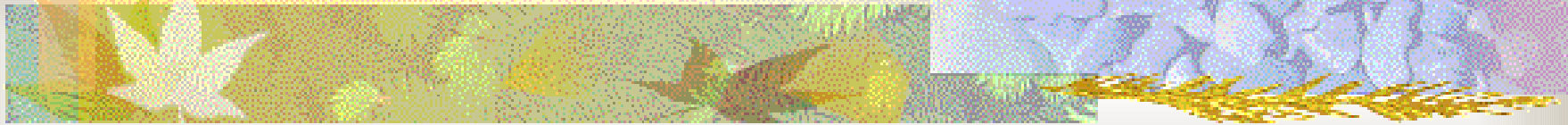




*la recherche documentaire en général
et en particulier sur l' internet.*



notions de base et astuces.



Les différentes étapes d'une recherche efficace : Définition des besoins (qualitatif et quantitatif) .

La première étape consiste à **faire le point** sur le sujet avant de commencer la recherche. On tente de répondre aux questions :

- ◆ Qui?
- ◆ Quand?(couverture chronologique)
- ◆ Quoi?
- ◆ Où?(champ géographique)
- ◆ Comment?
- ◆ Pourquoi?

On peut tenter de définir le sujet par la négative, se demander à quoi il est apparenté, quelles sont ses composantes, ses différentes formes, etc.

La quantité et le type d'information.

La décomposition du sujet permet de définir **la quantité et le type d'information** que l'on cherche. Cela est très utile pour choisir le type de document à utiliser (par exemple une information ponctuelle se recherche dans un ouvrage de référence : dictionnaire Le Petit Robert, Quid, encyclopédie Universalis).

TYPE DE TRAVAIL	QUANTITÉ	STYLE	SOURCE	NIVEAU
résumé	1	synthétique	dico, encyclo	simple
exposé	10	synthétique + exemples	+ ouvrage de références	simple, clair, pédagogique
travail de fin d'année	15-20	analytique	+ quelques livres et articles (sélection)	clair, avec vue d'ensemble complète
travail de fin d'études	20-25	analytique approfondi	+ dépouillement de plus de doc, différents centres de documentation.	rigueur, précision, détails (éventuellement avec un glossaire)

