

**Université Farhat Abbas Sétif**  
**Faculté des sciences**  
**Département d'Informatique**  
**Module : Ingénierie des connaissances**  
**Master 2 IFIA**

### **Objectifs de l'enseignement :**

Étude de la modélisation des connaissances dans ses dimensions acquisition, représentation et capitalisation. Présentation des voies de recherche et de l'intérêt actuel de l'industrie pour la modélisation des connaissances appliquée à la mémoire d'entreprise et à la capitalisation des connaissances. Montrer la complémentarité des approches ingénierie des connaissances et gestion des connaissances.

### **Contenu de la matière :**

**Chapitre 1 :** Ingénierie et Modélisation des connaissances.

**Chapitre 2 :** Points de vue dans les systèmes à objets et les graphes conceptuels.

**Chapitre 3 :** Extraction des connaissances à partir de données KDD ;

**Chapitre 4 :** Fouille de données et apprentissage automatique.

**Chapitre 5 :** L'approche CommonKADS,

**Chapitre 6 :** Gestion et Capitalisation des connaissances : Différents aspects de la mémoire d'entreprise: mémoire technique, mémoire de projet, mémoire métier. Retour d'expérience: REX, MEREX, Méthodes MKSM, SAGACE.

**Chapitre 7 :** Raisonnement à partir de cas, outils RàPC et exemples d'applications.

**Chapitre 8 :** Logique de description: LOOM.

### **Référence :**

- Vous trouvez la liste des références adéquates en fin de chaque chapitre (*bibliographie*).

## **Chapitre I: Ingénierie et modélisation des connaissances**

### **I.1. Introduction**

L'ingénierie des connaissances (IC) fournit une démarche d'analyse et de modélisation d'une résolution de problèmes [Charlet 2003]. Les travaux dans cette discipline fournissent des guides méthodologiques et de représentation de résolution de problèmes, [Aussenac et al. 1996]. Ces travaux puisent leur source dans des théories et des méthodes empruntées à diverses disciplines qui étudient l'activité rationnelle, comme la psychologie cognitive, l'ergonomie, la linguistique, la sociologie et l'intelligence artificielle. L'objectif de l'IC est de représenter le rôle que jouent les données et les informations dans une résolution de problèmes. Ces rôles sont appelés connaissance. En d'autres termes, on appelle connaissance le savoir et le savoir-faire utilisés par une personne ou un groupe de personnes pour réaliser une action dans un contexte donné.

### **I.2. La démarche de l'Ingénierie des Connaissances**

L'ingénierie des connaissances consiste à mettre en place les procédures permettant de sélectionner les informations les plus porteuses de sens, de les organiser autour de la sémantique qui leur a été associée a priori et d'associer des mécanismes d'interprétation automatique aux agents interprétant pour exploiter le système ainsi développé.

➤ **Définition :**

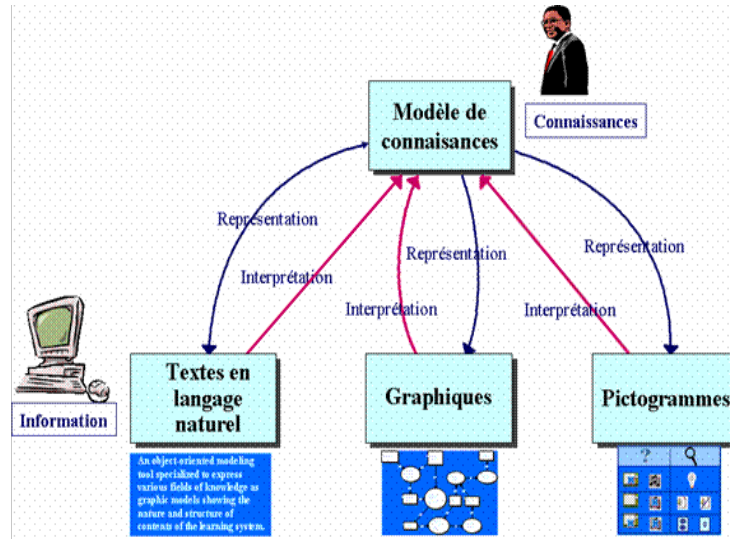
Une connaissance est la capacité d'exercer une action pour atteindre un but [Bach2004].

