

# Raisonner le Web Sémantique avec des graphes : Application à un cas industriel

Olivier Carloni

LIRMM, Université de Montpellier II,  
161, rue Ada, F-34392 Montpellier cedex - France  
carloni@lirmm.fr

Mondeca,  
3, cité Nollez, 75018 Paris  
<http://www.mondeca.com>

**Abstract.** Ce travail décrit un projet de recherche qui sera réalisé dans le cadre d'une thèse CIFRE en collaboration entre la société Mondeca et l'équipe RCR (Représentation des Connaissances et Raisonnements) du LIRMM.

## 1 Le contexte

### 1.1 La société

Mondeca est un éditeur de logiciel impliqué dans le Web Sémantique. Elle a développé ITM (Intelligent Topic Manager), un logiciel de gestion documentaire et d'ingénierie des connaissances. Un processus d'annotation semi-automatique permet d'extraire les informations pertinentes contenues dans les documents. Ses dernières sont alors représentées dans une base de connaissances que les utilisateurs peuvent consulter et enrichir. L'objectif de Mondeca pour les années à venir est d'améliorer la gestion des connaissances qu'offre déjà ITM, notamment en automatisant certaines vérifications et en munissant son outil de facultés d'inférence.

### 1.2 Le principal formalisme de description de la connaissance dans ITM

Les informations extraites des textes sont décrites dans la base de connaissances en utilisant le formalisme des Topic Maps. Une topic map [FIG01] se construit à partir de trois primitives typées : les topics, les roles et les associations. Les topics peuvent être reliés à des associations d'arité quelconque par l'intermédiaire de roles (tous les roles sont binaires). Chaque topic joue un rôle dans l'association qui le connecte à un autre topic. Les roles permettent, en quelque sorte, de nommer les "paramètres" des associations. Le type d'une association n'impose pas à son instance d'arité fixe. Chaque primitive du réseau topic map possède un identifiant unique et la structure d'un topic et d'une association permet de comporter des attributs.

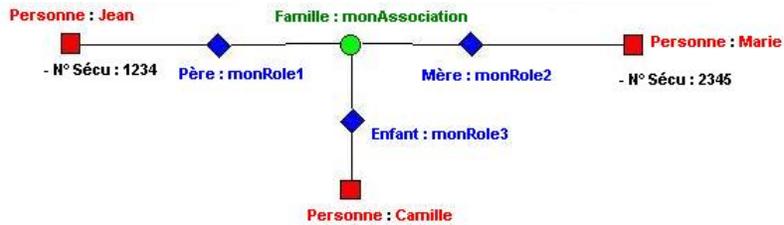


Fig. 1. Une topic map représentant une famille

### 1.3 Le modèle $M_{ITM}$ de représentation de la connaissance dans ITM

Le modèle  $M_{ITM}$  de représentation des connaissances défendu par Mondeca comporte trois niveaux de topic maps :

1. Le niveau “méta”, qui se définit lui même (définition réflexive) et comporte tous les constructeurs permettant de créer des ontologies de domaine pour chaque client. Ce niveau évolue très peu et est commun à tous les clients.
2. Le niveau “ontologique”. C’est à ce niveau que sont créées les ontologies de domaine pour chaque client en utilisant les constructeurs du niveau méta. On définit ici des types et des propriétés générales qui seront utilisées au niveau “factuel”. Il existe plusieurs sortes d’“ontologies” différentes : les ontologies de domaine spécifiques au client, mais aussi des ontologies construites pour gérer les services offerts par le logiciel. On retrouve, par exemple, les ontologies de “gestion documentaire”, de “gestion terminologique”, de “gestion des connaissances”.
3. Le niveau “factuel”. Les types et propriétés de l’ontologie permettent de définir les éléments de ce niveau. Les descriptions faites peuvent être du genre organisationnel (se réfère à l’ontologie de gestion documentaire), terminologique ou “faits du monde réel”.

La hiérarchie entre les niveaux est établie par une relation d’instanciation qui permet de typer les éléments d’un niveau en utilisant les types définis dans le niveau supérieur.

L’exemple FIG02 illustre la contrainte qu’exerce le réseau des topics décrit à un niveau, sur la topologie du réseau et la structure des éléments du niveau inférieur. “Jean” ne peut être connecté à l’association3 que si dans l’ontologie la classe de “Jean” est bien connectée par “type\_d’association\_permis” au topic représentant le type de l’association3 (i.e. “appartenir\_famille”).

Les ontologies d’ITM sont d’abord définies en OWL en utilisant l’outil d’édition Protégé [PR01].

Une fois l’ontologie établie en OWL pour un client (après étude des besoins client), cette dernière est traduite en Topic Maps pour être transférée dans ITM.