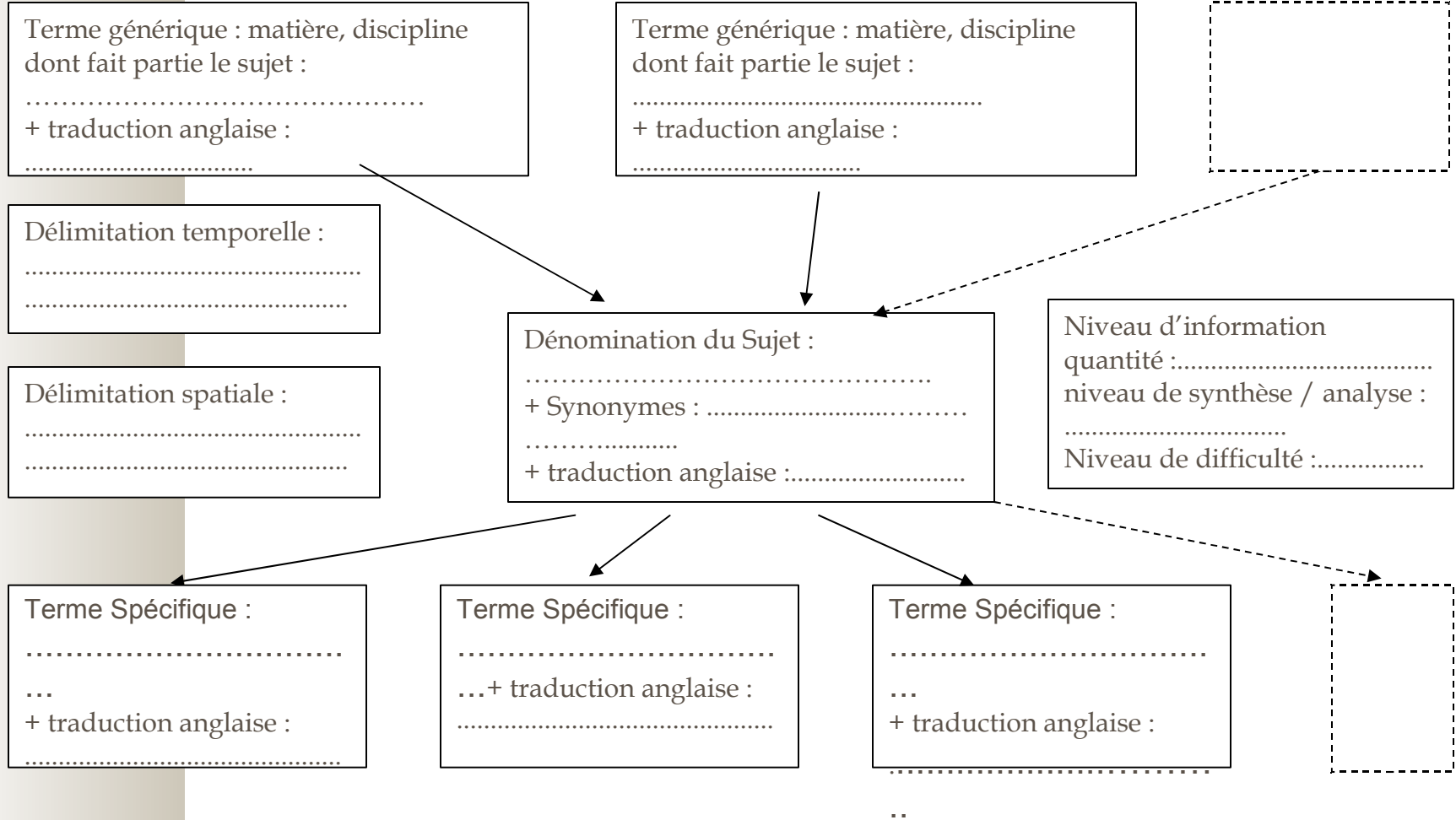


les synonymes et les traductions.

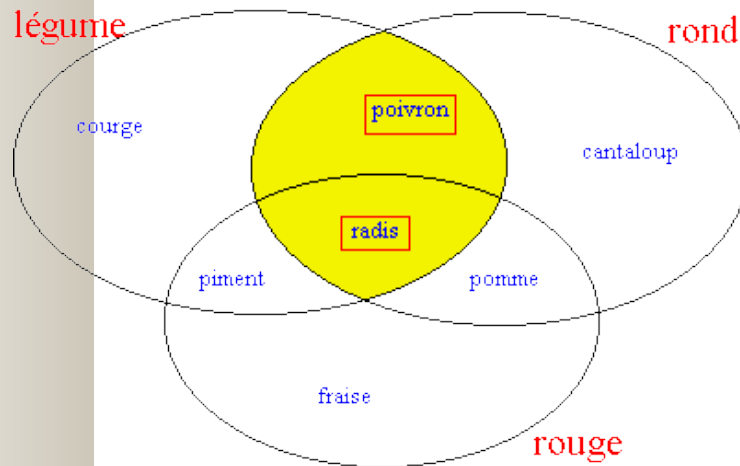
Enfin, il faut **transformer** le sujet **en mots-clés**.

Pour cela, il faut utiliser un dictionnaire pour repérer les synonymes ou les traductions pour la recherche dans d'autres langues (surtout l'anglais sur Internet). Si un de vos mots n'a pas d'équivalent dans le langage documentaire, élargissez à une notion plus générale (terme générique <> terme spécifique).

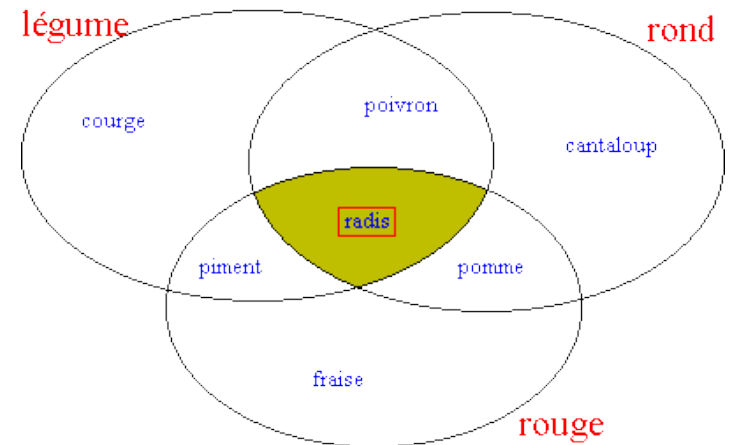
Tableau global des mots clés.



