

Projet BourbaKeM

Élément n°12

L'ingénierie des connaissances

Nada Matta¹

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. La démarche de l'IC comme toute ingénierie passe par plusieurs étapes : recueil, modélisation et représentation des connaissances.

2. Recueil des connaissances

Une familiarisation avec le domaine d'expertise et de l'activité est indispensable afin d'identifier sur les thématiques ainsi que les sources des connaissances : experts, documents, Base de données, etc. Un entretien libre avec l'expert peut être mené afin de dresser un schéma global de l'activité décrivant un ordonnancement des grandes étapes est dans certain cas, nécessaire comme une première phase du recueil.

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].

3. 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 ».

C'est à travers cette modélisation de l'activité au niveau rationnel, en distinguant la tâche prescrite du contrôle de l'activité réelle, que l'IC a fourni une flexibilité dans la modélisation des connaissances. Cette modélisation fournit un cadre permettant de se focaliser sur la description d'agents rationnels et de leur conduite avant leur transformation dans un format calculable, programmable dans un automate. C'est dans ce sens, que l'IC a permis non seulement de fournir des systèmes calculables (Systèmes à base de connaissances) mais également une représentation formelle de la connaissance, exploitable sous plusieurs formes (comme par exemple, un accès cognitif à l'information et au capital connaissance d'une entreprise, une structuration conceptuelle de différentes facettes d'une donnée, etc.).



Cette modélisation décrit un modèle abstrait qu'on appelle modèle conceptuel, cadre sémantique partagé par des utilisateurs et des concepteurs qui leur permet de communiquer. 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 formalisation des connaissances et dans certains 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). Citons par exemple, le langage CML « Conceptual Modelling Language » [Schreiber et al. 1994] qui permet de représenter le modèle de domaine sous forme de concepts et le modèle de raisonnement avec des tâches et des inférences.

4. Représentation des connaissances

Deux catégories de langages sont proposées pour la représentation des connaissances [Kayser 1997] :

1. Les *frames*, proches d'une représentation objet en informatique. Les concepts sont exprimés sous forme de classe hiérarchisée. Chaque classe comme (c'est le cas de la programmation orientée objets) est décrite par des attributs/Valeurs. Des instances représentent les exemples de classes.

2. Les *réseaux sémantiques*, formalisés sous forme de concepts, relations. Les graphes conceptuels sont proposés par Sowa [Sowa 91] pour traduire cette représentation en logique propositionnelle. Les relations entre les concepts permettent de décrire ses propriétés. La relation « Est-un », utilisée dans les langages orientés objet, est privilégiée afin d'illustrer l'héritage de propriétés. Par analogie, toutes les relations au niveau concept père sont héritées au niveau concepts fils.

Le réseau sémantique permet une puissance de raisonnement, exploité dans ce qu'on appelle actuellement « Ontologie ». Une ontologie, notion héritée de la philosophie grecque illustrant la science de l'être, est une représentation de concepts d'un domaine [Fensel 2001]. Elle est décrite sous forme d'un arbre d'héritage entre les concepts et un graphe décrivant les relations entre les concepts. Plusieurs niveaux d'ontologies ont été définis [Gangemi et al. 2000] : générique, spécifique à un domaine et relative à une application dans un domaine particulier. Plusieurs langages d'implémentation sont également définis comme RDF, OWL, etc. De même, des logiciels adossés à des moteurs d'inférences permettant de représenter les ontologies, sont proposés. Nous pouvons citer par exemple PROTEGE et CORESE.

L'exploitation de l'ontologie, utilisant ces logiciels permet par exemple de fournir une recherche intelligente des informations dans les documents, sur Internet. Ce que le Web sémantique [Berners-lee et al. 2001] propose.

5. Bibliographie de base

- [Aussenac-Gilles et al. 1996] : N. Aussenac-Gilles, P. Laublet, C. Reynaud, *Acquisition et Ingénierie des connaissances, Tendances Actuelles*, Cepadues Editions, 1996
- [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*, Chapitre 7 de l'ouvrage : Management et ingénierie des connaissances, modèles et méthodes, Traité IC2, Série Management et Gestion des STIC, Hermes-Lavoisier, 2008
- [Berners-lee et al. 2001] : Berners-Lee T., Hendler J. & Lassila O., *The Semantic Web A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American, May 17, 2001
- [Breuker & Van de Velde 1994] : Breuker and W. Van de Velde, *CommonKADS Library for expertise modelling, Reusable problem solving components*, (Eds), IOS press, Amsterdam, 1994
- [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.
- [Fayyad et al. 1996] : Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. *Advances in knowledge discovery and data mining*. (1996)
- [Feldman & Sanger 2007] : Feldman, R., & Sanger, J. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press. (2007)

[Fensel 2001] : Fensel, D: *Ontologies* (pp. 11-18). Springer Berlin Heidelberg, (2001)

[Gangemi et al. 2002] : Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). *Sweetening ontologies with DOLCE. In Knowledge engineering and knowledge management: Ontologies and the semantic Web* (pp. 166-181). Springer Berlin Heidelberg.

[Kayser 1997] : Daniel Kayser : *La représentation des connaissances*, Hermes, 1997

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

[Sowa 91] J. F. Sowa: *Principles of semantic networks: exploration in the representation of knowledge*, J.F. Sowa (Ed.), Morgan Kaufmann, 1991.

[Schreiber et al. 1994] Schreiber G., Wielinga B., Akkermans H., Van de Velde W., Anjewierden A. (1994) *CML: The commonKADS conceptual modelling language*. In: Steels L., Schreiber G., Van de Velde W. (eds) *A Future for Knowledge Acquisition. EKAW 1994. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, vol 867. Springer, Berlin, Heidelberg

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

6. L'auteur



Nada Matta est professeur en Ingénierie et Gestion des connaissances à l'Université de Technologie de Troyes.

Elle travaille sur la valorisation du savoir-faire, la traçabilité de projets et la capitalisation du retour d'expérience dans les activités coopératives comme la conception de produits, la gestion de crise, etc.

Plusieurs applications de ses recherches sont réalisées chez EADS, Dassault-Aviation, Décathlon, IFTH, Snecma, PSA, INRS, CSTB, SAMU10, ...

<http://matta.tech-cico.fr/>

E-mail: nada.matta@utt.fr