

**Le travail de recherche documentaire**

Un guide pour la documentation scientifique

<http://webdoc.snv.jussieu.fr>

**Marie-Agnès Gheeraert et Bernard Billoud**  
Université Pierre et Marie Curie

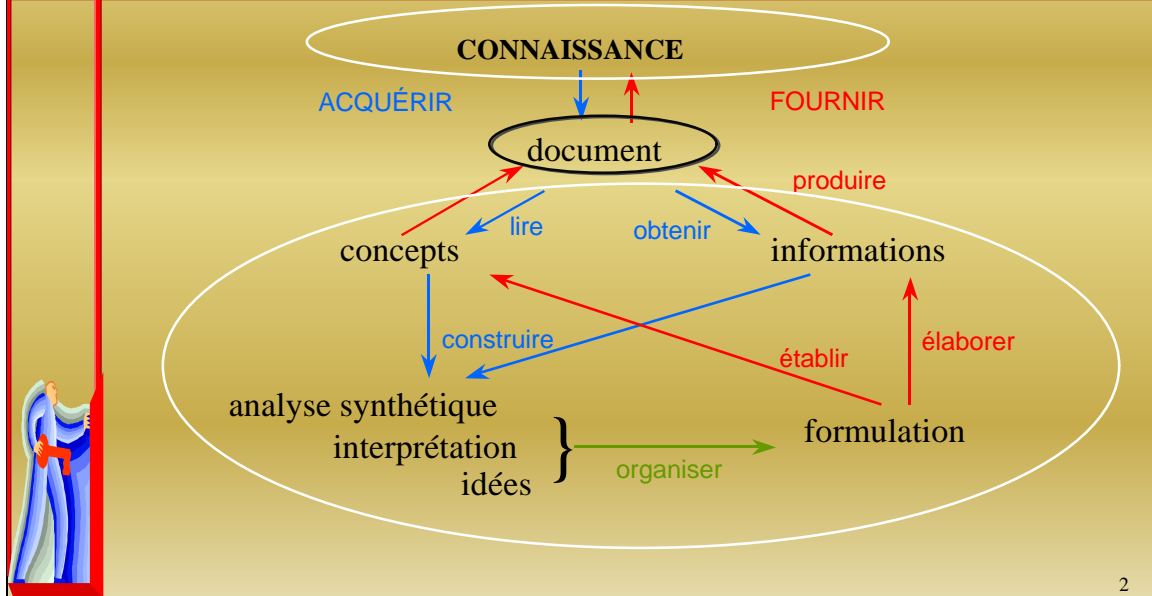
©UPMC

1

Devant la masse documentaire disponible en bibliothèque ou sur l'Internet, il est important d'acquérir une méthode apportant l'efficacité dans la recherche de documents.

Décomposer les processus mis en œuvre intuitivement au cours de nos recherches, les comprendre, les renforcer et les intégrer dans la méthode proposée ici.

## Qu'est-ce que se documenter ?



2

- Lorsqu'on débute une réflexion sur un sujet, même si l'on a quelques *idées*, on se *documente* pour en *acquérir* une meilleure *connaissance*. Nos *lectures* apportent *conceptes* et *informations*. La *construction* d'une *analyse synthétique* et d'une *interprétation* génère de nouvelles *idées*.
- Élaborer* une nouvelle *information* et *établir* de nouveaux *conceptes* en *organisant* et *formulant* nos idées lors de la rédaction (étayée et référencée). On *produit* alors un nouveau *document*, qui *fournit* ainsi de nouvelles *connaissances*.
- Le lien entre la démarche **individuelle** (partie inférieure du schéma) et la **société** (en haut) est le **document** (au centre).

