

## INDEXER, INTERROGER, MEME COMBAT...

### Sommaire

L'utilisation systématique de l'internet comme outil de recherche a, pendant un temps, laissé espérer que la recherche documentaire allait pouvoir se dégager de toute contrainte grâce au langage naturel et au plein texte. Si ces approches présentent un intérêt non négligeable, elles ont néanmoins leurs limites et on a dû convenir que l'indexation et les thésaurus restaient des valeurs sûres \*.

En effet, une recherche documentaire requiert un certain nombre de règles et de compétences sans lesquelles le hasard serait seul maître du jeu.

Cette démarche rigoureuse qui procède par phases, par croisements, par retours en arrière, donne à **l'interrogation** tout son sens.

Elle suppose aussi qu'un travail de préparation et de mise à disposition du document ait été au préalable effectué afin de donner toute sa pertinence à la chaîne documentaire au cours de laquelle **l'indexation** sera déterminante.

**Indexation et interrogation constituent les deux étapes complémentaires d'une même activité documentaire** : si un document n'est pas correctement indexé, il ne pourra pas être retrouvé ; si l'interrogation ne pose pas les bonnes questions, le résultat de la recherche sera décevant.

C'est pourquoi une bonne indexation aura pris en compte, par anticipation, les mécanismes de l'interrogation à venir : **l'indexeur doit se mettre à la place de l'interrogateur**.

De son côté, une bonne interrogation est fondée sur le capital confiance accordé plus ou moins consciemment au travail d'indexation.

Dans les deux cas, des stratégies et des outils ou des notions communs seront à mettre en oeuvre.

A la demande de certains collègues nouvellement arrivés dans la fonction, nous avons tenté ici de faire une présentation sommaire et non exhaustive de ces outils et notions.

En les inscrivant dans une démarche séquentielle, nous voulions suggérer le déroulement d'un processus dont l'organisation est laissée à l'appréciation de chacun.

\* Elargissement critique de la question : <http://www.savoirscdi.cndp.fr/culturepro/actualisation/bibliotheco/LangageFeyler.htm>

Un bon indexeur est aussi un bon interrogateur.

## PREALABLES A L'INDEXATION ET A L'INTERROGATION

Repères utiles, mais non exhaustifs :

- ◆ [le thésaurus](#)
- ◆ [l'indexation](#)
- ◆ [l'interrogation](#)
  - [le questionnement – la mobilisation des idées](#)
  - [les opérateurs booléens](#)
  - [l'autopostage](#)
  - [la troncature](#)
  - [des exemples](#)

Effectuer une recherche, en l'occurrence documentaire, c'est **se** poser des questions pour ensuite poser **une** question, c'est-à-dire **interroger**.

**Indexer**, c'est aussi se poser des questions, en se mettant à la place de la personne qui va interroger, ce qui suppose [l'analyse](#) des documents à traiter.

Dans les deux cas, le **thésaurus** sera un outil indispensable.

Pour permettre l'interrogation et l'indexation, un logiciel documentaire doit intégrer un thésaurus.

## LE THESAURUS

Qu'est-ce qu'un thésaurus ?

"Un thésaurus est **une liste structurée de concepts** destinés à représenter de manière **univoque** le **contenu** de documents et de questions dans un système documentaire déterminé et à assister l'utilisateur dans l'indexation des documents et des questions ; les concepts sont extraits d'une **liste finie**, établie *a priori* ; seuls les termes figurant dans cette liste peuvent être utilisés pour indexer les documents et les questions ; l'assistance à l'utilisateur est apportée par la structure sémantique du thésaurus : relations d'équivalence, de hiérarchie et d'association essentiellement." (Van Slype – 1987)

"*Langage documentaire* fondé sur une structuration hiérarchisée d'un ou plusieurs domaines de la connaissance et dans lesquels les notions sont représentées par des *termes* d'une ou plusieurs langues naturelles et les relations et les notions par des signes conventionnels". (AFNOR – 1989)

L'Education nationale utilise le thésaurus Motbis qui est un produit du CNDP.

