement des informations en français sur des sites francophones.

• **Nomade** [<http://www.nomade.fr>]

Un annuaire par soumission. C'est l'un des premiers et des principaux guides francophones. Il est complémentaire de Yahoo sur certains sujets et offre une présentation plus détaillée des sites.

• **Wanadoo** [<http://www.recherche.wanadoo.fr>]

Pour accéder directement à la partie « Guide » présente aussi sur le portail de recherche *Voilà*.

• **Looksmart France** [<http://www.looksmart.fr>]

La version française de cet annuaire américain date de 2001.

➤ **Les principaux annuaires de recherche anglophones**

• **Yahoo !** [<http://www.yahoo.com>] (*Yet another hierarchical organized oracle*)

C'est le plus ancien, le plus connu et le plus utilisé des annuaires. De nombreux sites organisés dans une liste hiérarchique très fine. Intéressant aussi pour les dépêches ; bon site grand public.

• **Looksmart** [<http://www.looksmart.com>]

La version « mère » propose bien plus de sites et de catégories que la toute récente version française.

• **Netguide** [<http://www.netguide.com>]

Répertoire très US avec un système de guides thématiques spécialisés sur quelques sujets (synthèse).

Il existe également des **répertoires ou annuaires sélectifs**, qui pour chaque rubrique ne présentent que les meilleurs sites (selon des critères spécifiques aux indexeurs), les « essentiels » en quelque sorte. Ces outils sont utiles pour repérer des sites de référence sur un sujet donné.

• **BonWeb** [<http://www.bonweb.com>]

Guide en français qui recense les 10 meilleurs sites par catégorie ; avec des commentaires évaluatifs, il concerne surtout des thèmes pour le grand public.

• **About.com** [<http://www.about.com>]

Guide anglophone qui regroupe 500 sujets différents suivis par des experts de leur domaine. L'objectif visé est la qualité des liens et des articles proposés.

Il existe enfin des **répertoires ou annuaires contributifs**, qui confient la responsabilité d'une ou

plusieurs catégories à des internautes bénévoles dont la compétence a été vérifiée.

• **Open Directory Project** [<http://dmoz.org>]

Cet outil s'appuie sur un grand nombre de volontaires bénévoles pour construire un gigantesque répertoire du web devenu désormais incontournable. Slogan : "humans do it better".

## Les moteurs ou robots (search engine ou "bots" en anglais)

Le moteur indexe automatiquement (par le biais d'un robot qui sillonne inlassablement le réseau) un très grand nombre de pages web mais ne les classe pas. C'est un système de **collecte automatique suivie de l'indexation également automatique** de ressources. L'index du moteur de recherche contient donc tous les termes des pages web visitées (il ne reste pas au niveau du site).

Tout robot est essentiellement composé de 3 parties :

1/ le module de **collecte automatique** des ressources (un logiciel : le spider visite de façon permanente des milliers de sites web et en stocke le contenu sous forme de copies de page). Il possède des algorithmes spécifiques pour examiner périodiquement des millions de pages sans tourner en boucle.

2/ le module d'**indexation**, au sens constitution de fichiers inverses (l'indexeur ou crawler analyse de façon plus ou moins fine l'information des millions de documents rapatriés et crée un fichier d'indexage. C'est ce fichier qui sera consulté lors de la recherche par mot-clé.)

**L'indexeur est propre à chaque robot** et les contenus indexés sont très variables (contenu intégral des pages, titres html, premiers paragraphes ou liens). Cette indexation est associée à une **pondération** des mots qui tient compte à la fois de critères statistiques (parfois lexicaux) mais aussi de l'encodage html (balise méta ou meta-tags).

3/ le module de **recherche d'information** (le searcher) gérant aussi l'interface avec l'utilisateur. L'utilisateur pose une requête sous forme d'une liste de mots séparés par des opérateurs implicites ou choisis. Il y a très souvent 2 modes de recherche (simple et avancée).

### Pour autant, il ne faut pas croire que les robots peuvent "tout" récupérer :

- d'abord parce qu'ils ne peuvent pas dénicher des informations à l'intérieur de bases de données organisées de type professionnelles (on parle alors du web invisible).
- ensuite parce que le web grossit actuellement plus vite que les capacités d'indexation des moteurs.

### Avantages pour l'utilisateur :

- La couverture plus exhaustive (mais toujours relative) de l'espace public Internet avec des documents très diversifiés,
- la possibilité de récupérer des informations présentes dans les pages d'un site web donc de pouvoir faire des requêtes très précises et pointues,
- et des options de recherche assez sophistiquées pour affiner une requête.

### Points faibles :

- la grande hétérogénéité des réponses et le risque fréquent de doublons,

- le nombre souvent ingérable de réponses d'où un bruit énorme, mais aussi du silence,
- savoir qu'en mode avancé, la syntaxe d'interrogation demande parfois un bon apprentissage pour être efficace,
- et enfin, une mise à jour de plus en plus lente du fait de la croissance rapide du web.

👉 **Un truc** : *on le sait, les moteurs n'indexent au mieux qu'un tiers du web ; il faut donc avoir conscience du silence possible et **consulter différents robots** car souvent ils se complètent.*

### ➤ *Les principaux moteurs de recherche francophones*

- **Ecila** [<http://www.ecila.fr>]

Il n'indexe que les pages web françaises en cherchant dans le titre, la description et les mots-clés (balises méta)

- **Voila** [<http://www.voila.fr>]

Le principal portail français comme outil de recherche développé par France Télécom, il rassemble un guide et un moteur, et plein d'outils de recherche spécialisés (annuaires tél, Paris en photo, etc.)