Les opérateurs logiques (Booléens). ET (AND)

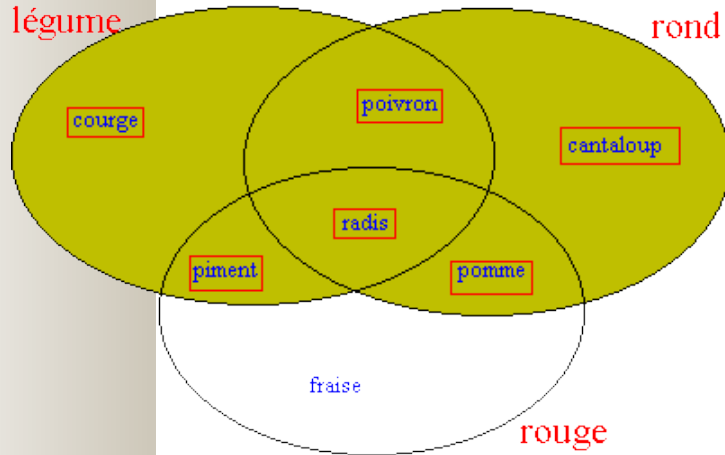


Légume ET Rond

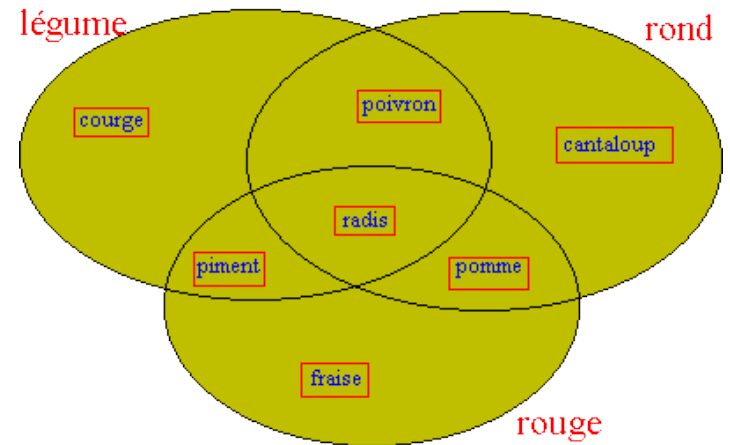


Légume ET Rond ET Rouge

. Les opérateurs logiques (Booléens). OU (OR)