## Qu'est-ce que le document ?

- Le document est le mode individuel d'accès à la connaissance
- le document est le mode collectif de diffusion de la connaissance

*L'aspect physique du document n'est pas le plus important.*

*Pourtant :*

- La reproduction (par l'imprimeur), la diffusion (par l'éditeur), la conservation (par le bibliothécaire) d'un document confèrent à son contenu la propriété de s'exporter dans l'espace et dans le temps



Cependant il existe d'autres façons d'acquérir des connaissances: suivre un enseignement, observer, expérimenter etc.. Le document se caractérise par la propriété du langage écrit d'une part, qui permet la communication en différé entre deux individus non co-présents, et d'autre part, l'apport des technologies mises en œuvre autour de la conservation et la diffusion, en sa qualité de média.

## Les ressources documentaires et les supports de l'information



texte	livres, magazines	lecture linéaire
image	photos, dessins	observation
planches	affiches, tracts	lecture synthétique
son	cassettes audio, CD	écoute, linéaire
vidéo	cassettes vidéo, DVD	écoute et observation, linéaire
hypertexte	Internet, CD	lecture, non linéaire par «nœuds», active
hypermédia	Internet	lecture synthétique, active

Le document, véhicule de l'information, se retrouve sous différents supports.

## La recherche documentaire: le thésaurus personnel

- la méthode: dresser une liste de mots avec termes apparentés caractérisant notre recherche.  
Ex: léopard ou panthère, chasse, carnivore, relation prédateur-proie
- la démarche: se définir un thésaurus personnel à l'aide d'un dictionnaire ou d'un thésaurus:
  - permet de clarifier sa problématique
  - permet d'approcher le classement de la bibliothèque
  - permet une recherche documentaire plus efficace car mieux ciblée



Débutons notre recherche. Notre sujet concerne le comportement de chasse du léopard. Comment trouver et sélectionner sa documentation ? Par où commencer ?

## Les langages documentaires: descripteurs et thésaurus

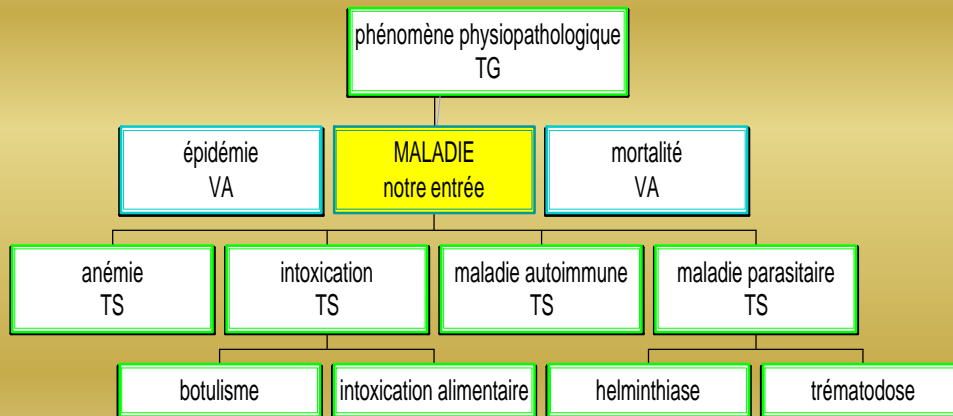
- Un **descripteur** ou mot-clé (un mot ou une expression) est choisi pour représenter un concept. C'est un terme dit préférentiel (ex: *durée*), par rapport à un autre, dit non préférentiel, comme un synonyme (ex : *longueur* est ambigu).
- Le thésaurus est une liste alphabétique de concepts (termes préférentiels) et de renvois (termes non préférentiels) couvrant un domaine particulier et présentant trois types de relations sémantiques: la relation d'équivalence, la relation hiérarchique et la relation associative.



Les professionnels de l'information, bibliothécaires et documentalistes, utilisent les langages documentaires, vocabulaire normalisé et structuré pour l'indexation. L'indexation ou analyse documentaire est la description structurale et sémantique du document.

le classement et le rangement des ouvrages dans la bibliothèque est fonction de leur cote de classification, établie à partir de l'indexation.

## Les langages documentaires: extrait d'un thésaurus



7

Un thésaurus établit les relations de filiation entre les concepts représentés par les termes: il peut donc prendre la forme d'un arbre généalogique, comme dans cet exemple issu de la biologie médicale.