- **AltaVista France** [<http://www.altavista.fr>]

Version française récente du célèbre moteur américain, qui ne concerne ici que des sites français.

### ➤ *Les principaux moteurs de recherche anglophones*

- **Alta Vista** [<http://www.altavista.com>] ou [<http://www.av.com>]

Le plus connu des moteurs : il indexe le texte intégral des sites, propose la recherche dans de nombreux fichiers multimedia et offre des possibilités de sélection assez sophistiquées.

- **HotBot** [<http://www.hotbot.com>]

Ce moteur présente une bonne palette d'options en recherche plus guidée que le précédent avec des menus déroulants.

- **Northern Light** [<http://www.northernlight.com>]

Un moteur innovant qui classe les résultats dans des dossiers contextuels distincts et indexe aussi des sources rarement présentes sur Internet.

- **Google** [<http://www.google.com>]

Moteur récent qui privilégie les sites les plus cités dans les liens (indice de popularité). De plus, il indexe les fichiers en format PDF, et archive l'ensemble des pages qu'il indexe.

- **Fast** [<http://www.alltheweb.com>]

Moteur apparu en 99 qui affiche une grande rapidité et revendique la première place pour le nombre de pages indexées.

## **Les méta-moteurs (ou méta-robots)**

Ils permettent **l'interrogation simultanée à partir d'une seule requête** de plusieurs outils de recherche, moteurs ou annuaires. Ces possibilités d'interrogation multiples sont intéressantes dans le cas de recherche ciblées et spécialisées qui ne génèrent pas un trop grand nombre de réponses, mais ne sont guère opérantes sur des termes trop génériques. Si on gagne ainsi en couverture des index balayés, on perd parfois en rapidité et les requêtes doivent rester simples pour s'adapter à tous les outils. En effet, les méta-moteurs ne maintiennent pas eux-même de base de données, ils transmettent juste la requête à d'autres outils mais grâce à l'efficacité de leur travail (interrogation en parallèle, élimination des doublons, présentation unique...), on les assimile parfois à une toute

première génération d'agents intelligents.

### **Quelques méta-moteurs**

- **MetaCrawler** [<http://www.metacrawler.com>]

Interrogation simultanée de moteurs ou de répertoires.

- **Savvy Search** [<http://www.search.com>]

Interrogation simultanée de 24 outils généraux de recherche, et des regroupements thématiques d'outils bien appréciables.

- **Profusion** [<http://www.profusion.com>]

Interrogation simultanée de 9 outils, et de sources sur le web invisible.

- **Ixquick** [<http://www.ixquick.com>]

Puissant et rapide, il travaille sur les 10 meilleures réponses proposées par des outils francophones.

### **Les méta-moteurs clients ou off-line**

Semblables dans leur fonctionnement aux outils ci-dessus, ce sont par contre des logiciels qui **s'installent sur le disque dur de l'ordinateur**. Ils permettent d'une part d'effectuer une recherche sur internet et d'autre part de rapatrier les documents trouvés sur le poste de travail pour une consultation hors ligne. Les plus connus sont **Copernic** (pour PC) et **Sherlock** (pour Mac). Certains offrent des particularités comme **Umap** qui constitue un thésaurus à partir de l'indexation des pages trouvées (le corpus) et génère sur une carte la représentation graphique de l'ensemble des mots du thésaurus. Ces outils sont surtout utiles dans le cadre d'une veille documentaire.

## **La tendance actuelle des outils de recherche d'information**

Devant la formidable expansion du web, les outils traditionnels ont beaucoup de mal à suivre cette croissance pour continuer à indexer une part raisonnable de l'ensemble des sites et de leur contenu. De plus, leur démarche marketing leur déconseille d'afficher des résultats nuls pour certaines requêtes. Ils ont donc développé des stratégies d'alliance afin d'offrir un meilleur service, mais parfois de façon bien peu transparente pour l'utilisateur.

### **❖ L'alliance quasi-systématique répertoire / moteur**

Désormais, les principaux outils généralistes ont adopté les uns après les autres cette stratégie. Ainsi quand les répertoires ne trouvent dans leur index aucune réponse à la requête posée, ils affichent les réponses venant d'un moteur (*par exemple, Yahoo.fr relaye sur Google et Nomade sur Inktomi*). A l'inverse, un moteur intègre très souvent désormais les catégories d'un répertoire, de façon visible ou plus discrète (*comme AltaVista France qui utilise Looksmart et Google qui sous-traite avec l'Open Directory Project*).

### **❖ Le portail combinant plusieurs outils de recherche**

L'exemple le plus évident en France, est Voilà qui, en réponse à une requête affichera :

- des données extraites de l'Encyclopédie Hachette
- des sites web venant du Guide (répertoire)
- des pages web tirées du Moteur

☛ **A noter :** *Pour savoir quel type d'outil vous répond, repérez vite si les résultats affichés sont des pages web (URL plus ou moins détaillée) ou des adresses de site avec leur rubrique indiquée.*

## Des lieux pleins de ressources :

### les outils et portails spécialisés, les sites fédérateurs

Devant la profusion des ressources évolutives sur le net, on assiste également à une **tendance à la spécialisation** avec des outils thématiques ou des démarches qui ne tentent plus d'être encyclopédiques mais de circonscrire intelligemment et utilement des secteurs d'information.