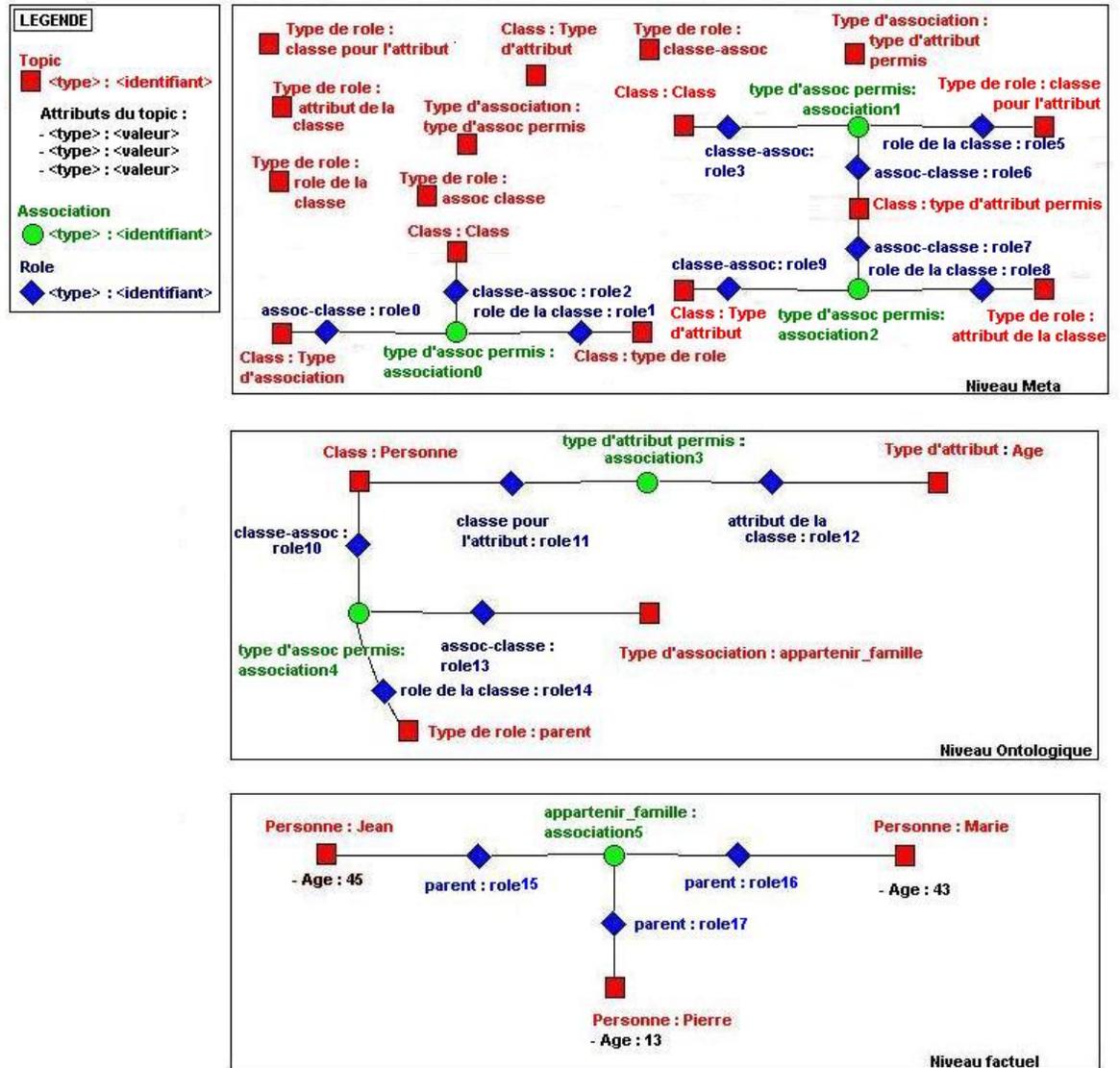


Fig. 2. Le modèle  $M_{ITM}$  avec son niveau méta, ontologique et factuel.

Le formalisme des Topic Maps a été délibérément conçu pour s'affranchir d'une sémantique trop contraignante et n'a donc pas de sémantique formelle. OWL devrait permettre de lui en fournir une.

Pourtant, Mondeca souhaite pourvoir son outil de capacités de raisonnement permettant notamment :

1. de contrôler la cohérence des descriptions faites dans la base par rapport au modèle (exemple : vérifier les faits par rapport à l'ontologie),
2. de déduire à partir de connaissances implicites exprimées sous forme de règles, de nouvelles représentations ou des réponses à des requêtes.

## 2 Les problématiques de recherche

### 2.1 L'hypothèse de recherche

Nous faisons l'hypothèse que les graphes conceptuels sont un bon candidat pour doter ITM de facultés d'inférence.