➤ **Qu'est ce que la RDC?**

Représenter des connaissances propres à un domaine particulier consiste à décrire et coder les entités de ce domaine sous une forme qui puisse être exploitée en machine [KAY 97].

➤ **Représenter une entité :**

Les connaissances : concepts, propriétés, relations, prédicats, hiérarchies, réseaux sémantiques, schémas, règles, méthodes heuristiques. La figure I.1 illustre la représentation et l'interprétation des informations par une personne. L'interprétation consiste à transformer les informations en connaissances internes dans sa mémoire, ce qu'on peut aussi appeler « apprentissage ». La représentation des connaissances est le processus inverse par lequel une personne produit des informations utilisables par d'autres, en utilisant un système de représentation, ce qu'on peut aussi appeler « **extraction des connaissances** ».



**Figure I.1** : Interprétation des informations – Représentation des connaissances

L'ingénierie des connaissances débouche sur une instrumentation technique permettant d'associer les inscriptions de connaissances sous une forme adaptée à des mécanismes d'interprétation prédéfinis (déduction ontologique, déduction analogique par exemple). La démarche de l'IC comme toute ingénierie passe par plusieurs étapes : recueil, modélisation et Reueil des connaissances représentation des connaissances.

### II.2.1 Recueil des connaissances

Le recueil est basé sur la formalisation et la consolidation des savoir-faire et des connaissances. Il tire sa source des connaissances implicites, acquises par l'expérience des spécialistes de l'entité, des documents qui contiennent déjà un savoir formalisé, des opérations passées. Il consolide le langage commun. **La méthodologie de Recueil** décrit ainsi les aspects d'initialisation, de capitalisation et de validation des connaissances. Plusieurs techniques de recueil existent. Leur application dépend des sources des connaissances Par exemple :

- Le recueil dirigé par des questionnements de l'expert sur les étapes identifiées dans l'étape précédente et en dressant des classifications des objets, des problèmes et des stratégies.
- L'extraction de concepts à partir de texte (« TextMining ») [Feldman & Sanger 2007]. Des algorithmes de détection d'occurrence de termes sont d'abord appliqués à des corpus de texte. L'expert peut identifier les concepts parmi les termes identifiés et dresser des classifications de ces concepts.
- L'agrégation de données par d'application d'algorithmes de « DataMining » [Fayyad et al. 1996].

## II.2.2 La modélisation des connaissances

La modélisation des connaissances a marqué la nouvelle génération des techniques d'IC. La plupart des approches se basent sur une représentation au niveau connaissances (Knowledge level) au sens de Newell [Newell 1982]. Ce niveau permet de modéliser un agent rationnel en rendant explicite le *corps* (« le Quoi »), les *actions* (le Comment) et les lois de conduite (« le Pourquoi »). Dans ce type de modélisation le modèle de domaine représente le « Quoi » et le modèle de *raisonnement* représente le « Pourquoi et Comment ».

### ➤ Les systèmes à base de connaissances :

Ils peuvent représenter et traiter des principes et des règles de décisions, des taxonomies, des théories, des processus et des méthodes mémorisées dans l'ordinateur. Ils sont capables d'aider l'utilisateur à accomplir des tâches de façon plus intelligente (voir figure I.2.).

Alan Newell a proposé une réponse de type "génie logiciel" pour le développement de systèmes à base de connaissances. Il s'agit d'exprimer dans un langage intermédiaire la représentation des connaissances utiles. Pour faciliter la réutilisation, les connaissances sont classées selon les buts du raisonnement, selon le type d'inférence attendu, etc. Ainsi défini avec un expert à un niveau de langage qui lui est accessible, il suffit ensuite de "dériver" le système à base de connaissances (au niveau symbolique de la machine). Cette dérivation exploite le fait qu'il existe des principes invariants pour la résolution de problème, le diagnostic, la recherche d'information, l'aide à la décision, la planification, etc.