➤ Du côté des outils, on peut citer :

#### Les répertoires thématiques

Comme un annuaire, ils référencent des sites selon un thème, une région ou un axe particulier. En fonction de vos centres d'intérêt, ce sont des repères à garder précieusement dans vos signets car ils sont souvent le point de départ conseillé pour une recherche de type navigationnelle intelligente.

Exemples : **Indexa** [<http://www.indexa.fr>] pour les entreprises françaises avec un site web  
**Les 2 savoies** [<http://www.les2savoies.com>] annuaire région Savoie – Mont-Blanc  
**Strategic Road** [<http://www.strategic-road.com>] info. économique et géopolitique

#### Les moteurs spécialisés

Ils ne recherchent que sur un corpus bien défini. (Voir aussi « pour chercher ailleurs que sur le web » infra).

Exemples : **Ditto** [<http://www.ditto.com>] pour les photos et images  
**Orientation** [<http://www.orientation.com>] recherche par zone géographique

➤ Du côté des démarches, on peut citer :

#### Les portails spécialisés

En accès libre, ou après enregistrement pour certains, ces portails sont toujours thématiques et visent à faciliter l'accès à nombre de services et ressources depuis une page d'accueil synthétique.

Exemples : **Service Public** [<http://www.service-public.fr>] pour l'administration française  
**Legifrance** [<http://www.legifrance.gouv.fr>] pour l'info. juridique (codes, lois, JO)  
**PlasticWay** [<http://www.plasticway.com>] sur la plasturgie française

#### Les sites fédérateurs ou de référence

Ce sont des recensements de sites de référence ou de ressources Internet sur un thème ou visant un objectif précis. Parfois réputés et volumineux, parfois très pointus, ils sont souvent l'oeuvre

d'organismes, de documentalistes...ou tout simplement de passionnés d'un sujet. En fait, en les consultant, il faut partir du principe que quelqu'un a peut-être déjà collecté thématiquement des ressources sur le sujet recherché ! Attention toutefois à rester attentif aux dates de mise à jour.

### **Quelques exemples**

- L'ADBS propose un "**guide thématique des Webs**" qui recense plusieurs méta-pages francophones, par thèmes. [<http://www.adbs.fr/adbs/sitespro/gthwebs/html/index.htm>]
- **Les Signets de la BnF** [<http://www.bnf.fr/web-bnf/liens/accueil.htm>]  
Une sélection de plus de 1 600 sites utiles ou de qualité. Entrée thématique + outils Internet.
- **Sapristi !** [<http://csidoc.insa-lyon.fr/sapristi/digest.html>] pour l'Information Scientifique et Technique
- **AlphaSearch** [<http://www.calvin.edu/library/searreso/internet/as/index.htm>] recense en anglais des sites fédérateurs de qualités dans 35 disciplines universitaires.

### **Pour chercher ailleurs que sur le web**

Pour explorer sur le net d'autres espaces d'information que le web, il y a des outils dédiés comme :

- **Francopholistes** [<http://www.francopholistes.com>] pour les listes de diffusion francophones
- **Liszt** [<http://www.liszt.com>] pour les listes de diffusion internationales
- **Voila News** [<http://www.news.voila.fr>] pour les newsgroups francophones
- **Ex-Deja** (Google) [<http://groups.google.com>] pour les newsgroups anglophones
- **Shareware** [<http://shareware.cnet.com>] pour les logiciels
- **Tucows** [<http://www.tucows.com>] pour les logiciels

### **Les listes d'outils de recherche**

Enfin, il existe des recensements, plus ou moins complets et à jour, des différents outils de recherche disponibles sur Internet.

Exemples : **Beaucoup** [<http://www.beaucoup.com>]  
**7alpha** [<http://www.7alpha.com>]  
**The Big Hub** [<http://www.thebighub.com>]

☛ **Un truc** : pour découvrir, explorer et apprendre à connaître tous ces outils, voir la page récapitulative tenue à jour par Jean-Pierre Lardy et intitulée : "**Vite, tous les outils ...**" [<http://www.adbs.fr/adbs/sitespro/lardy/outils.htm>]

Ces quelques pistes ne sont données ici qu'à titre indicatif car il ne s'agit nullement de connaître tous les outils existants, mais plutôt d'utiliser la catégorie d'outils la plus efficace par rapport au

type d'information recherchée. Une bonne pratique des outils de recherche principaux paraît préférable à une pratique moyenne de dizaine d'outils différents. Il faut toujours chercher le bon point de départ, et ne pas oublier qu'Internet peut parfois simplement contribuer à trouver la réponse à la question : *Où trouver ... ?*

## Méthodologie de la recherche d'informations

*"La recherche d'information est un art et non une science."*

Voici quelques clés indispensables à toute recherche d'information. On peut ensuite bien évidemment les affiner en fonction des types et des niveaux de complexité des requêtes.

### Définir la demande et adapter la stratégie

C'est une étape indispensable. Il faut définir ce que l'on recherche pour sélectionner de façon pertinente les meilleurs outils en fonction de cette typologie. Une question est toujours dans un contexte et demandée par quelqu'un. Ex. adresse mail, dépêches d'agence, dossier de fond, noms de sociétés, cartes....