## Les langages documentaires: la relation d'équivalence dans un thésaurus

- un même terme peut référer à plusieurs concepts et un même concept peut être désigné par plusieurs termes.
  - Les synonymes ( traduction directe par un descripteur)
    - ❖ variante orthographique (CLE = CLEF)
    - ❖ traduction étrangères d'usage courant (LOGICIEL = SOFTWARE)
    - ❖ évolution du langage (BIODIVERSITE = DIVERSITE BIOLOGIQUE)
    - ❖ synonymie sémantique (VITAMINE C = ACIDE ASCORBIQUE)
    - ❖ synonymie d'usage (THIOL = THIOALCOOL)
    - ❖ jargon ou langage populaire (BREUVAGE = BOISSON).
  - Les quasi synonymes (significations différentes mais notions voisines)
    - ❖ notions jugées trop fines pour le thésaurus en question, renvoyés sur un terme plus général
    - ❖ variantes d'un même concept (FLUIDITE = VISCOSITE)
    - ❖ termes antonymes (STABILITE = INSTABILITE)





## Les langages documentaires: les relations hiérarchiques dans un thésaurus

- La relation hiérarchique **sémantique**  
exprime les rapports de généralisation et de spécialisation entre notions
  - chaque terme ou notion spécifique représente un cas particulier du terme générique et hérite des caractéristiques de ce dernier.
    - ❖ TG : SCIENCES NATURELLES . TS : BOTANIQUE
  - l'organisation peut être soit monohiérarchique soit polyhiérarchique
    - ❖ TASTE BUDS : BT LINGUAL EPITHELIEUM ; BT SENSE ORGANS ; BT TASTE ; BT TONGUE
- la relation hiérarchique **partitive**
  - CORPS : TS TETE ; TS THORAX ; TS ABODOMEN
- la relation hiérarchique **d'instance**
  - REGIONS MONTAGNEUSES : TS ALPES



## Les langages documentaires: les relations d'association dans un thésaurus

- Un concept est associé à un autre comme étant relatif au même sujet, et pour en suggérer un autre aspect.

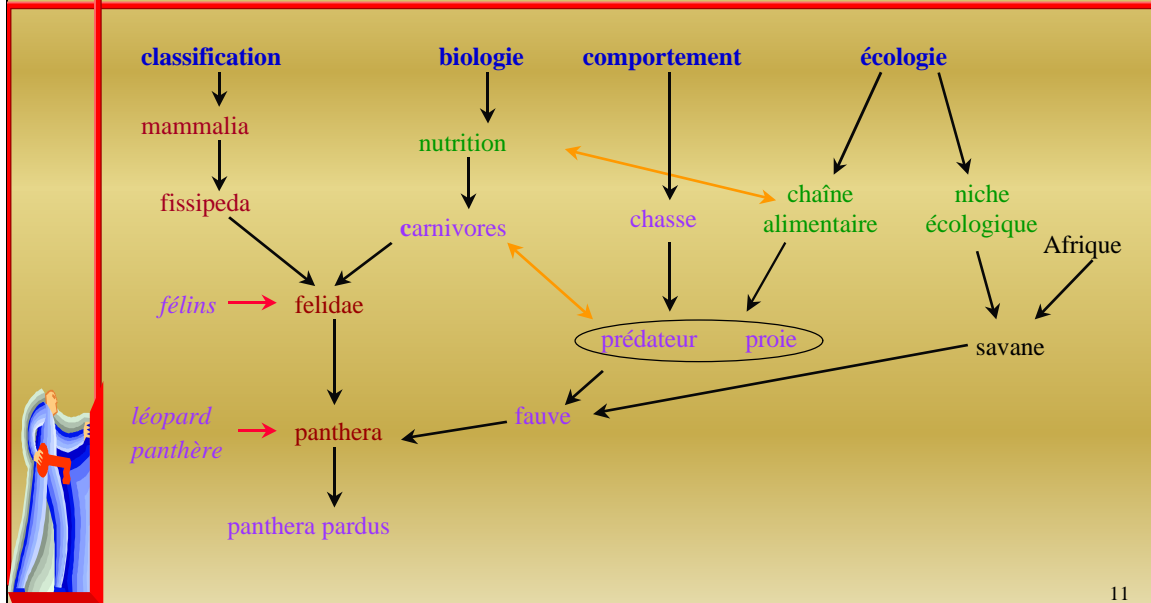
discipline - sujet - objet - propriété - procédé - application - processus - moyen - effet - mesure – évènement - historique - personnalité : NUCLEAIRE - ENERGIE - URANIUM - RADIOACTIVITE - ENRICHISSEMENT - ELECTRICITE - REACTION EN CHAINE - REACTEUR - RADIATION - BECQUEREL - THCHERNOBYL - MARIE CURIE