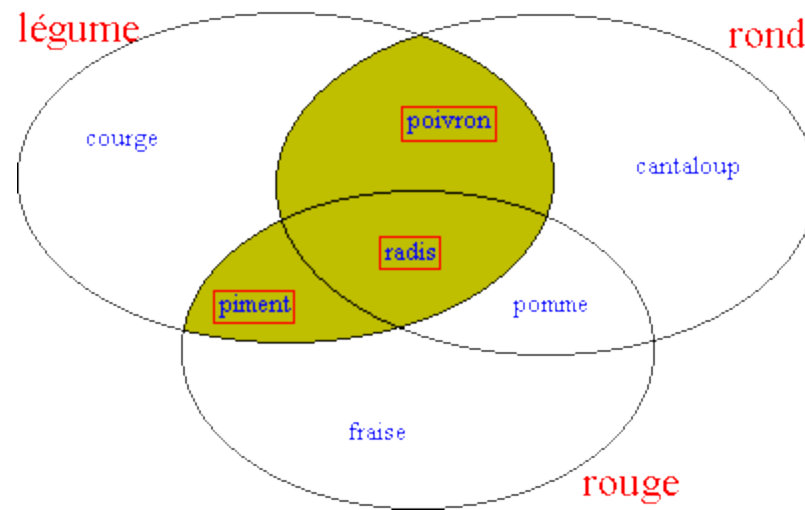


Légume OU Rond



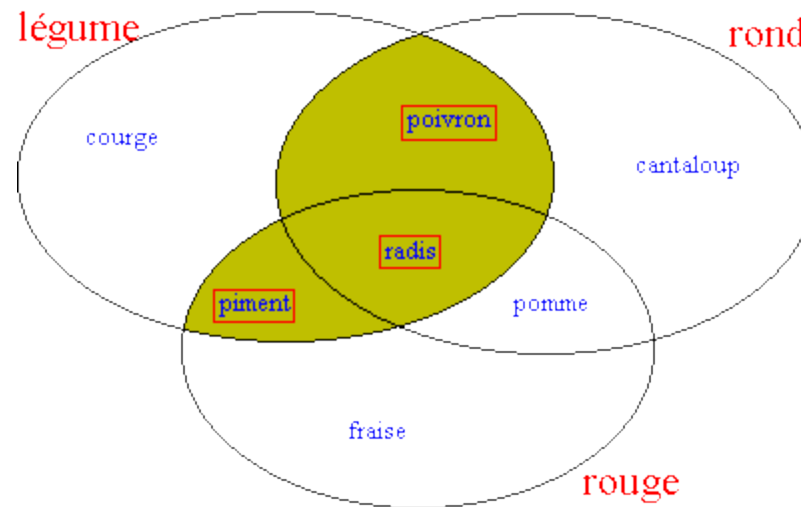
Légume OU Rond OU Rouge

Combinaison à différents niveaux



Légume ? Rond ? Rouge

Combinaison à différents niveaux : la solution



Légume ET (Rond OU Rouge)



Exemple : Quel est l'apport du numérique en photographie astronomique ?

QUI ? s.o.

Où ? s.o.

Quand ? s.o.

Quoi ? numérique en photo astronomique

Comment ? relation entre les deux

Synonymes et termes proches :

Numérique -> digital

Photographie astronomique -> céleste, astres, ciel étoiles, planètes, galaxie, télescope, etc.

Terme générique : photographie scientifique

Traductions :

Numérique -> digital

Photographie astronomique -> astronomic photography et surtout : celestial photography



7.4. Exemple en cinéma :

- Sujet : la direction photo dans le cinéma allemand et français de l'entre-deux guerres :

(« direction photo » OR eclaireage OR light OR cinematographer) AND (cinema) AND (allema* or Franc* OR german*) AND (192* or 193* or « entre-deux guerres» OR interwar OR « between the wars »)



1. *Internet...*

est composé d'ensembles de pages regroupées sur des sites et éventuellement reliées entre elles.

La « page » est le fichier qui s'affiche en une fois sur l'ordinateur. C'est l'unité de base d'internet.

Un « site » est un ensemble de pages reliées entre elles (page d'accueil et pages liées).

Un « serveur » est un ordinateur qui héberge un ou plusieurs sites.

Les liens entre les pages permettent de naviguer sur internet, ce sont les « liens hypertextes » .

Statistiques de noms de domaine de Whois source

		Daily Changes(last 24hrs)			
Active	Deleted	New	Expired	Transferred	TLD
51,066,573	128,932,837	1,272,968	1,366,561	138,283	.Com
7,364,451	14,072,793	55,672	134,206	25,399	.Net
4,703,735	11,844,244	158,052	20,228	3,351	.Org
2,780,724	3,557,197	3,219	2,175	8,752	.Info
1,383,455	779,954	1,078	998	1,079	.Biz
1,006,175	788,172	746	826	1,734	.Us
68,436,948	160,636,774	1,491,735	1,524,994	178,598	Total
Last Updated : 2006-05-09					

<http://www.internetworldstats.com/stats.htm>

<http://www.domaintools.com/internet-statistics/>



Les informations sur internet.

- ◆ **Quantité astronomique**

(estimation de plusieurs centaines de milliards de pages)

- ◆ **Grande volatilité de l'information**

(manque de permanence dans le contenu ou la forme)

- ◆ **Qualité extrêmement variable**

(liberté d'expression, peu ou pas d'étapes de validation)

- ◆ **Difficulté croissante pour s'y retrouver**

(choix et utilisation de l'outil, les faiblesses des outils eux-mêmes)



Le Web visible

Une petite partie des informations est indexée (environ 8 milliards de pages) :

- =>celles qui sont facilement accessible aux outils
- =>celles qui font la démarche pour se faire connaître des outils

Beaucoup de sites commerciaux mais aussi :

- Des sites émanant d'organismes, d'institutions (bibliothèques, universités, asbl, organismes nationaux et internationaux,...) ou d'états (ministères)
- Des sites personnels



Les outils : 1 les répertoires (directories)

Les *répertoires* (ou annuaires) de sites web sont réalisés par une équipe de personnes qui sélectionnent et rassemblent dans des catégories et des sous-catégories thématiques les sites qui leur paraissent les plus intéressants.