- **Définir le sujet précisément** (*quoi ?*) sujet traité, thématique, quels mots-clé ?
- **Définir l'étendue et le contexte** (*pourquoi ?*) bien délimiter son sujet, voir son utilité
- **Définir la forme** (*comment ?*) Texte ou images. Dépêches, article, documents, liens ...

Il s'agit ici d'une analyse classique de la question, à laquelle il convient d'ajouter une analyse spécifique au contexte Internet afin de définir une stratégie logique d'action (quelle démarche adopter, quels outils utiliser...?)

### Connaître les opérateurs et les requêtes en mode avancé

Tout comme l'interrogation des banques de données classiques, les moteurs de recherche offrent souvent des possibilités d'interrogation selon deux modes : requête simple ou requête avancée. Savoir que lorsqu'on lance une requête, le moteur fait une équation de recherche de type Bdd sans le montrer expressément. Seul problème, les syntaxes de requêtes avancées sont spécifiques à chaque moteur.

#### ❖ Règles de base (valables sur la plupart des outils de recherche)

Un mot est une chaîne de caractères alphanumériques, délimitée par un caractère de ponctuation, un blanc ou autre. Peu d'outils intègrent la reconnaissance lexicale (graphie inexacte, faute de frappe...)

- ☛ toujours vérifier l'exactitude des termes rentrés (*Washinton au lieu de Washington*)

#### ❖ Majuscules / minuscules

Certains outils respectent la casse exacte des termes. Donc, pour récupérer le maximum d'informations, toujours penser à taper la requête en minuscules car le robot cherchera sur toutes les possibilités.

En majuscules, il se conformera souvent à l'exacte copie du mot tel qu'il a été tapé. (Sauf si l'outil accepte les 2 indifféremment comme "Voilà").

- ☛ *FRANCE = uniquement les sites avec FRANCE et non France  
France ou france = pourra sortir ces mêmes graphies + FRANCE*

Noter que certains outils (Hotbot) proposent une catégorie « nom de personne »

## ❖ Les opérateurs

- Les plus connus sont les **opérateurs booléens**.

Il s'agit de mots ou de signes (ET, OU, SAUF / AND, OR, NOT +, -) destinés à filtrer logiquement la recherche. Souvent en majuscules (et penser à l'anglais)

Le **ET** permet de croiser des notions (intersection d'ensembles)

*Exemple : UE **ET** Suisse*

Le **OU** permet d'élargir des notions (synonymie souvent)

*Exemple : Suisse **OU** Confédération helvétique*

Le **SAUF** permet de retrancher des notions (exclusion)

*Exemple : UE **SAUF** France*

- **+ et - (signe "plus" et "moins")**

Le signe **+** signale au robot que vous voulez que le mot figure obligatoirement dans les documents retrouvés. Toujours le coller au mot concerné.

*Exemple : +jazz +piano*

Le signe **-** revient à un SAUF : le terme est exclu de la recherche.

*Exemple : +café +boisson -bar*

- Pour la **troncature**

C'est souvent l'**astérisque (\*)** qui symbolise la troncature (Joker ou wildcard) mais ses effets sont différents d'un moteur à l'autre et à utiliser donc avec prudence. Selon les cas, la troncature est limitée à un seul caractère (*photo\** = *photo* ou *photos*) ou bien à plusieurs caractères (*photo\** = *aussi photographe, photographie ...*)

☛ un des usages les plus classiques : à l'intérieur du terme

*Exemple : colo\*r permet de récupérer tant color que colour*

- Les **expressions** ("phrase" en anglais)

L'usage des **guillemets** (quote) est de plus en plus répandu. Il est très pratique pour indiquer qu'on recherche une expression complète et non un ensemble de termes dissociés.

*Exemples : "virtual library" - "journaux électroniques" - "vins de la loire"*

- Les zones "**champs**"

Il peut être utile de limiter sa recherche au champs titre (title: pour AltaVista) ou bien à l'adresse URL.... Des recherches spécifiques peuvent être demandées sur certaines parties des documents. Le nom du champ doit être tapé en minuscule suivi de 2 points et du terme de la requête, le tout sans espace.

*Exemples : url:fr - link:adbs.fr - title:musique*

☛ **Un truc :** penser à consulter l'aide en ligne pour connaître la syntaxe exacte des opérateurs sur l'outil que vous utilisez.

## Savoir EVALUER les sources

L'utilisation croissante de l'Internet comme outil ouvert de diffusion de l'information rend indispensable l'évaluation de la qualité de ce qui est trouvé. Il faut avoir un regard critique, ne pas croire à tout ce qu'on y lit et surtout bien repérer la nature des émetteurs d'informations.

Voici quelques questions indispensables à se poser :

- **qui sont les responsables ?**

Y a-t'il une affiliation institutionnelle ? si oui, a-t-on de l'information sur cette institution ? L'auteur est-il bien identifié ? Peut-on le contacter ? (adresse e-mail)

Toujours penser à chercher la page d'accueil (homepage) quand on navigue avec les liens ou les moteurs de recherche qui renvoient directement sur des pages internes.

- **quel contenu ?**

L'information est-elle récente ? Quelle en est la mise à jour ? Les anciens documents sont-ils archivés ? Y a-t'il des liens vers des sites intéressants ?

- **quelle organisation de l'information ?**

Y a-t'il une table des matières ? L'organisation est-elle logique ? Peut-on se déplacer facilement dans le document ? Les mises à jour sont-elles notées et facilement identifiables ? Les pages sont-elles datées ?