Les recommandations essentielles sur son utilisation figurent dans les pages introductives ou dans le *lisez.moi* du produit.

<http://www.cndp.fr/lesScripts/bandeau/bandeau.asp?bas=http://www.cndp.fr/motbis/accueil.htm>

# L'INDEXATION

L'indexation est **l'action** d'indexer **ou le résultat** de cette opération.

Indexer, c'est représenter (décrire) par les éléments d'un langage documentaire – ou naturel – des données résultant de **l'analyse du contenu** d'un document ou d'une question.

## Analyser un document

1/ C'est se poser un certain nombre de [questions](#) sur le contenu et la forme de ce document.

L'indexation ne portera que **sur le contenu**.

**Quoi ?** de quoi s'agit-il ? (le sujet général)

**Quand ?** à quelle époque ?

**Où ?** dans quel lieu ?

**De quel point de vue ?** (essentiel)

Pour affiner l'indexation, l'utilisation de [l'autopostage](#) peut être d'un grand intérêt.

2/ s'appuyer sur un certain nombre de **repères** :

- le titre et le sous-titre (s'ils sont pertinents)
- le sommaire et la table des matières
- la 4<sup>e</sup> de couverture
- l'introduction et la conclusion

NB : proscrire toute lecture subjective ou réductrice

3/ dégager des **mots-clés** correspondant aux concepts (langage naturel) : cf. ci-contre

4/ les traduire en **descripteurs** (langage documentaire)

5/ **anticiper** la recherche ; **se mettre à la place de l'utilisateur**.

## Du mot-clé au descripteur

### Quoi ?

Par exemple, pour indexer un document traitant de l'astrophysique, le mot-clé qui vient naturellement à l'esprit est *astrophysique* ; ce dernier n'est pourtant pas retenu par Motbis qui propose le descripteur *astrophysique : science*.

### Quand ?

Par exemple, pour indexer un document traitant de l'histoire du cinéma, le mot-clé qui vient naturellement à l'esprit pour désigner la période est *histoire* ; alors que le descripteur requis est *20<sup>e</sup> siècle* et/ou *21<sup>e</sup> siècle*.

### De quel point de vue ? - essentiel

La mort en Egypte peut être certes considérée sous l'angle biologique, mais elle le sera bien davantage par rapport aux funérailles et ce sera le descripteur *rite funéraire* qui sera retenu.

D'autres questions sur **la nature et la forme du document** peuvent se poser :

**Pour qui ?** (destinataire)

**Qui ?**

**Comment ?**

**Sous quelle forme ?**

etc.

(cf. [le questionnement et la mobilisation des idées](#))

## Risques à éviter

- le **bruit** documentaire (après l'interrogation, des documents sélectionnés ne sont pas pertinents et "encombrent" la recherche),
- le **silence** documentaire (après l'interrogation, des documents pertinents existant dans la base n'apparaissent pas).

## Précautions à prendre

(entre autres)

- respecter la **spécificité verticale** : indexation au même niveau que le concept retenu  
[Par exemple, si le document porte sur la Grèce, il faudra indexer avec Grèce et non avec Europe qui est un terme générique, donc placé "au-dessus".]
- respecter la **spécificité horizontale** (descripteurs précoordonnés) : utiliser un descripteur composé s'il existe dans le thésaurus, et seulement dans ce cas  
[Par exemple, si le document traite de la pollution des mers, indexer avec *pollution des mers* et non avec *pollution et mer*.]
- indexer **au plus spécifique**, au plus précis : [l'autopostage](#) permet ensuite d'interroger sur le terme générique, mais aussi sur son ou ses termes spécifiques.

**Indexer, c'est se mettre à la place de l'utilisateur ;  
c'est donc aussi interroger (par anticipation).**

# L'INTERROGATION

Dans l'idéal, afin d'effectuer une bonne interrogation, la connaissance de quelques notions techniques de base peut faciliter les choses :

- [base de données](#)
- [opérateurs booléens](#)
- [autopostage](#)

## LA BASE DE DONNEES

Une base de données est **un ensemble d'informations sur un domaine particulier**. On dira, par exemple, que l'ensemble des notices renvoyant aux documents présents dans le CDI constitue la base documentaire du CDI.