(www.yahoo.fr)

<http://www.dmoz.org/World/Français/>

www.lii.org (avec commentaires)

www.blogolist.com (weblog)



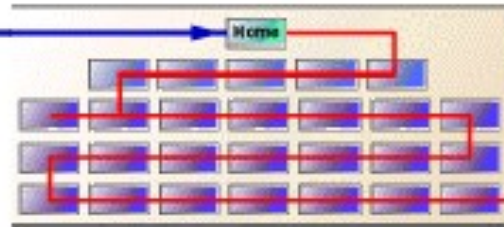
Les outils : 2 : les moteurs (webcrawlers, search engines).

Les *moteurs* sont des programmes (robots) qui recherchent automatiquement une chaîne de caractères à travers un certain nombre de pages indexées sur le réseau.

La qualité du moteur dépend du nombre de pages indexées, des possibilités de requêtes, de la présentation des résultats (tris, exclusion des doublons) et de l'évaluation éventuelle des pages trouvées.



Web spider



Science	Flowers
Can	Art
Travel	Stamps
Weather	Football
Schools	Soccer
Travel	Tennis
Theaters	Swimming
Animals	Cruises
Tractors	Rail
Doctors	Car
Music	Tre
Word	Re
Records	Re

Builds list of words and notes where they were found




Builds index based on its own system of weighting

Encodes the data to save space



Stores data for users to access



Les différents moteurs ont des index différents car ils n'utilisent pas les mêmes critères et parcourent le réseau à des fréquences différentes. Selon la quantité de requêtes qu'ils reçoivent à un moment donné, ils consacreront un temps plus ou moins long à chaque requête.

Au sein même de Google, les data center peuvent avoir des index qui varient l'un par rapport à l'autre! Le plus complet étant <http://66.249.93.104/>

Pour être complet, il faut donc interroger différents moteurs à différents moments!

Pas de confiance aveugle en GOOGLE!

Voir :

<http://www.lesmoteursderecherche.com/index.htm>



Les faiblesses des moteurs de recherche :

- ◆ Ils n'indexent que le début des pages;
 - ◆ Ils n'indexent que du texte (les noms et informations « metadonnées » sur les images et les sons)
 - ◆ Ils n'indexent que les premiers niveaux des sites.
 - ◆ Ils n'indexent que ce qui n'est pas protégé par mots de passe (privé ou payant)
 - ◆ Ils n'indexent pas (ou mal) les pages dynamiques des bases de données
 - ◆ Ils n'indexent pas tous les types de fichiers (.pdf, .doc, etc)
- voir : <http://searchenginez.com/filetype.html>

Vous pouvez lire le texte ci-dessous :

Efficacité économique et réinsertion sociale

La REVUE 27/02/1988

L'Atelier protégé « FAC-AP », de Molenbeek, se lance dans l'édition de livres (sur la Révolution française et la protection animale)

Créé depuis 1982, et établi depuis janvier 86 au 14 de la rue Vandennepereboom à Molenbeek, l'atelier protégé « FAC-AP » occupe actuellement 70 personnes, handicapées physiques ou mentales. Jusqu'ici, il concentrait ses activités sur le conditionnement le façonnage, l'impression d'emballages et de dépliants. Il se lance aujourd'hui dans l'édition pure et simple, avec deux livres qui viennent à leur heure, un album « Bicentenaire de la Révolution française » et un recueil juridique « Le droit de la protection animale en Belgique ». De l'un comme de l'autre, les responsables de l'atelier espèrent beaucoup, pour assurer la viabilité de l'atelier et son développement, dans la double optique qui régit son fonctionnement : allier effi-

cacité économique et réinsertion sociale.

Avec ses 70 travailleurs, l'Atelier protégé « FAC-AP » est une des plus importantes institutions protégées de la région bruxelloise. Mais ses responsables actuels, MM. Janray et De Gorter, ne cachent pas leur objectif, occuper, d'ici un à deux ans, 150 à 200 travailleurs. Ce n'est certes pas une sinécure, car, disent-ils, gérer un atelier protégé nécessite l'équilibre délicat de deux impératifs parfois contradictoires : efficacité économique et réinsertion sociale. « Sans subsides, il est évident que nous n'en sortirions pas, pour l'instant », dit M. Janray. « Nous sommes reconnus et soutenus par le Fonds national de reclassement des handicapés, mais celui-ci tarde à nous payer,

Ainsi, actuellement, un million de subsides sont en attente. Quand à la Province, pour la première fois, elle nous alloue 500.000 francs de subsides. Quant à l'O.N.E.M., il ne nous paie pas depuis un an les 500.000 francs qu'il nous doit pour les 5 C.M.T. (chômeurs mis au travail) que nous employons. »

★ A côté de ces C.M.T. et de l'équipe d'encadrement (4 personnes), l'atelier emploie des handicapés physiques légers ou graves, ainsi que des handicapés mentaux. Le salaire horaire de base est de 101,5 francs. Avec l'ancienneté et la qualification, il peut monter à 180 francs. Pour la plupart, ce salaire s'ajoute aux allocations qu'ils touchent du Fonds national de reclassement.