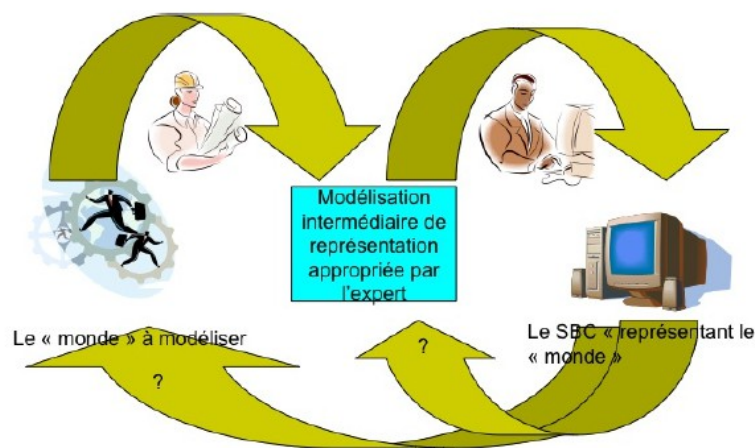


Figure I.2. Modélisation des connaissances.

### ➤ Difficultés :

- Acquisition des données.
- Organisation des données.
- Pas de classement universel des différents types de connaissances.

- Ambiguïté des termes.
- Contextualisation.
- Validation des données

➤ **Différentes approches :**

En IC, le modèle conceptuel est une représentation partagée entre l'expert (source des connaissances), l'ingénieur de la connaissance, les utilisateurs destinataires de la normalisation des connaissances et dans certaines cas, l'informaticien programmeur d'un système à base de connaissances.

Différentes approches définies en ingénierie des connaissances (CommonKADS [Breuker & Van de Velde 1994], KOD [Vogel 1982], MASK [Aries et al. 2008], etc.) offrent des guides permettant de fournir une telle représentation. Les techniques avancées dans ces approches peuvent être vues comme une grammaire de modélisation (appelée également langage de modélisation), un vocabulaire (dans lequel des primitives spécifiques à certains types de domaines sont définies) et une démarche (guidant l'ingénieur de la connaissance à définir le vocabulaire spécifique d'une application donnée en utilisant la grammaire définie).

### **II.2.3 La représentation des connaissances**

Il existe plusieurs modes de représentation des connaissances. Chaque mode de représentation possède des caractéristiques permettant une adaptation optimale à certaines réalités à modéliser. Que ce soit la caractérisation d'objets, l'étude de processus, la construction de système décisionnel ou logique, il existe un mode de représentation plus approprié pour la modélisation du système. Les arbres et les réseaux sémantiques, les ordinogrammes, les diagrammes causaux, les arbres de décisions et les modèles orientés objets sont autant de modes de représentation des connaissances qui sont fondés sur une approche graphique [Paquette2002] [Paquette2006].

• **Arbre sémantique :**

L'arbre sémantique est très utilisé en pédagogie. Il sert à mettre en évidence les éléments de connaissance sous la forme d'une structure hiérarchique partant d'un objet général et d'un classement par spécialisation des classes en sous-classes et de classes en individus (Figure I.3). Il n'y a pas de liens entre les branches de la hiérarchie (liens croisés). Ce type de représentation est très utilisé aussi en création de site Web.

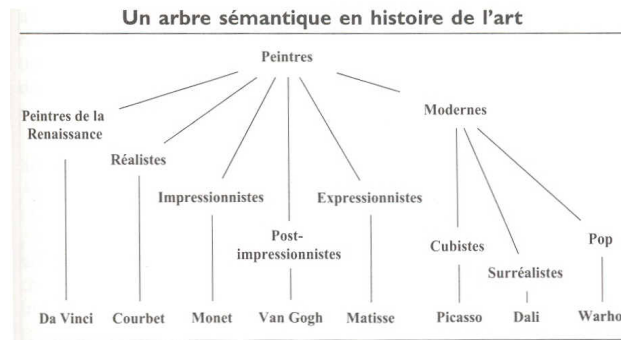


Figure I.3 : Exemple d'arbre sémantique.