Cette base présente **une structure** : cette dernière est organisée en un certain nombre de champs ou de rubriques. Chaque rubrique renseigne sur un élément ou un aspect du document.

La base du CDI pourrait avoir la forme suivante :

1. type de support
2. nature du support
3. ISBN ou ISSN
4. article du titre
5. titre propre
6. ...
7. ...
8. Motbis
9. indexation libre
10. date de saisie
11. origine de la donnée
12. etc.

Sa transposition à l'écran laissera apparaître les mêmes champs que l'on pourra renseigner ou sur lesquels on pourra interroger.

On peut interroger :

- sur un champ et sur un seul terme
- sur un champ et sur plusieurs termes
- sur plusieurs champs

cf. [exemples d'interrogation](#)

Dans les deux dernières hypothèses, il faut utiliser [les opérateurs booléens](#).

En outre, comme pour l'indexation, il est indispensable de traduire un sujet en mots-clés, puis en descripteurs et de le délimiter. [La mobilisation des idées et le questionnement](#) sont des outils très utiles. On peut alors formuler une requête, poser une question.

## LE QUESTIONNEMENT

### Nous aussi, nous nous interrogeons...

Le questionnement n'est qu'une étape, préalable à l'indexation et à la consultation de documents. Il concerne la documentation, mais aussi l'ensemble des disciplines, à travers, par exemple, la confection d'une rédaction ou d'une dissertation en français ou en économie, la lecture d'une image ou la préparation d'un exposé.

Les documents présentés ici sont le résultat d'activités pratiques réalisées avec des élèves. Ils ne sont en rien une fin en eux-mêmes. Ils ne représentent pas non plus des réponses immédiates ou des solutions toutes faites. Ils n'ont en aucun cas de valeur modélisante. Ils constituent seulement des pistes de réflexion dans la mise en place d'autres outils adaptés à un contexte particulier.

Le questionnement « standard » à partir des 3QOCP est certes déjà intéressant. Néanmoins, une mobilisation des idées enrichie par l'apport d'une réflexion personnelle débouchera sur des formulations d'hypothèses générant elles-mêmes leur propre critique (l'erreur aussi est constructive).

Cette démarche n'est pas sans risques. Il « suffira » de la confronter à la relativité des savoirs, à la notion de points de vue, sans s'interdire d'autres ouvertures. On retrouve par là même la nécessité de « cerner un sujet », d'éviter le hors-sujet, etc.

Une simplification excessive de la démarche trahirait sa finalité et interdirait l'accès à une seconde phase d'apprentissage.

