

# LANGAGES DOCUMENTAIRES

## RAPPEL ET APPROFONDISSEMENT FICHE « ANALYSE DOCUMENTAIRE »

- L'analyse documentaire se trouve au cœur du système documentaire.  
→ elle permet d'identifier le sujet et les données contenus par un document ET permet de produire un résumé et/ou une indexation.
- L'indexation est une opération consistant à décrire et à caractériser un document  
→ description des concepts, puis leur transcription en langage documentaire.
- L'indexation permet d'augmenter la pertinence, la cohérence et l'objectivité d'un document dans une recherche par des indexations de type :
  1. *analytique* = basé sur des langage documentaire,
  2. *signalétique* = les notions sont symbolisées par des domaines disciplinaires (ex. : cote),
  3. *automatique* = recherche d'occurrence par des machines.

## Les langages documentaires

### Définition (AFNOR) :

- Les langages documentaires sont des langages artificiels constitués de représentations de notions et de relations entre ces notions, destinés, dans un système documentaire, à formaliser des données contenues dans les documents et les demandes des utilisateurs.

### Utilité des langages documentaires :

- Ils mettent de faire face à des problèmes d'ordre :
  1. *Linguistique* : la synonymie et la polysémie, issues du langage naturel, provoquent des ambiguïtés et des risques de silence (absence de résultats).  
**Ex. :** « Calcul » renvoie aussi bien aux mathématiques qu'à la pathologie rénale.
  2. *Documentaire* : chaque individu peut analyser documentairement un document à sa façon.  
→ Pb de l'hétérogénéité des analyses et des interprétations, ainsi que des références culturelles.

→ L'utilisation d'un langage normalisé permet de rendre l'analyse documentaire objective et universellement compréhensible.

### Typologie des langages documentaires :

- Langages classificatoires = associés à des codes (Bacon, Dewey, Otlet, Ranganathan)  
→ Aussi appelés langage à structure hiérarchique
- Langages combinatoires = association de plusieurs descripteurs (langage naturel normalisé)  
→ peuvent être pré-coordonné (la combinaison des termes est déjà établie a priori de l'indexation),  
→ peuvent être post-coordonné (la combinaison des termes est établie au moment de la recherche documentaire, c'est-à-dire a posteriori) = usage des booléen « ET ».

## Débat autour des langages documentaires

- Les langages documentaires sont des outils garants de l'accès à l'information.  
→ Ils organisent la connaissance et sont des médiations en évolution.
- L'outil informatique est venu aider l'indexation documentaire, voire il y est intégré.
- Les outils du web ont transformé la recherche d'information (leur mode d'accès) : recherche en plein texte.
- Plusieurs outils se sont développés autour du TAL (traitement automatique du langage) pour proposer une recherche en langage naturel.  
→ C'est une interprétation basée sur la phrase construite plutôt que sur un mot-clé.  
**Ex. :** théorie physique quantique Vs. « quelles sont les théories en physique quantique ? »
- L'indexation automatique vient aider l'indexation humaine depuis l'explosion numérique.  
→ Trop d'informations à indexer en peu de temps.
- L'engouement pour l'indexation automatique a conduit à prédire la fin de l'indexation classique et des langages documentaires au profit du traitement automatique de l'information.
- Les langages documentaires continuent à connaître une évolution avec un filigrane une problématique récurrente sur leur utilisation.