• **Le réseau sémantique :**

Le réseau sémantique (appelé aussi carte conceptuelle et graphe entité-relations) (Figure.I.4) permet de mettre en relation des concepts et des faits par la dénomination de liens. Cette représentation est très utilisée en informatique, notamment pour la conception de bases de données, mettant ainsi en relation les tables qui regroupent les concepts décrits par leurs attributs. En intelligence artificielle, le réseau sémantique permet de créer une base de faits qui caractérise un domaine donné.

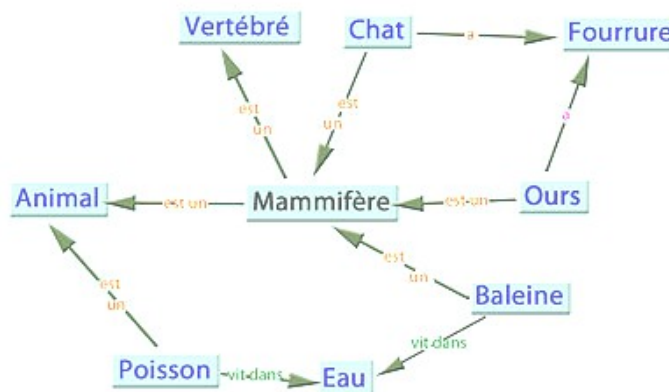


Figure I.4 : Exemple de réseau sémantique.

• **L'ordinogramme :**

L'ordinogramme est utilisé afin de modéliser des procédures. Jusqu'ici, les graphiques représentaient des concepts et leurs relations. Dans un ordinogramme, les objets sont des actions ou des décisions, et les liens représentent le flux d'informations d'une action ou d'une décision à l'autre. La notion de séquence est ainsi introduite dans ce type de représentation (Figure I.5).

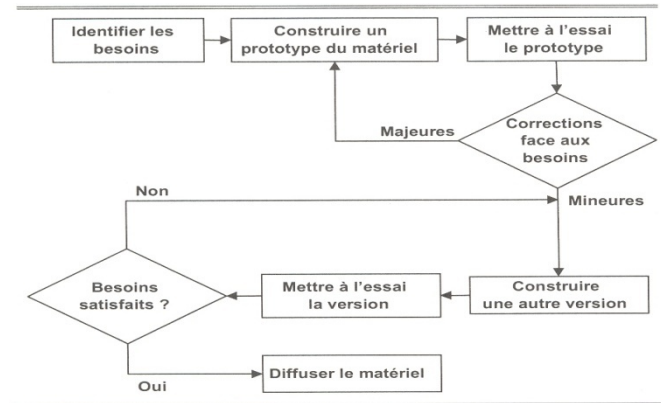


Figure I.5 : Exemple de réseau sémantique.

• **Diagramme causal et arbre de décision**

Le diagramme causal et l’arbre de décision sont très utilisés dans les systèmes d’aide au diagnostic ou à la prise de décision et dans les systèmes experts, en établissant des relations de cause à effet entre différents facteurs. La lecture du diagramme se fait de gauche à droite. Lorsqu’un facteur est observé, les possibilités résultantes sont indiquées par des liens. Dans l’exemple de la figure I.6, le non-démarrage de la voiture peut être occasionné par la défaillance du carburateur, du moteur ou de l’alimentation. À son tour, la défaillance du moteur peut dépendre de la mauvaise lubrification et celle-ci peut dépendre à son tour de la température ou encore du niveau ou de la pression d’huile.