## Mémoriser sa stratégie et enrichir son carnet

Le net est tellement labyrinthique et inégal, qu'il nécessite des outils personnels de repérage de l'information pertinente. C'est à chacun de construire son propre carnet de signets ou favoris, de recenser les sites ressources les plus pertinents selon ses centres d'intérêts.

Penser donc à  **systématiquement**  enregistrer un site intéressant en  **signet ou favori** . L'opération est rapide et permet souvent d'éviter une recherche ultérieure identique. En contre-partie, il faut régulièrement faire du ménage dans ses signets et les avoir classés au bon endroit.

☛ **Un truc :** les liens hypertexte utilisés sont d'une couleur différente (possibilité de configuration de la couleur et du temps de sauvegarde dans les "Options"). C'est parfois pratique quand on est sur un sommaire ou une liste chargée de nombreux liens à explorer.

## CONCLUSION

L'Internet n'est qu'une source d'information parmi d'autres qui peut être très utile pour certains sujets, presque inutile pour d'autres. Lors d'une recherche documentaire exhaustive, il faut la coupler avec d'autres outils. C'est aussi une ressource qui évolue vite et qu'il faut suivre

attentivement sous peine d'y perdre ses repères.

## Pour en savoir encore plus !

### ➤ BIBLIOGRAPHIE

- **Trouver l'info sur Internet** / Olivier Andrieu - Eyrolles, mars 1998 - 440 p.
- **Guide des sites internet pour l'entreprise** / Armelle Thomas - Dunod, 1998
- **La recherche intelligente sur Internet** / Henry Samier, Victor Sandoval – Hermes, 1998
- **Recherche d'Information sur Internet (Risi)** / Jean-Pierre Lardy - ADBS, Coll. Sciences de l'information, série Recherches et documents, 2000 - 102 pages *Des mises à jour de l'ouvrage sont en ligne sur le site de l'ADBS, Cf. infra*

### ➤ LIEUX RESSOURCES

#### ▪ francophones

- **GIRI** [<http://www.bibl.ulaval.ca/vitrine/giri>]  
Guide d'initiation à la recherche et guide des outils indispensables à cette recherche
- **Sapristi !** [<http://csidoc.insa-lyon.fr/sapristi/digest.html>]  
Guide méthodologique pour rechercher des pistes selon la typologie de l'information
- **RIIs** [<http://www.adbs.fr/adbs/viepro/sinfoint/lardy/risi.htm>]  
Méthodologie de recherche et outils (Cf. le livre supra)
- **Abondance** [<http://www.abondance.com>]  
Site consacré à l'actualité des outils de recherche et au référencement (lettre hebdomadaire)
- **LBQ** [<http://www.sciencepresse.qc.ca/lbq/lbq.html>]  
La Lettre du Bibliothécaire Québécois est une revue électronique bimestrielle sur l'actualité de la recherche d'information dans le domaine I – D et bibliothèque.

#### ▪ anglophones

- **Search Engine Watch** [<http://searchenginewatch.com>]  
Le site de référence sur l'actualité des outils de recherche, avec une lettre mensuelle gratuite.
- **Pandia** [<http://www.pandia.com>]  
Guide et actualité sur la recherche d'information

### ➤ LISTES de DISCUSSION

- **ADBS-INFO** [<http://listes.cru.fr/wws/info/adbs-info>]
- **BIBLIO-FR** [<http://www.cru.fr/listes/biblio-fr@cru.fr>]
- **Synthèses** de ces 2 listes sur le site de l'ENSSIB

[[http://www.enssib.fr/bibliotheque/cadre\\_syntheses.html](http://www.enssib.fr/bibliotheque/cadre_syntheses.html) ]

## GLOSSAIRE

### AGENT INTELLIGENT

Logiciel visant à faciliter la recherche et la gestion de l'information sur l'Internet.

### ALGORITHME DE PERTINENCE

*Relevancy algorithm* : Méthode utilisée par un moteur de recherche ou un répertoire pour relier les mots-clés d'une requête avec le contenu de chaque page, de telle sorte que les pages Web trouvées correspondent bien aux termes utilisés dans la requête. Chaque outil de recherche est susceptible d'utiliser un algorithme différent et de le changer ou de l'améliorer. Voir aussi : critères de tri.

### ANNUAIRE

Voir Répertoire

### APPLET JAVA

Programme en langage Java, téléchargeable et exécutable sur n'importe quel type de processeur visant à améliorer l'interactivité des pages. Les principaux navigateurs sont capables de lire et d'exécuter les applets Java. Il est possible que la présence de ce programme stoppe l'indexation d'une page par le robot d'un moteur de recherche.

### ARAIGNEE (*traduction littérale de Spider*)

C'est la partie du moteur de recherche qui "surfe" sur le net, enregistre les URLs, classe les mots-clés et le texte de chaque page qu'il trouve.

### ASPIRATEUR DE SITE

Outil permettant de copier un site Web à distance sur un disque dur ou CD Rom pour le relire ensuite en local.

### BRUIT

Réponse(s) non pertinente(s) fournie(s) lors d'une recherche d'information.

### CADRE

Voir Frame

### CGI

Acronyme de *Common Gateway Interface*, logiciel qui facilite la communication entre un serveur Web et des programmes fonctionnant hors de ce serveur ; par exemple, des programmes qui traitent des formulaires interactifs ou qui recherchent des informations dans des bases de données sur le serveur, suite à la requête d'un utilisateur.