Le but de ce travail est donc d'évaluer dans quelle mesure la famille SG permet d'opérationnaliser les raisonnements dans ITM. La famille SG [BA01] des graphes conceptuels s'obtient à partir des graphes conceptuels simples essentiellement définis par :

- Le support : Il est composé d'une hiérarchie de types de concepts et de types de relations primitifs, d'ensembles de types incompatibles et d'un ensemble de marqueurs individuels typés.
- Les graphes conceptuels simples : Ce sont des graphes bipartis comportant des sommets relations et des sommets concepts qui sont étiquetés à partir du support. Les sommets concepts sont étiquetés par une conjonction correcte de types primitifs (pas de types incompatibles) et par un marqueur qui est soit le marqueur individuel, soit le marqueur générique.
- La projection : C'est un homomorphisme de graphes. La projection permet d'effectuer des raisonnements logiquement fondés car elle est adéquate et complète avec la déduction en logique du premier ordre sur les formules existentielles conjonctives.

A partir de ce modèle de base, on peut obtenir d'autres modèles plus expressifs en introduisant des règles ou des contraintes.

### 2.2 Sémantique formelle

Un des travaux théoriques importants est d'établir la sémantique formelle des représentations faites dans ITM à partir du modèle intuitif de représentation des connaissances ( $M_{ITM}$ ) et du modèle décrit en OWL. Les principaux constructeurs de OWL utilisés sont ceux de OWL-Lite, mais le découpage n'est pas si simple et une petite partie de OWL-DL et OWL-Full est nécessaire pour une

représentation des connaissances en OWL cohérente avec le modèle  $M_{ITM}$ . Cette formalisation est en cours.

Lorsque la sémantique formelle est bien définie, il faut établir quelle est la portion des descriptions de la base qui peut être transcrite en graphes conceptuels tout en préservant cette sémantique. Il semble, à ce stade de l'étude, qu'une grande partie des descriptions peuvent être importées sans problème dans le modèle des graphes conceptuels simples.

### 2.3 Transformations en graphes conceptuels

Pour pouvoir raisonner en graphes conceptuels sur la connaissance décrite dans ITM, il faut définir des transformations syntaxiques du langage des topic maps vers le langage des graphes conceptuels tout en préservant le modèle de représentation des connaissances de Mondeca.

La transformation n'est pas unique car le modèle  $M_{ITM}$  comporte 3 niveaux conceptuels (les niveaux méta, ontologique et factuel) alors que celui des graphes conceptuels en comporte 2 (le support et le graphe), il existe alors trois ou quatre transformations potentiellement correctes et utiles.

### 2.4 Extensions pour ITM

Les graphes conceptuels permettront aussi d'enrichir le langage de représentation des connaissances de ITM. Par exemple, les graphes conceptuels simples permettent de représenter un individu non identifié par le marqueur générique, ce qui n'est pas possible pour l'instant dans ITM. Les règles et les contraintes de la famille SG augmenteront l'expressivité du modèle. Les extensions réalisées devront se positionner par rapport aux standards existants ou émergents du Web Sémantique (RDF, OWL, le tout jeune RuleML, ...)

### 2.5 Passage à l'échelle

ITM gère des bases de connaissances dont la taille peut atteindre le million de topics. Il est important de garantir un temps de réponse acceptable. Il faudra déterminer si le passage à l'échelle du moteur d'inférence se fait correctement.

CoGITaNT [CO01] est un des outils qui a été développé pour permettre des inférences en graphes conceptuels et TooCoM [TO01] permet de visualiser, d'éditer et d'opérationnaliser des ontologies. L'étude doit démontrer si dans l'état actuel de CoGITaNT (architecture, organisation des données, algorithmes), on peut gérer des graphes massifs en un temps raisonnable.

TooCoM permettra de visualiser et d'éditer les axiomes des ontologies avant d'évaluer les performances, dans CoGITaNT, d'une inférence à partir de ces axiomes.

## References

- [PR01] Editeur d'ontologies Protégé. Reference web : <http://protege.stanford.edu>
- [CO01] CoGITaNT (COncceptual Graphs Integrated Tools Allowing Nested Typed Graphs) Reference web : <http://cogitant.sourceforge.net>
- [BA01] Jean-François Baget, Marie-Laure Mugnier. The SG Family: Extensions of Simple Conceptual Graphs, Proceedings of the seventeenth International Joint Conference on Artificial Intelligence ( IJCAI'2001), Seattle, Washington, USA, August 2001, vol. I, pages 205-210, Morgan Kaufmann, 2001.
- [TO01] Frederic Fürst, Michel Leclère, Francky Trichet: Operationalizing Domain Ontologies: A Method and a Tool. ECAI 2004: 318-322.