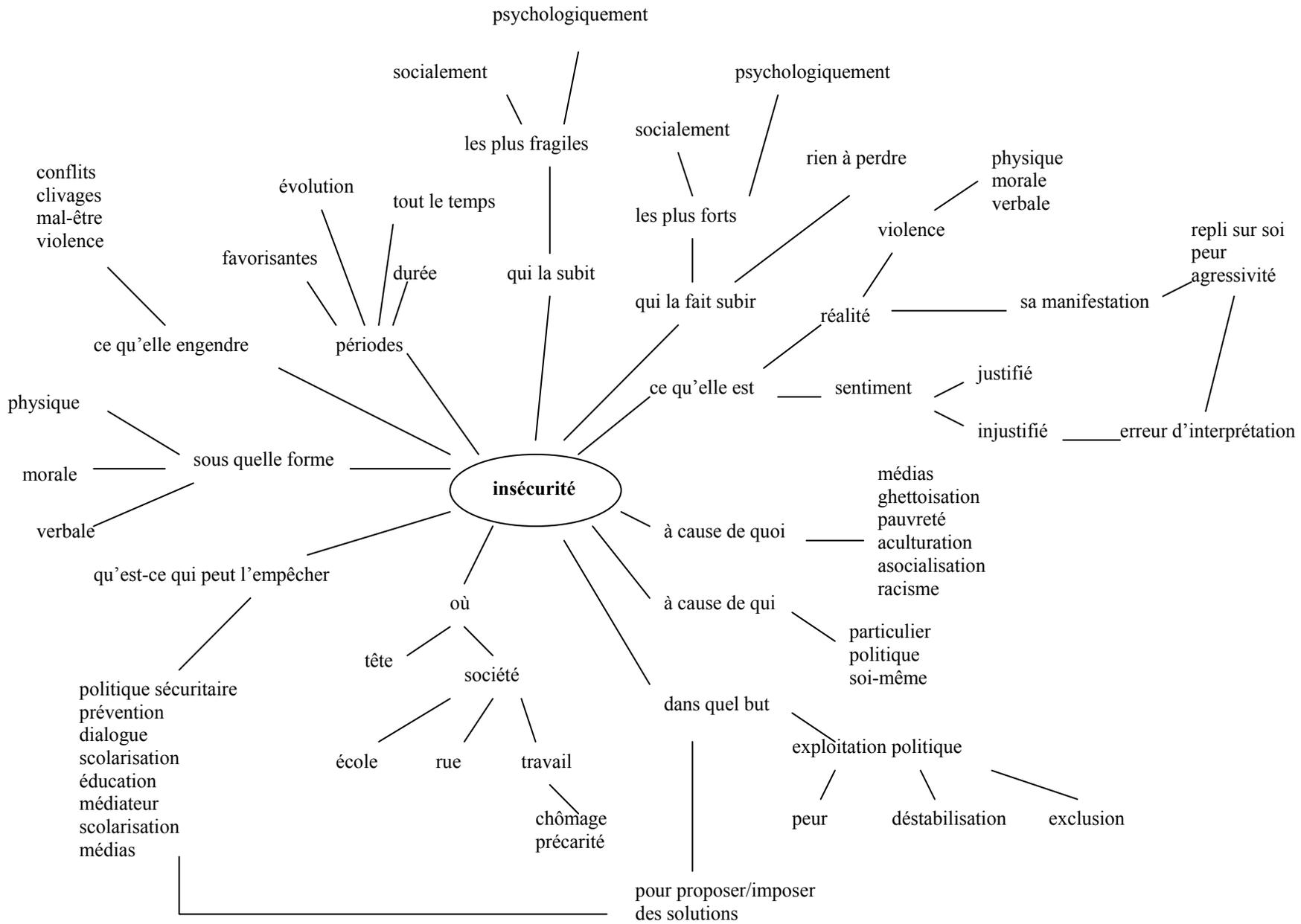
Les exemples suivants ont pour objectif :

- d'illustrer de manière séquentielle l'évolution possible d'un même thème, celui de l'insécurité, qui relève de l'enseignement de l'ECJS , avec :
  - un questionnement conceptuel
  - un questionnement cumulant diverses formes d'interrogation
  - un exemple d'application en classe où l'on voit que certains arguments peuvent valoir dans des domaines opposés, voire contradictoires
- de présenter un thème au moyen de verbes
- de montrer l'utilisation possible du questionnement dans les disciplines

**Le caractère réducteur du travail présenté ici peut se comprendre aisément : il n'a de sens que par rapport au contexte particulier dans lequel il s'inscrit et au discours qui l'accompagne dans la classe.**