### CONTENU DYNAMIQUE

Il s'agit de pages Web avec des informations qui changent ou sont changées automatiquement en fonction d'une base de données ou d'éléments provenant de l'utilisateur. (ex. suffixe .asp, .cfm, .cgi ou .shtml dans l'URL). Voir aussi Page dynamique

### CRITERE DE TRI

Façon automatique de sélectionner les résultats retournés par le moteur de recherche, afin de présenter en début de liste ceux qui correspondent le mieux à la requête. On distingue généralement le tri par pertinence du tri par popularité. Voir aussi : algorithme de pertinence.

## **CRYPTAGE**

Moyen de rendre secrète la communication informatique grâce à des logiciels d'encodage de données. Seul le possesseur de la clé de décodage peut interpréter le message. Aussi appelé "chiffrement".

## **DENSITE DES MOTS CLES**

Une des propriétés qui permet d'indiquer l'importance de certains mots dans le texte d'une page Web. Certains outils de recherche utilisent cette propriété pour le tri. La formule de calcul = nombre d'occurrences du terme demandé / nombre de termes de la page en question, une fois éliminés les mots vides.

## **DIRECTORY**

Voir Répertoire

## **EN-TETE**

*heading tags* : ce sont les commandes qui se trouvent en tête des pages html. Certains moteurs de recherche donnent plus d'importance et de poids au texte qui s'y trouve.

## **EQUATION DE RECHERCHE**

Formulation d'une question sous forme mots clés reliés par des termes logiques (ou opérateurs). Voir aussi opérateurs booléens.

## **FRAME**

*Synonymes : trame, cadre* . Il s'agit d'une technique de programmation en html qui permet de diviser la fenêtre du navigateur en plusieurs zones, puis de charger une page web dans chacun des cadres obtenus, les documents agissant l'un sur l'autre à travers différentes sous-fenêtres. Les frames posent souvent de gros problèmes d'indexation aux moteurs de recherche.

## **JAVA**

Langage multiplateforme, créé par Sun, capable de s'exécuter à l'intérieur d'une page Web. Souvent utilisé pour créer des applets.

## **JAVASCRIPT**

Langage simple interprété qui permet d'exécuter des petites tâches au sein des pages html.

## **KEYWORD**

Voir Mot-clé

## **LIEN A L'ARRIVEE**

Un lien hypertexte vers une page particulière venant de quelque part et apportant du trafic à cette page. Les liens à l'arrivée sont souvent un instrument de mesure pour connaître la popularité d'une page.

## **LIEN MORT**

Un lien qui ne mène plus à une page ou à un site, soit parce que le serveur est en panne, soit parce que la page a été déplacée ou bien n'existe plus. La plupart des outils de recherche ont

des techniques pour ôter de telles pages de leur liste automatiquement. Mais l'Internet continuant à croître quotidiennement, il leur devient de plus en plus difficile de contrôler régulièrement toutes ces pages.

### **META-MOTEUR ou METAMOTEUR**

Un outil qui, pour une même requête, interroge simultanément plusieurs moteurs de recherche et/ou répertoires et compile les résultats avant de les présenter. Parfois qualifiés d'agents "semi-intelligents".

### **META TAG**

Construction placée dans l'en-tête html des pages Web, fournissant des informations qui ne sont pas visibles par les navigateurs. Les plus courants des meta-tags sont KEYWORDS et DESCRIPTION.

### **MOT CLE**

*ou key-word* : Mot ou groupe de mot, éventuellement dans une forme lexicographique normalisée, choisi dans le titre, les meta tags ou le texte d'un document, caractérisé par le contenu et permettant la recherche de ce document.

### **MOTEUR DE RECHERCHE**

*ou Search Engine et chercheur au Québec* : Programme qui indexe le contenu de différentes ressources Internet, et plus particulièrement de sites Web, et qui permet à l'internaute de rechercher de l'information selon différents paramètres, en se servant de mots clés, et d'avoir accès à l'information ainsi trouvée.

Mode de fonctionnement : des robots logiciels (appelés crawlers ou spiders) scrutent le Web, vont de page en page et sauvegardent alors le contenu texte des pages rencontrées, constituant ainsi un "index", c'est-à-dire une collection plus ou moins grande de pages Web. Le robot logiciel repasse selon des délais plus ou moins fréquents sur les pages qu'il a indexées au préalable, pour en sauvegarder une version plus récente. On dit alors qu'il "rafraîchit sa base (ou son index)". Lorsque l'internaute saisit un mot clé dans le formulaire proposé, le moteur va en rechercher les occurrences dans son index, i.e. dans le contenu texte des pages Web sauvegardées au préalable. Une fois le "lot" de pages contenant le terme demandé identifié, le moteur classe les pages par ordre de pertinence, selon un ordre et un algorithme spécifique. (voir algorithme de pertinence et aussi critère de tri). A tort, le terme "moteur de recherche" est souvent utilisé tant pour un répertoire que pour un vrai moteur. Voir aussi : outil de recherche.

### **MOTEUR THEMATIQUE**

*Synonyme : moteur spécialisé* : Il procède par catégorisation automatique de pages, généralement à partir de catégories prédéfinies et de mots-clés préétablis.