Ce n'est sans doute pas, comme on dit, la gloire. Mais M. Janray insiste sur le fait que la situation de l'atelier, n'est pas non plus très confortable. Ainsi, l'achat du bâtiment actuel a coûté 9 millions : 5 millions payés par le Fonds national, les 4 autres payés par l'Atelier sous forme d'un emprunt.

D'où la seconde exigence, celle de rentabilité économique. Dans ce but, l'atelier, qui au départ ne fabriquait que du matériel de bureau, a diversifié ses activités. Ainsi, au 14 de la rue Vandennepereboom, on trouve d'abord un atelier de façonnage et de conditionnement, fabriquant toutes sortes d'emballages et de boîtes. L'atelier s'est aussi spécialisé dans le traitement du plastique,

qu'il s'agit de souder (pochettes, chemises, parasols) ou d'emballages. Il existe aussi un secteur « fournitures scolaires ».

Mais le fleuron de l'atelier est sans doute son imprimerie, petite mais très complète. A partir de matériel acquis par occasion, par prêts d'autres ateliers et par douelle dispose de la photocomposition et est informatisée. Ce qui lui permet, à côté des travaux classiques (impression de dépliants, de sachets, d'emballages), de s'adonner à des tâches plus spécialisées. A commencer par l'emballage et l'impression de jeux « Wall Street junior », un jeu de bourse « pour rire », conçu en Belgique il y a deux ans, et qui, on s'en doute, connaît son petit succès depuis quelques mois : en 2 mois, 6.400

exemplaires sont sortis de l'atelier, et 5.000 sont en cours de fabrication, destinés à l'exportation, notamment en France.

Cette imprimerie se lance maintenant dans un nouveau pari et franchit un nouveau pas : l'impression est l'édition de livres à part entière. En l'occurrence, un album « Bicentenaire de la Révolution française » et un recueil sur « Le droit de la protection animale en Belgique ». L'un comme l'autre, on ne peut le nier, viennent à leur heure. En tout cas, les responsables de l'atelier comptent sur ces deux publications pour conforter la viabilité de l'atelier, et mieux le faire connaître à l'extérieur, à l'occasion de ce qui est une « première » en Belgique pour un atelier protégé.

Laurent WILEN.



Rassemblement des compétences pour faire progresser l'atelier : de gauche à droite, M. Adant (auteur de « La Protection animale »), M. Rohr (promoteur du jeu « Wall Street junior »), Janray (directeur de l'atelier) et Gillet (coauteur du « Droit de la Protection animale »).



• L'atelier de fabrication des boîtes.

Mais pas google :

[Connexion](#)



[Web](#) [Images](#) [Groupes](#) [Annuaire](#) [Actualités](#) [plus »](#)

"efficacité économique et réinsertion sociale"

Rechercher

[Recherche avancée](#)

[Préférences](#)

Rechercher dans : Web Pages francophones Pages : Belgique

Web

Conseil : Pour obtenir plus de résultats, supprimez les guillemets.

Aucun document ne correspond aux termes de recherche spécifiés ("**efficacité économique et réinsertion sociale**").

Suggestions :

- Vérifiez l'orthographe des termes de recherche.
- Essayez d'autres mots.
- Utilisez des mots plus généraux.

[Page d'accueil de Google](#) - [Publicité](#) - [Solutions d'entreprise](#) - [À propos de Google](#)

©2006 Google

Pourtant l'information existe sur internet à l'adresse:
<http://www.arows3.com/RF07-Press-Lanterne.jpg>



6.3.