**Tableau des codes :**

Relation	Français	Anglais	symboles
Equivalence	Employer EM	Use US	-->
	Employé Pour EP	Use For UF	==
Hiérarchique	Terme générique TG	Broader Term BT	<
	Terme spécifique TS	Narower Term NT	>
Associative	Terme Associé TA ou voir Aussi VA	Related Term RT	---



## La recherche documentaire: exemple de construction d'un thésaurus



•En s'aidant du dictionnaire ou d'un thésaurus:

❖ 1ère étape: rechercher les descripteurs correspondant à nos termes ;

❖ 2ème étape: rechercher les termes génériques et les termes spécifiques relatifs à ces mots-clés,

Remarque: pour "*Panthera pardus*", le léopard, la position systématique de l'espèce constitue un cas particulier mais incontournable de hiérarchie sémantique. Nous commencerons donc par le situer dans la classification des espèces.

❖ 2ème étape toujours: établir les relations sémantiques hiérarchiques entre les descripteurs recensés (flèches orientées) ;

❖ 3ème étape: établir les relations d'association entre les descripteurs (doubles flèches).

La liste de mots-clés pourra être enrichie tout au long de la recherche documentaire, et à mesure des découvertes.

En définitive, il apparaît que l'étude du comportement de chasse du léopard recouvre des notions issues de systématique et de biologie, d'éthologie et d'écologie.

## Les langages documentaires: définition du thésaurus

- «Du point de vue de sa fonction, un thésaurus est un instrument de contrôle de la terminologie utilisé pour **transposer** en un langage plus strict (**langage documentaire**, langage d'information) le **langage naturel** employé dans les documents et par les indexeurs ou les utilisateurs. Du point de vue de sa structure, le thésaurus est un **vocabulaire**, contrôlé et dynamique, de termes ayant entre eux des **relations sémantiques** et **génériques** et qui s'applique à un domaine particulier de la connaissance» (Chaumier, 1988).



## Les langages documentaires: le catalogage

- Le classement, le rangement et l'indexation (**cote**) des ouvrages de la collection d'une bibliothèque se font à partir des descripteurs (résultat de l'analyse documentaire).
  - On retrouve facilement cette classification par thèmes dans le classement par rayonnage
  - La description complète (et standardisée) des ouvrages: les **notices bibliographiques** sont répertoriées sur
    - le catalogue
    - le fichier manuel (rangé en double par auteurs et par sujets)
    - les bases de données bibliographiques



## Les langages documentaires: les notices bibliographiques

Champ	livres	articles scientifiques	périodiques
<b>Numéro d'identification</b>	<b>ISBN</b> : <i>International Standart Book Number</i>		<b>ISSN</b> pour <i>Serial Number</i>
<b>Auteur(s)</b>	Auteur(s) de l'ouvrage	Auteur(s) de l'article	
<b>Titre</b>	Titre de l'ouvrage	Titre de l'article	Titre de la revue
<b>Source</b>	Maison d'édition (et/ou série, et/ou collection)	Titre du journal de publication, numéro (et volume ou tome)	Société d'édition (et/ou série, et/ou collection)
<b>Année</b>	Date de publication	Date de publication	
<b>Pages</b>	Nombre de pages	Pages de début et de fin d'article	
<b>Cote et disponibilité</b>	Cote et situation dans la bibliothèque		Bibliothèque et cote dans celle-ci
<b>Mots clés et/ou résumé</b>	sujets traités dans l'ouvrage extraits de l'analyse documentaire lors de l' <b>indexation</b> par les documentalistes		

14

La description structurale du document, la notice bibliographique, comporte plusieurs rubriques (en principe toujours celles-ci) dont le numéro d'identification unique pour un livre (ISBN) ou un périodique (ISSN). Les rubriques de la fiche constituent différents les **champs** qui structurent la base de données bibliographique.