### **OPERATEUR BOOLEEN**

Pour effectuer une recherche par mots clés, on couple souvent une suite de mots grâce à des opérateurs booléens. Venant du nom de George Boole (mathématicien britannique) ces opérateurs permettent d'élargir ou de restreindre la recherche en combinant certains mots et en excluant d'autres. Il existe plusieurs opérateurs booléens : le **ET** (les deux mots saisis figurent obligatoirement dans la réponse), le **OU** (la réponse comporte soit l'un des mots saisis, soit l'autre, soit les deux) et le **SAUF** (qui exclue le terme en question de la réponse).

## OUTIL DE RECHERCHE

Terme générique pour tout service de recherche d'information sur le Web, combinant bien souvent désormais les procédés d'un répertoire et ceux d'un moteur de recherche, plus parfois de moteurs spécialisés.

## P to P

*Point to Point, Peer to Peer ou encore People to People.* Type de connexion qui met en communication deux interlocuteurs, et seulement deux. Des applications comme Napster ou Gnutella emploient cette technique et permettent à des utilisateurs de permuter, d'échanger des fichiers directement entre eux, sans passer par un serveur d'hébergement.

## PAGE DYNAMIQUE

Page HTML dont le contenu n'est pas situé dans un fichier enregistré sur le serveur mais générées "à la volée" par une application informatique à partir d'un modèle de document HTML en accédant à des informations situées dans une (ou des) base(s) de données. Les techniques utilisées sont variables :

- CGI,
- langage de script,
- API propriétaires permettant de créer un lien entre la base de données et le serveur HTTP.

## PAGE STATIQUE

Page HTML dont le contenu est situé dans un fichier figé, enregistré sur le serveur Web.

## POPULARITE

*Synonyme : notoriété.* Mesure le nombre et la qualité des liens pointant vers une page particulière. Plusieurs moteurs de recherche utilisent de plus en plus ce procédé dans le processus de tri.

## PORTAIL

Terme générique pour désigner un site qui sert de point d'entrée sur l'Internet pour un nombre important d'utilisateurs. Un site portail offre une multitude de services différents depuis la page d'accueil.

## POSITIONNEMENT

*ou Ranking :* Processus de classement des sites, des pages Web dans un moteur de recherche ou un répertoire afin que les sites les plus pertinents apparaissent en premier sur la page résultat lors d'une requête.

## REGROUPEMENT

*ou Cluster ?* Affichage d'une seule adresse pour chaque site Web sur la page des résultats d'un outil de recherche. Cette méthode permet d'éviter qu'un petit nombre de sites occupe toutes les premières positions de résultats et en facilite la lecture pour l'utilisateur.

## REPERTOIRE

*Synonymes : catalogue, index thématique, liste thématique. Synonyme communément utilisé à tort : annuaire (un répertoire n'est pas annuel !).*

Liste de sites Web classés dans des catégories thématiques. Le classement est effectué et géré par des personnes physiques en fonction de diverses informations soit fournies au moment du

référencement par le concepteur du site, soit déduites après la visite du site par les indexeurs du répertoire. L'unité de classement est le site et non la page. La valeur ajoutée d'un répertoire tient en la qualité de son système de classification et à l'insertion éventuelle de commentaires et descriptions enrichies pour chaque site référencé. La plupart des répertoires proposent une recherche par mot-clé sur le titre des sites, les mots de la description, et les catégories concernées. On peut distinguer plusieurs types de répertoires : Cf. infra.

### **REPertoire GENERALISTE**

Répertoire ayant vocation à indexer tous les sites et qui n'effectuent une censure que sur la base de principes prédéfinis (par exemple, sont exclus des sites manifestement illégaux ou dont le référencement cherche à induire l'internaute en erreur). Exemple : Yahoo, Nomade

### **REPertoire SPECIALISE**

Répertoire dont les sites répertoriés relèvent tous d'un domaine ou d'un secteur particulier (le vin, le tourisme, le sport, l'agriculture, etc.). Un répertoire spécialisé peut, par exemple, ne prendre en compte que les entreprises d'un secteur, ou les produits d'un domaine. Ne pas confondre avec un moteur thématique. Exemple : Qualisteam.com dans le domaine bancaire et financier.

### **REPertoire SELECTIF**

Répertoire dont les gestionnaires privilégient les sites de meilleure qualité et excluent les sites qu'ils n'estiment pas suffisamment intéressants (exemple : bonweb.com).

### **REPertoire CONTRIBUTIF**

*Synonymes : répertoire ouvert,- open directory* : Répertoire dont l'enrichissement est effectué par différentes équipes d'internautes. Ces répertoires confient la responsabilité d'une ou plusieurs catégories soit à des internautes experts reconnus dans leur domaine et rémunérés pour leur prestation (exemple : About.com), soit à des internautes bénévoles dont la compétence dans le domaine couvert par cette catégorie a été vérifiée. Ces internautes reçoivent alors les demandes de référencement de leur catégorie, décident ou non de référencer les sites et, le cas échéant, rédigent eux-même la description du site (exemple : dmoz - Open Directory Project).

### **REPertoires D'OUTILS DE RECHERCHE**

*Synonymes : listes de listes, répertoires de répertoires - Répertoires spécialisés dans le référencement de répertoires et d'outils de recherche (moteurs de recherche, méta-moteurs, etc.).*

### **REQUETE**

*Synonymes : query, terme recherché* . Mot, expression ou groupe de mots employés pour interroger un outil de recherche afin de localiser des pages sur le sujet recherché.

### **ROBOT**

Programmes de navigation qui suivent les liens hypertextes des pages Web mais qui ne sont pas directement sous contrôle humain. Exemples : les spiders ou araignées des moteurs de recherche.