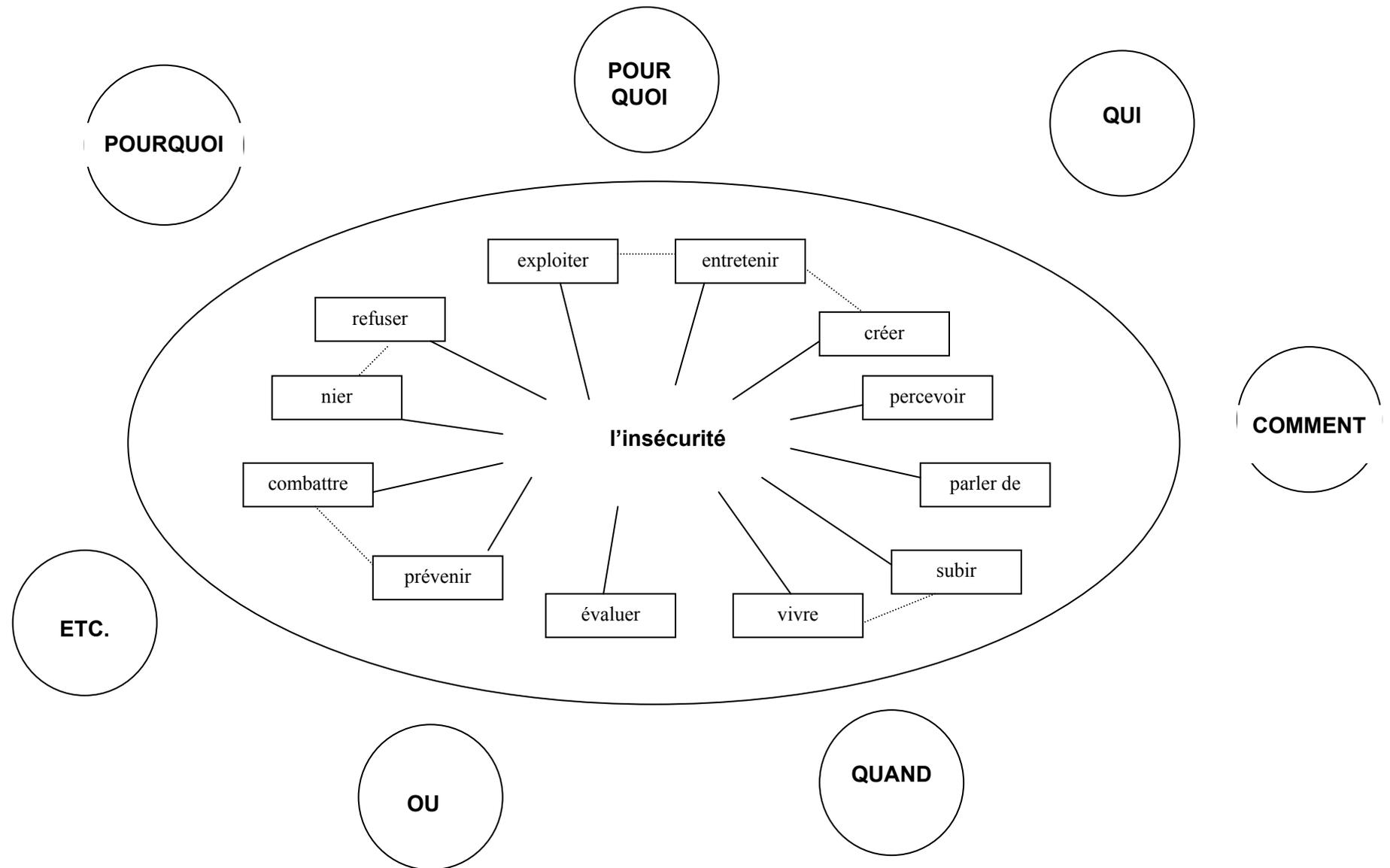
Dans tous les cas, on confirmera l'importance du questionnement dans le cadre de la recherche documentaire et des liens étroits que le CDI et le professeur-documentaliste entretiennent avec les disciplines et les enseignants de discipline.