Le contenu des notices diffère légèrement selon la nature de l'ouvrage : livre, périodique ou article paru dans une revue.

## La recherche documentaire: Où trouver les notices bibliographiques ?

- ouvrages contenus dans la bibliothèque
  - catalogues sous format papier
  - bases de données sur poste ou en ligne
    - ❖ Toutes les bibliothèques universitaires (SUDOC) <http://www.sudoc.abes.fr/>, Bibliothèque Centrale du Muséum Nat. d'Histoire Naturelle <http://muscat.mnhn.fr/mnhnbib/>, Médiathèque de la Cité des Sciences <http://www.cite-sciences.fr/>
- références bibliographiques d'articles scientifiques, bases de données sur poste ou en ligne
  - ❖ *BIOSIS* (biologie générale), Medline (biologie médicale) <http://www4.ncbi.nlm.nih.gov/PubMed/>, PASCAL <http://www.inist.fr/>
- la base de données *Myriade* contient le Catalogue Collectif National des Périodiques qui recense toutes les revues disponibles dans les bibliothèques de France. Sur CDRom ou 36 17 CCN



## Les langages documentaires: les langages d'interrogation

- requêtes simples:  
interroger une notice à l'aide d'un descripteur par champ (ex: auteur: *Darwin* ; sujet: *évolution*)
- requêtes composées:  
à l'aide de plusieurs descripteurs et d'opérateurs logiques
  - opérateurs booléens: **ET, OU, SAUF**
  - opérateurs de comparaison: =, <, > (ex: pour l'année)
  - troncature: impose la recherche seulement sur une portion de mot (ex: *évolu\** pour *évolution, évolutif(ve), évoluer...*)
  - phrase: permet d'interroger sur des expressions: (ex: «*sélection naturelle*»)



De nombreuses notices sont stockées dans les bases de données bibliographiques.

On retrouvera les ouvrages intéressant notre recherche en formulant des **requêtes** pour interroger la base.



## La recherche documentaire les principes de l'interrogation

- le principe des **moteurs de recherche** (Alta Vista): donne une liste de sites où a été repérée la plus grande **fréquence** d'apparition de *ce mot* dans une collection de textes
  - interrogation sur un langage naturel
- le principe des moteurs d'indexation ou **annuaires thématiques** (Yahoo!): donne une liste de sites référencés par des **documentalistes** sur *ce sujet*
  - interrogation sur un langage documentaire
- les **bases de données bibliographiques** proposent les deux principes d'interrogation:
  - langage documentaire: champs MOTS-CLES, SUJETS
  - langage naturel: champs RÉSUMÉ, TITRE
  - on peut imposer le champ de recherche



Les **moteurs de recherche** font appel à des **robots** qui vont opérer une comparaison et rechercher la **similitude** entre le ou les mots (CQFD une **chaîne de caractères**) souhaités et une collection de textes (Alta Vista).

Les **moteurs d'indexation** renvoient à la liste correspondant au terme d'indexation de **même sens** ou de sens voisin dans le **thésaurus** (Yahoo).