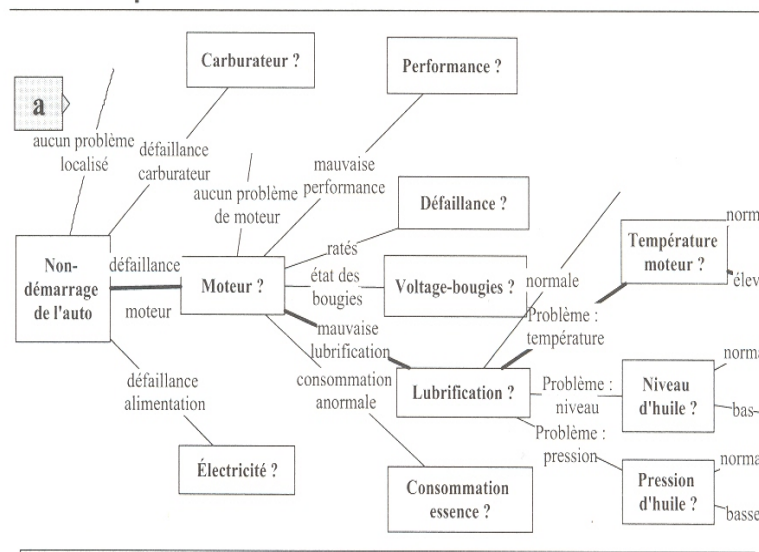
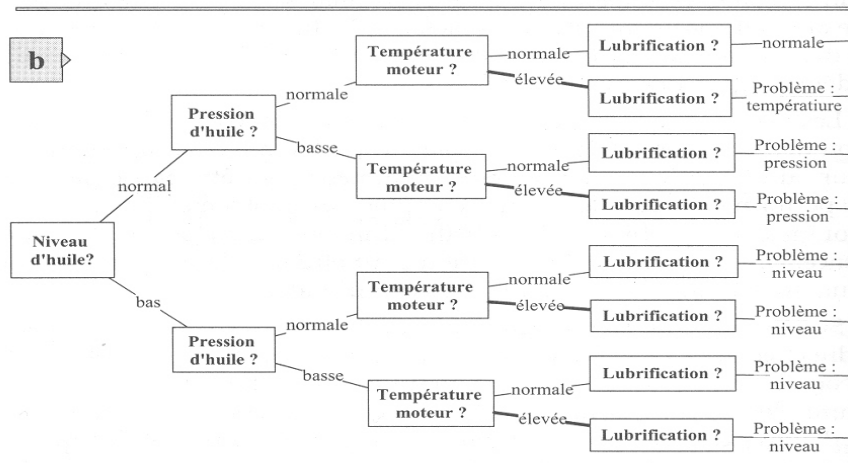


Figure I.6 : Exemple de diagramme causal

L’arbre de décision, dont la structure est déduite du diagramme causal, sert d’outil d’aide à la prise de décision. Il permet d’extraire une base de règles et, inversement, une base de règles peut être représentée graphiquement par un arbre de décision. La Figure I.7 présente l’arbre de décision extrait de la partie du diagramme causal qui concerne la lubrification.



**Figure I.7:** Exemple d'arbre de décision.

En fait, un arbre de décision est équivalent à un ensemble de règles qui constitueront la base de connaissances du système expert. Dans ce cas-ci, on obtient de l'arbre de décision de la Figure I.7 les règles suivantes.

- Si *Niveau-huile = normal* et *Pression-huile = normale* et *Temp-moteur = normale*  
 Alors *Lubrification = normale*
- Si *Niveau-huile = normal* et *Pression-huile = normale* et *Temp-moteur = élevée*  
 Alors *Lubrification = Problème de température*
- Si *Niveau-huile = normal* et *Pression-huile = basse*  
 Alors *Lubrification = Problème de pression d'huile huile*
  - Si *Niveau-huile = bas*  
 Alors *Lubrification = Problème de niveau d'huile*



- **La modélisation orientée objet :**

La modélisation orientée objet s'est développée à la fin des années 1980 notamment lors de l'émergence de la programmation par objet. La figure I.8 présente un modèle orienté objet d'un scénario pédagogique indiquant que le scénario prend la forme d'un modèle MOT (Modèle par Objets Typés) et qu'il est composé de plusieurs objets modélisés. Ces objets peuvent être des ressources, des règles, des liens ou des activités. Modèle orienté objet.