**QUESTIONNEMENT**

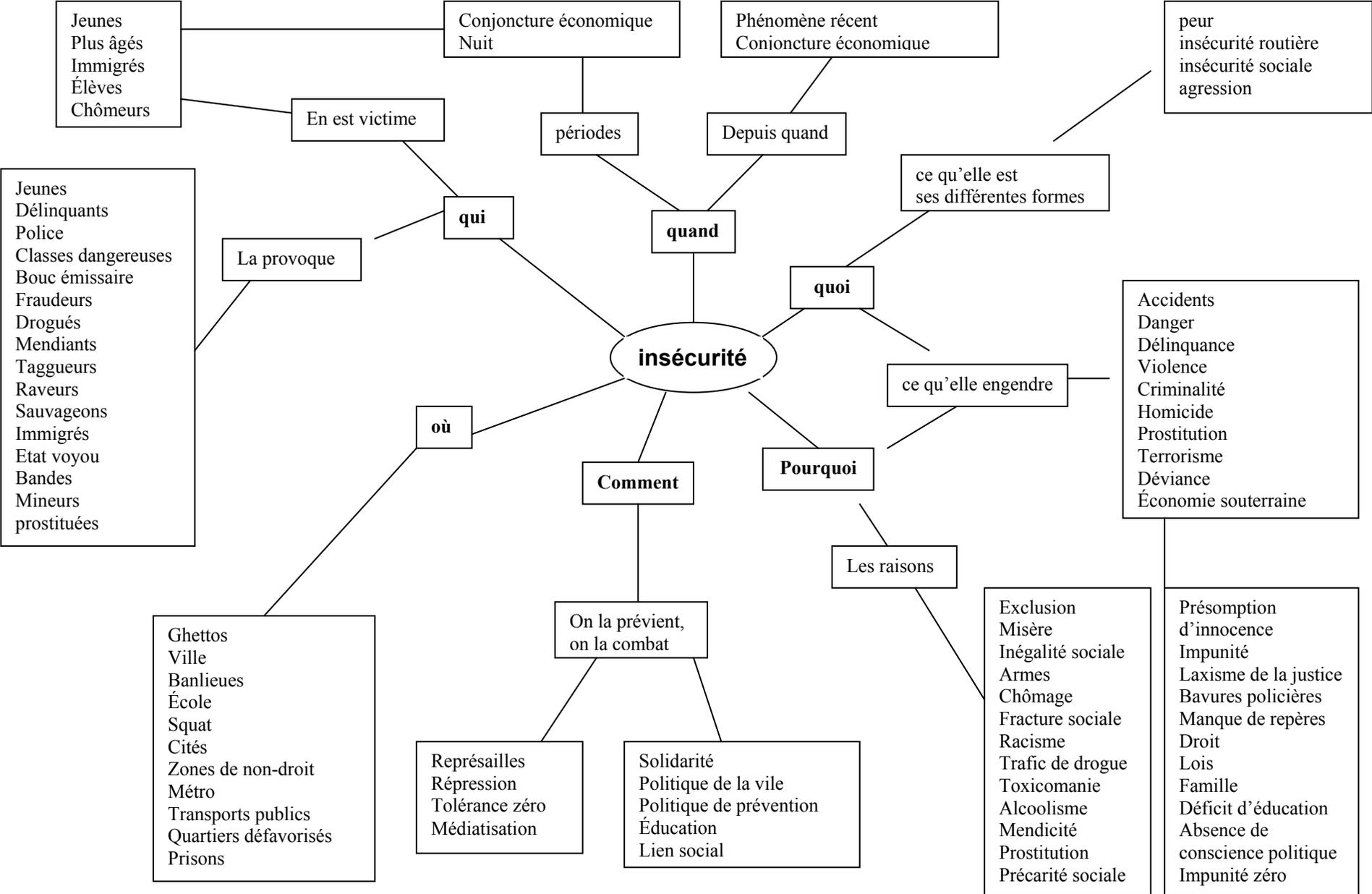


Afin d'élargir le champ des possibles et de multiplier les points de vue, il est aussi souvent essentiel de recourir à l'utilisation de **verbes**, soit comme première entrée, soit comme relais aux questions, sans oublier l'établissement de liens entre les différents éléments.

Si l'on reste sur le thème de l'insécurité :



**QUESTIONNEMENT**  
**esquissé avec une classe de 1<sup>re</sup> ES**  
**par Clotilde CHAUVIN**

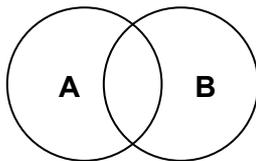


## LES OPERATEURS BOOLEENS

Ces opérateurs doivent leur qualificatif au mathématicien George BOOLE :  
ET, OU, SAUF.