## Exemple de recherche de documentation: 1- Alta Vista

requête	Résultats	sur 10 items : rés. pertinents	Résultats non pertinents
leopard	3 365	1	1 chien, 1 cirque, 1 voyage, 1 prix, 1 véhicule, 1 militaire, 2 littérature), 1 divinité
panthere	2 863	0	1 zoo, 1 cirque, 1 photos, 1 foot, 1 math, 1 M. Ayme, 1 champignon, 2 déco, 1 Afrique
leopard OR panthere	8 602	0	2 p perso, 1 équipe, 1 tarzan, 1 déco, 1 blagues, 1 coloriages, 1 icônes, 1 Afrique, 1 ville
felins ET afrique	225	1	1 tarzan, 1 safari, 1 antiraciste, 1 fourrures, 1 film, 1 art, 1 Afrique, 2 bâtiment
« panthera pardus »	37	4	2 safari, 1 paleo, 1 panthere rose, 1 décret, 1 zoo
feli* ET panther* ET leopard	53	5	1 cirque, 1 comte, 1 équipe, 1 film, 1 vétérinaire

18

Nous recherchons de la documentation sur le comportement de chasse du Léopard.

Nous effectuerons une première recherche de documents sur l'Internet à l'aide d'Alta Vista et une seconde sur la base de données bibliographique « Zoological Record ».

Nous savons *a priori* que la documentation que nous trouverons à partir des notices de Zoological Record sera scientifique et spécialisée, tandis que la documentation Web sera plus générale et peut-être moins sûre. On considère alors sur le net comme « résultat pertinent » un document traitant du mode de vie du léopard. Et, sur Zoological Record, on considère « résultat pertinent » un document traitant du comportement de chasse du léopard.

## Exemple de recherche de documentation: 2- Zoological Record

requête	Résultats	sur 10 items : rés. pertinents	Résultats non pertinents
leopard	163	2	1 (maladie feline), 1 poisson, 2 reptile, 1 amphibien, 2 "leopard seals", 1 fourrure
leopard OU panthere	591		
Leopard ET felidae ET panthera	62	5	1 fourrure, 1 nourriture, 3 insémination artificielle
« panthera pardus »	84		
« panthera pardus » ET behaviour	4	3	1



19

Nous allons ainsi tester l'efficacité d'une part de nos mots-clés, et d'autre part l'efficacité de les combiner à l'aide d'opérateurs logiques. Il s'agit également **d'adapter notre stratégie de recherche au type d'interrogation et au fonds documentaire.**

L'apparition de résultats non-pertinents (bien que vraisemblable, ex: léopard des mers) est appelé le « **bruit** ». A l'inverse, le « **silence** » qualifie la réponse pertinente non retrouvée. Dans le premier cas, on **affinera** la recherche en utilisant des termes plus spécifiques ou des opérateurs restrictifs. Dans le second on **élargira** le champ de recherche avec des termes plus généraux.

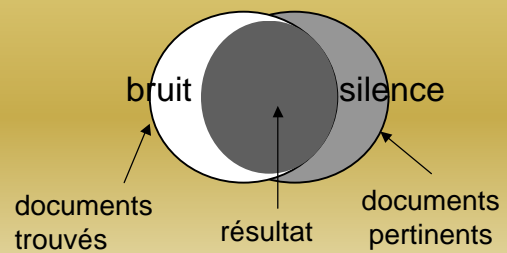
## Les langages documentaires

### Les apports

- La construction et l'utilisation du thésaurus
  - éclaire les idées et la problématique
  - apporte une logique relationnelle entre les concepts considérés
- Selon les résultats obtenus:
  - la recherche sur le terme générique élargit le champ de résultats
  - la recherche sur le terme spécifique ou sur une combinaison de termes précise le champ de résultats



- l'interrogation sur le langage documentaire permet de diminuer le bruit et le silence et augmente la pertinence des résultats



## Les langages documentaires

### Les limites

- comparé au langage naturel:
  - pose le problème de la rigidité du langage: il est peu évolutif, l'enrichissement en est retardé car il existe un intermédiaire: le travail des documentalistes
  - perte d'information:
    - ❖ par la diminution de la variabilité du langage
    - ❖ par la perte des nuances de sens
- Ni les langages documentaires (thésaurus, notices...), ni les langages d'interrogation (troncature, booléens...) ne sont proposés de manière standard: il n'existe pas de véritable **code** universel !  
=> on doit à chaque fois consulter le mode d'emploi.