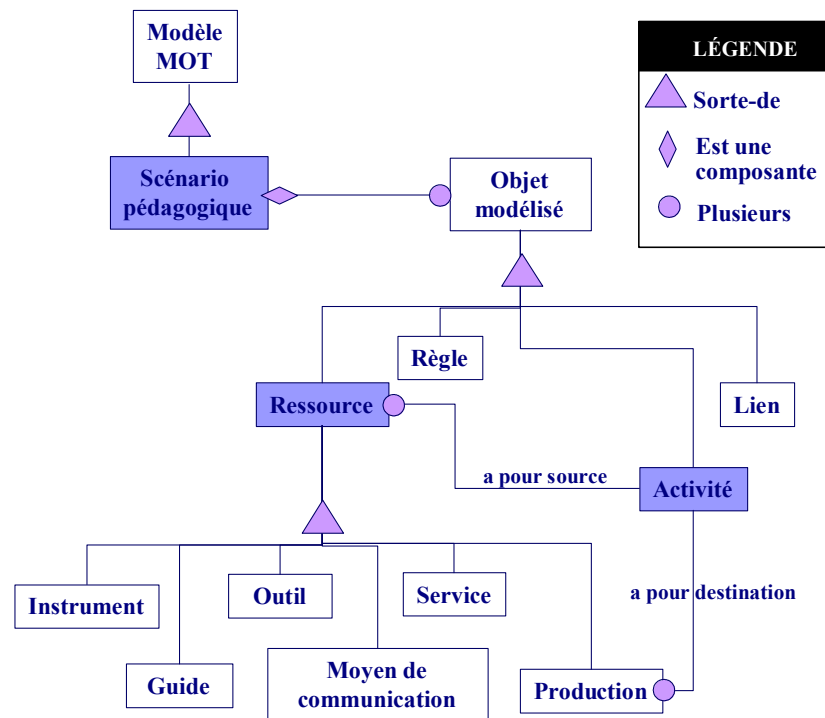


Figure 2 : Exemple de modélisation orientée objet.

- **Graphe conceptuel :**

Les graphes conceptuels sont une représentation à la fois computationnelle et interprétable, compromis entre la logique trop souvent abstraite et le langage naturel trop souvent illogique (voir chapitre II).

➤ **Ontologie et représentation des connaissances**

Pour assurer un usage plus large et mieux intégré des modèles de connaissances, on doit à la fois assurer une généralité plus grande de la représentation et une lisibilité à la fois par les humains et par les logiciels. L'utilisation des ontologies fournit une spécification aux modèles de connaissances. Une ontologie définit les concepts d'un domaine de connaissances, sous forme de *classes* d'objets, d'*instances* de concepts ou d'objets d'un domaine, les *relations* entre classes et objets et les *propriétés* (ou axiomes) des objets et des relations. Au surplus, les synergies entre les domaines de connaissances sont possibles par la création de renvois d'une ontologie vers une autre.

## Biographies :

[Charlet, 2003] Charlet J. *L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales*. Pierre et Marie Curie – Paris VI, 2003.

[Aussenac-Gilles et al. 1996] : N. Aussenac-Gilles, P. Laublet, C. Reynaud, *Acquisition et Ingénierie des connaissances, Tendances Actuelles*, Cepadues Editions, 1996.

[Bach 2004] Bachimont, B.. Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle. Mémoire de HDR, Université de Technologie de Compiègne, 2004.

[Feldman & Sanger 2007] : Feldman, R., & Sanger, J. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press. (2007).

[Fayyad et al. 1996] : Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. *Advances in knowledge discovery and data mining*. (1996).

[Newell 1982] L. Newell: *The Knowledge level*, Artificial Intelligence Journal, 19(2), 1982.

[Vogel 88] C. Vogel : *Génie Cognitif*, Ed. Masson, Paris 1988.

[Aries S. et al, 2008] : Aries S., Le Blanc B., Ermine J-L, *MASK : une méthode d'ingénierie des connaissances pour l'analyse et la structuration des connaissances*, Série Management et Gestion des STIC, Hermes-Lavoisier, 2008.

[Paquette2002] Paquette, G. (2002). Modélisation des connaissances et des compétences. Montréal, Presses de l'université du Québec.

[Paquette2006] Paquette, G., Héon, M., & Basque, J. (2006). La modélisation des connaissances en milieu organisationnel. *Centre de recherche LICEF, Québec, Canada*.