Ils permettent d'établir des relations entre des ensembles.

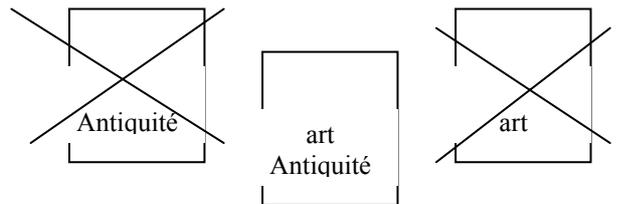
1/- **ET** permet de sélectionner des éléments communs à plusieurs ensembles ; l'opération réalisée est une **intersection**.



A **ET** B

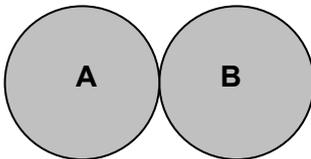
La partie centrale correspond  
aux éléments communs à A et B

Soit, par exemple, les ensembles suivants :  
A est l'ensemble des notices contenant dans  
la rubrique du thésaurus Motbis le terme art  
B est l'ensemble des notices contenant dans  
la rubrique du thésaurus Motbis le terme  
Antiquité.  
Je recherche des notices  
correspondant au thème de *l'art  
antique*.



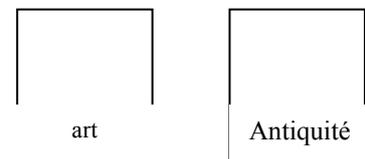
Si l'on formule la question suivante :  
*art ET Antiquité*,  
seront sélectionnées les seules  
notices contenant **à la fois**  
ces deux descripteurs.

2/- **OU** permet de sélectionner la somme des  
éléments se trouvant séparément dans plusieurs  
ensembles ; l'opération réalisée est une **union**.



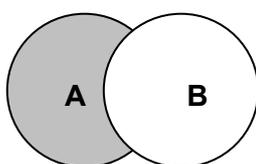
A **OU** B

La partie grisée correspond  
à la somme des éléments  
appartenant soit à A soit à B



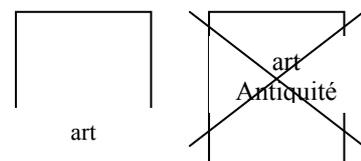
Si l'on formule la question suivante :  
*art OU Antiquité*,  
seront sélectionnées les notices  
contenant l'un **ou** l'autre de ces  
descripteurs (ou les deux, bien sûr)

3/- **SAUF** permet de sélectionner des éléments  
d'un ou de plusieurs ensembles, en excluant les  
éléments se trouvant dans un ou d'autres  
ensembles ; l'opération réalisée est une  
**exclusion**.



A **SAUF** B

Seule la partie grisée  
répond à la question



Si l'on formule la question suivante :  
*art SAUF Antiquité*,  
seront sélectionnées les seules  
notices contenant *art*,  
**mais non Antiquité**.

## L'AUTOPOSTAGE

Dès qu'une base comprend un nombre de notices relativement important, cette fonction est essentielle pour obtenir des réponses précises.

On devra toujours la prendre en compte au moment de l'indexation.

### Principes généraux

En se positionnant sur un terme, l'autopostage permet d'interroger sur les différents niveaux de ses spécifiques

#### Position initiale d'interrogation

##### biologie cellulaire

- ↪ cellule
  - cytoplasme
  - membrane plasmique
  - noyau cellulaire
    - chromosome
    - chromosome sexuel
    - gène : biologie
- division cellulaire
- méiose
- mitose
- échange cellulaire

#### Résultat de l'interrogation

##### Si niveau 1 demandé

- cellule
  - cytoplasme
  - membrane plasmique
  - noyau cellulaire

##### biologie cellulaire

- cellule
  - ↪ cytoplasme
  - membrane plasmique
  - noyau cellulaire
    - chromosome
    - chromosome sexuel
    - gène : biologie
- division cellulaire
- méiose
- mitose
- échange cellulaire