**ROBOTS.TXT**

Fichier texte déposé dans le répertoire principal d'un site Web pour interdire l'accès aux robots de certaines pages ou sous-répertoires du site.

**SILENCE**

Désigne l'ensemble des documents pertinents non retrouvés lors d'une recherche.

**SITE FEDERATEUR**

*Synonymes : site de référence, portail spécialisé*

Site spécialisé sur un thème précis (ex. le vin) proposant plusieurs types de ressources. Par exemple : répertoire spécialisé, liens vers des répertoires ou pages de liens spécialisées, articles en texte intégral ou bibliographie en ligne, actualités du secteur, événements du secteur, accès à base(s) de données, etc. Le fin du fin d'un site fédérateur consiste à créer une communauté autour de lui (via forums, newsletters, etc) pour devenir le point de référence du domaine.

**SPAMDEXING**

Création ou modification d'un document avec l'intention de tromper un répertoire ou un système de classement automatique. Toute technique visant à augmenter la position potentielle d'un site aux dépens de la qualité du corpus de l'outil de recherche peut également être considérée comme du spamdexing.

**SPIDER**

Voir Araignée

**TECHNIQUES DE POSITIONNEMENT**

Stratégies qui consistent pour le concepteur d'un site ou le référenceur à modifier sa page Web afin que les moteurs de recherche traitent la page comme la plus appropriée pour une requête spécifique, ou un ensemble de requêtes.

**TRI PAR PERTINENCE**

Méthode de classement automatique des résultats retournés par le moteur de recherche qui s'appuie sur le calcul d'un score pour chaque réponse. La pertinence est alors basée sur des facteurs comme :

- le poids d'un mot déterminé par sa place dans le document
- la densité : fréquence d'occurrence dans un document par rapport à la taille du document
- le poids d'un mot dans la base et sa fréquence d'occurrence dans toute la base
- la correspondance d'expression : similarité entre l'expression de la requête et l'expression correspondante dans un document
- relation de proximité : proximité des termes de la requête entre eux et dans le document.

Cette technique est apparue avec la 1<sup>ère</sup> génération de moteurs de recherche (depuis 1994) et présente l'inconvénient d'être facile à détourner par les référenceurs peu scrupuleux (Cf. Spamdexing). Elle est utilisée par des moteurs comme AltaVista, Excite, Inktomi, Voila...

**TRI PAR POPULARITE**

Méthode de classement automatique des résultats retournés par le moteur de recherche qui

s'appuie soit sur le principe de citation soit sur la mesure de l'audience. Dans le premier cas, l'importance d'une page est évaluée en fonction des liens hypertexte qui pointent vers elle et en fonction de la nature du document qui la cite. Le tri est alors indépendant du contenu, mais les documents récents ou peu cités par les autres sont défavorisés. Google avec son système de PageRank en est l'exemple le plus connu.

Dans le second cas, l'importance d'une page est fonction du nombre de visites reçues lors d'une requête sur un moteur de recherche. C'est l'analyse du comportement de l'internaute qui détermine la popularité d'une page par rapport à un mot-clé. Solution DirectHit utilisée par HotBot et LookSmart par exemple.

## URL

*Uniform Resource Locator* : adresse d'un site Web.

## WEB (ou WWW ou Web)

"Toile d'araignée mondiale" : outil logiciel multimédia et hypertexte permettant d'effectuer des recherches de tous types sur le réseau, l'accès à l'information recherchée et sa visualisation. Cet interface graphique, accessible via Netscape ou Internet Explorer, a permis l'explosion actuelle d'Internet. Les utilisateurs peuvent y créer, y éditer ou y rechercher des documents. La taille du Web est en constante augmentation, dépassant allègrement le milliard de sites.

## WEB INVISIBLE

Expression qui sous-entend "la part du Web invisible pour les moteurs de recherche" : l'ensemble des pages non localisables et/ou non indexables par ces outils. Le Web invisible correspond à plusieurs types de ressources :

- Documents dans des formats différents du html (par exemple pdf, word, etc. )
- Pages situées à l'intérieur d'une frame (cadre)
- Pages dont les caractéristiques techniques rendent difficiles, sinon impossible l'indexation par les moteurs : javascrrips modifiant le contenu, technologies propriétaires (par exemple flash, active X, java)
- Pages qui n'ont fait l'objet ni d'un référencement direct , ni d'aucun lien d'une autre page.
- Pages nécessitant une identification de la part de l'internaute
- Pages dont le contenu indique aux moteurs qu'ils ne doivent pas l'indexer
- Page produite à partir de bases de données ou d'applications, et dont l'URL comporte des paramètres non exploitables par la plupart des moteurs
- Page produite à partir de données saisies par l'utilisateur via un formulaire html. Exemple : les résultats de l'interrogation d'une base de données avec des critères de recherche entrés par l'utilisateur.

## XML

Pour eXtensible Markup Language : langage de description et d'échange de documents structurés. Il est le résultat de la coopération d'entreprises et de chercheurs partenaires du World Wide Web Consortium (W3C) dont l'objectif a été de définir un formalisme permettant de regrouper les concepts d'hypertextes, de bases de données, de formats d'échange et de publication.

\*\*\*

*Ces définitions ne sont pas normatives. Pour certaines, elles sont issues de discussions passionnées échangées entre les formateurs Internet de l'ADBS. Elles visent avant tout à éclaircir un vocabulaire spécialisé et encore fluctuant.*