## La collecte et la préparation de l'information

- collecter l'information:
  - la sélection des ouvrages constitue une pré-lecture
  - on emploie différents niveaux de lecture:
    - ❖ lecture complète
    - ❖ lecture du résumé et/ou de la table des matières...
- condenser et présenter l'information:
  - l'extraction des concepts, idées, et informations
    - ❖ réduction d'un texte par sa synthèse (condensation)
    - ❖ collecte de citations, d'éléments d'information, de données, d'illustrations etc.



## L 'organisation de l'information

- indexer et référencer l'information:
  - indexer pour repérer et organiser sa documentation
  - relever les **références** afin de **retrouver**

Suivant le modèle AUTEUR. Date. *Titre*. Source.

- Références bibliographiques « classiques » :
  - ❖ AUTEUR. Date. *Titre*. Source (Édition ou version. Titre de périodique et/ou éditeur. Collection. Numéro normalisé. Lieu de publication.
- Références bibliographiques électroniques :
  - ❖ AUTEUR. Date. *Titre*. [type de support (CDRom, disquette etc)]. Éditeur, lieu de publication.
  - ❖ AUTEUR. Date de mise à jour ou de révision (jj/mm/année). *Titre*. Éditeur. Available online at <http://...>



## Conclusion sur la recherche documentaire: une démarche, une méthode, des outils

- Quel sujet traité ?
  - élaborer et enrichir son thésaurus personnel tout au long du travail
- Quel type d'information ?
  - « générale » ?
  - « scientifique » ?
- Quel type de réalisation ?
  - narration
  - monographie
  - théorie, traité
  - compte-rendu ou rapport
  - exposé etc.

inventaire

sélection

analyse

- s'aider d'un dictionnaire ou d'un thésaurus déjà existant dans le domaine considéré
- en bibliothèque ou sur l'Internet:
  - ouvrages généraux (livres, CD-ROM..)
- en bibliothèque spécialisée:
  - publications scientifiques
- source de l'information:
  - compiler, comparer, analyser la bibliographie
  - interpréter données, résultats



Tout au long de ce travail, nous avons suivi la même démarche et nous sommes posés ces mêmes questions :

Qu'est-ce qu'on a à disposition? (inventaire) ; Qu'est-ce qu'on prend? (sélection) ; Qu'est-ce qu'on en retire? (analyse).

Répondre à ces trois questions nous aidera certainement pour répondre *in fine* à : Qu'est-ce qu'on cherche ?



## Thésaurus et dictionnaires

- Thésaurus
  - MeSH Terms (Medline Subject Heading).  
<http://www4.ncbi.nlm.nih.gov/PubMed/>
  - Life Science Thesaurus  
<http://www.csa.com/helpV3/lscthes.html>
- Dictionnaires en biologie «*Terminology Collection*»:  
<http://www.uwasa.fi/comm/termino/collect/special/biology.html>



## Références

- **Chaumier, J.** (1988). Le traitement linguistique de l'information. 3e éd. mise à jour et augmentée. Paris: Entreprise Moderne d'Édition.
- **Filippi D.** Comment présenter des références bibliographiques. Université paris 8.  
<http://www-bu.univ-paris8.fr/Formation/RefBiblio.html>
- **Gheereart M-A.** (2000). Les thésaurus. UPMC.  
<http://haydn.cicrp.jussieu.fr/documentation/thesaurus/>
- **Gosselin, C.** (1995). L'information et le travail de recherche. Éducatechnologiques. Vol 2, n° 1. Université Laval, Québec, Canada.  
[http://www.fse.ulaval.ca/fac/ten/reveduc/html/vol2/no1/educ\\_tdm.html](http://www.fse.ulaval.ca/fac/ten/reveduc/html/vol2/no1/educ_tdm.html)
- **MEDIADIX.** Initiation à la recherche documentaire sur l'Internet et à la récupération de données. Université Paris X.  
<http://www.u-paris10.fr/mediadix/doc/inetrd/>