##### Si niveau 2 demandé

- cytoplasme
- chromosome
- chromosome sexuel
- gène : biologie

##### ↪ biologie cellulaire

- cellule
  - cytoplasme
  - membrane plasmique
  - noyau cellulaire
    - chromosome
    - chromosome sexuel
    - gène : biologie
- division cellulaire
- méiose
- mitose
- échange cellulaire

##### Si niveau 3 demandé

- ↪ biologie cellulaire
  - cellule
    - cytoplasme
    - membrane plasmique
    - noyau cellulaire
      - chromosome
  - division cellulaire
  - méiose
  - mitose
  - échange cellulaire

Selon le même principe, certains logiciels proposent en outre un autopostage ascendant, voire un autopostage étendu (étendu aux termes associés).

Cette présentation générale est à adapter au logiciel utilisé.

Exemple d'autopostage sur BCDI

<http://www.ac-amiens.fr/academie/pedagogie/Documentalistes/bcdi/bcdi.htm#Recherche>

## LA TRONCATURE

Procédé permettant la suppression ou l'omission d'une chaîne de caractères dont on veut ignorer la valeur (AFNOR 1987).

Si on tape cell\* (troncature à droite), on obtiendra des réponses sur **cellulaire**, **cellule**, **cellulite**, **celluloïd**, **cellulose**, etc.

Si on tape \*ite (troncature à gauche), on obtiendra des réponses sur **cellulite**, **moite**, **suite**, **fuite**, etc.

(On peut aussi parfois opérer une troncature à l'intérieur d'un mot ; si on tape : \*cour\*, on obtiendra des réponses sur **recours**, **secourisme**, etc.

Les possibilités varient d'un logiciel à l'autre.

## EXEMPLES D'INTERROGATION A PARTIR DE TROIS DESCRIPTEURS

(voir aussi [opérateurs booléens](#))

- **Soit une recherche** à partir du sujet suivant :

*Des réserves naturelles pour protéger les animaux d'Afrique*

Descripteurs retenus : *réserve naturelle, protection des animaux, Afrique.*

Le contrôle après thésaurus étant effectué sur chacun des termes, on aura à poser la question suivante :

*réserve naturelle ET protection des animaux ET Afrique*

Les notices sélectionnées comprendront ces trois descripteurs.

- **Soit la recherche suivante** à effectuer :

*Tous les documents de Pierre Miquel ou de Jacqueline de Romilly traitant de l'Antiquité.*

Descripteurs ou mots-clés retenus : *Miquel, Pierre ; Romilly, Jacqueline de ; Antiquité.*

Le contrôle par thésaurus étant fait sur *Antiquité*, on aura à poser la question suivante :

*(Auteurs = Miquel, Pierre OU Auteurs = Romilly, Jacqueline de) ET Antiquité*

[on pourra aussi utiliser la troncature sur les noms d'auteurs et l'[autopostage](#) sur le descripteur Motbis]

Les notices sélectionnées comprendront :

soit (OU) *Miquel, Pierre ET Antiquité*,

soit (OU) *Romilly, Jacqueline ET Antiquité*

[Si l'on avait omis les parenthèses, on aurait formulé la question suivante :

*Auteurs = Miquel, Pierre OU Auteurs = Romilly, Jacqueline de ET Antiquité,*

ce qui aurait entraîné deux opérations, pour deux recherches distinctes (sachant que l'opérateur **ET** – comme le **SAUF** – est prioritaire sur le **OU**) :

– une recherche sur Auteurs = Romilly, Jacqueline de **ET** Antiquité

– une recherche sur : Auteurs = Miquel, Pierre,

ces deux éléments étant reliées par un **OU**.]

Les notices sélectionnées auraient donc comporté :

soit *Romilly, Jacqueline de ET Antiquité*,

soit *Miquel, Pierre* (ce qui, tel quel, ne répondait pas à la question initiale).]