Voici quelques moteurs généralistes :

www.google.fr

www.alltheweb.com

www.altavista.fr

[http://www.ask.com/](http://www.ask.com)

Des moteurs plus spécialisés :

www.sciencedirect.com

www.voilà.fr (documentation francophone)

www.blogpulse.com (blog)

Des répertoires de moteurs :

www.searchengines.com

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



Quelques astuces à propos des moteurs :

- ◆ Renseignez vous sur l'opérateur booléen par défaut ou utilisez la recherche avancée
- ◆ Utilisez les mots les plus caractéristiques du sujet
- ◆ Ecrivez les mots en minuscules et sans accent
- ◆ Utilisez le « OU » pour lier les synonymes, les traductions ou les différentes formes d'un mot si vous hésitez sur son orthographe.
- ◆ L'ordre des mots a parfois de l'importance (le plus important d'abord)
- ◆ Utilisez les touches CTRL F pour retrouver un mot dans une page




La meilleure utilisation de Google : recherche avancée

Expression exacte entre guillemets :

ex. photo de mode <-> « photo de mode »

Recherche dans les titres et les adresses :

ex. : « photo de mode » AND (intitle: liens OR inurl: liens)



Les métamoteurs ou « metacrawlers ». Ils ne traitent pas eux-mêmes la requête mais l'envoient à un ensemble plus ou moins complet de moteurs de recherche. Par exemple, le programme *Copernic*.

Quelques exemples :

www.profusion.com

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

www.kartoo.fr (présentation des résultats en cartes)

Il existe des sites de spécialistes en recherche d'information :

<http://c.asselin.free.fr/>



Les encyclopédies et sites de références :

encyclopédie libre wikipédia :

<http://fr.wikipedia.org/wiki/Accueil>

Sites de dictionnaires :

<http://www.lexilogos.com/index.htm>

<http://les-dictionnaires.com/>

<http://www.yourdictionary.com/languages.html>

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

Pour les synonymes :

<http://elsap1.unicaen.fr/cgi-bin/cherches.cgi>



Les catalogues en ligne de bibliothèques :

www.worldcat.org



Les moyens de chercher dans l'invisible ou « web profond »

Il existe quand même quelques outils :

Répertoires sélectifs ou thématiques et/ou de sources gratuites :

<http://signets.bnf.fr/>

<http://infomine.ucr.edu/>

www.invisible-web.net

www.completeplanet.com (70.000 bd)

<http://www.culture.gouv.fr/culture/bdd/index.html>

<http://urfist.uni-lyon1.fr>

Pour la « littérature grise » :

<http://whitepapers.zdnet.com/>

Archives ouvertes :

www.ccsd.cnrs.fr

www.doaj.org

<http://arxiv.org>

Archive :

www.archive.org

les bibliothèques numériques :

http://affordance.typepad.com/mon_weblog/2006/09/bibliothques_nu.htm

12. Evaluer les résultats de la recherche :

Attitude très courante et pourtant fausse :

je l'ai lu dans un livre

je l'ai vu à la télé

je l'ai trouvé sur internet

c'est vrai!



4 éléments sont à analyser :

fiabilité de la source et de l'auteur :

quel type de contenu : un article? Site commercial, post de blog, forum, extrait de livre
par ex. Retaper le nom de l'auteur dans google pour voir sa notoriété

Qui cite la page? Type d'extension?

qualité du contenu

cohérence de l'organisation

pertinence : le résultat correspond-il à mon besoin documentaire?




Sur Quel site suis-je? : site trouvé grâce à un lien sponsorisé dans une liste de résultats?

Importance du suffixe :

Ils sont répartis en deux grandes familles.

D'un côté, les suffixes génériques, qui comportent au moins trois lettres :com,.edu,.net,.mil,.gov,.int,.biz,..info,.coop,.aero,.museum,.

Tous les autres suffixes comportent deux lettres et sont dits géographiques.
Il y en a 240 en tout, généralement un par Etat et pour certains territoires rattachés à un pays possédant déjà son propre suffixe.

- 
- ◆ Les sources sont-elles citées?
 - ◆ Les pages sont-elles rédigées sans fautes? (Cela peut indiquer si la page a été relue par une tierce personne ou si la responsabilité d'une organisation sérieuse est engagée).
 - ◆ Les graphiques et schémas sont-ils clairement présentés avec explications, unités, etc.?
 - ◆ Quelle est la fraîcheur de l'information? (Date de mise à jour?)
 - ◆ Y-a-t-il des publicités? Si oui, sont-elles clairement séparées du contenu du document?
 - ◆ Peut-on vérifier le sérieux de l'auteur et /ou de l'organisation qui héberge la page?
 - ◆ L'organisation ou l'auteur ont-ils intérêt à ce qu'une solution s'impose face aux autres?



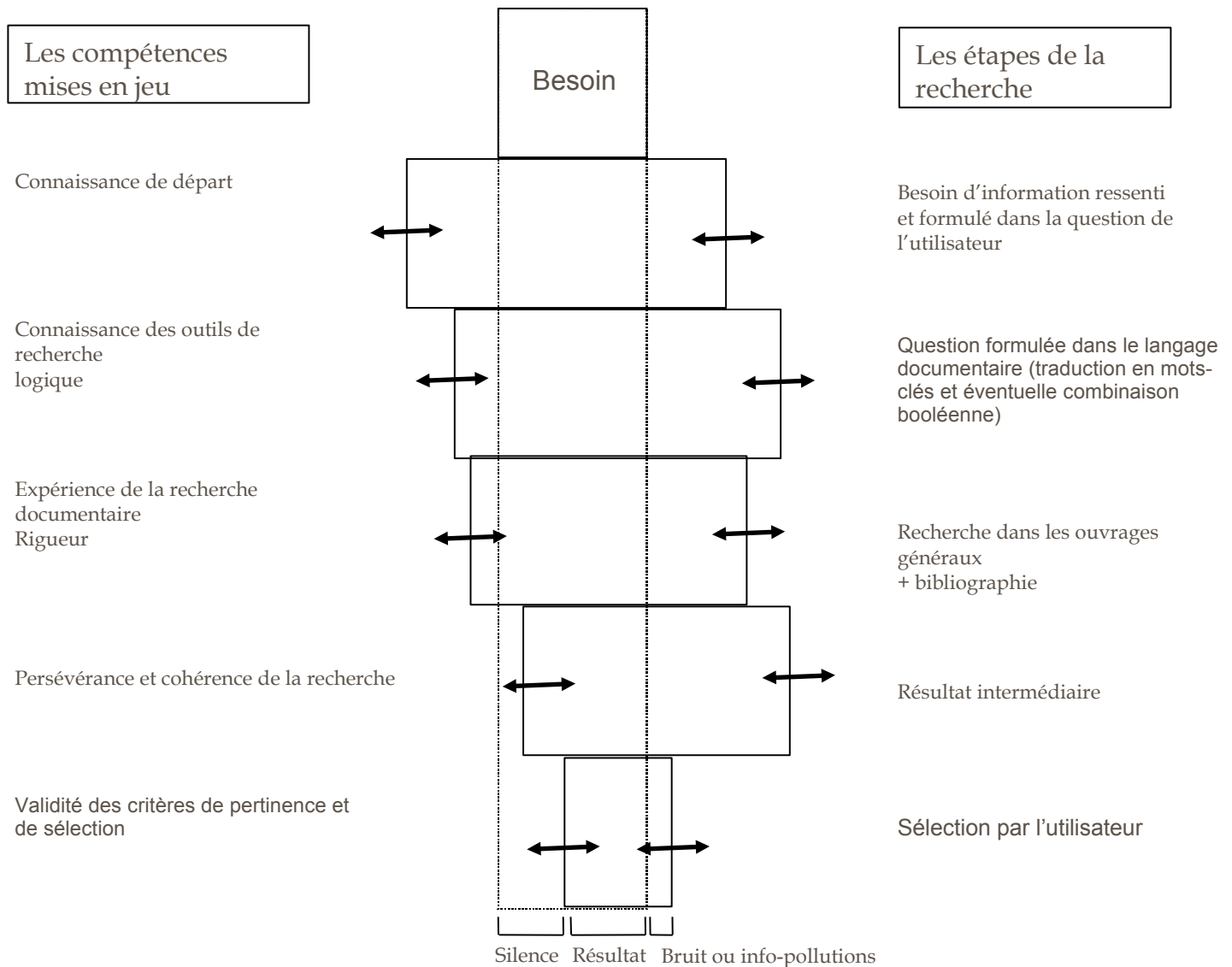
Le suivi d'une recherche : la veille

but : se tenir au courant des nouveautés dans un domaine
Relancer régulièrement les recherches (en limitant par date pour obtenir les nouveautés)

les RSS (Really simple syndication) : envoi automatique et régulier de résumés sur un sujet ou d'une source.
(après inscription et installation du logiciel nécessaire)

Ex :

Tableau récapitulatif





Résumé

- ◆ Internet est complexe.
- ◆ Utilisation des outils adaptés à vos besoins (moteurs et/ou annuaires).
- ◆ Connaître les outils que l'on utilise (recherche avancée).
- ◆ Ne pas confondre ET et OU.
- ◆ Ne pas oublier le web invisible.
- ◆ Définition claire des besoins documentaires avec synonymes et traductions.
- ◆ Évaluation **critique** des résultats.