
Construction d'ontologies à partir de textes : une approche basée sur l'Ingénierie Dirigée par les Modèles

Henry Valéry Téguiak

Laboratoire d'Informatique Scientifique et Industrielle (LISI)
Ecole Nationale Supérieure de Mécanique et d'Aérotechnique (ENSMA),
1, avenue Clément Ader, BP 40109,
86960 FUTUROSCOPE Cedex France

teguiakh@ensma.fr

MOTS-CLÉS : Méta-modélisation, MOF, Ontologie, Ingénierie Dirigée par les Modèles.

KEYWORDS: Meta-modeling, MOF architecture, Ontology, Model Driven Engineering

La plupart des outils de construction d'ontologies existant permettent de construire des ontologies en précisant la représentation des concepts et leurs sémantiques. Ces derniers ne précisent ni la manière de trouver les concepts ni la façon de décrire leur signification. Ces outils supposent que l'ontologie de domaine que l'on souhaite représenter existe déjà. Or, la conception d'une telle ontologie s'avère particulièrement difficile, surtout si l'on souhaite qu'elle fasse l'objet de consensus dans une communauté assez large. Un moyen très largement utilisé pour atteindre cet objectif est de partir d'éléments préexistants dans le domaine : corpus textuels, taxonomies, fragments d'ontologie préexistants, et de les exploiter comme base informationnelle pour définir progressivement l'ontologie du domaine. C'est cette problématique que tente de résoudre le projet ANR DaFOE4App au sein duquel s'inscrit notre thèse qui vise à proposer une plate-forme suffisamment flexible de sorte à supporter divers scénarii de construction d'ontologies.

Dans le cadre de notre thèse, nous abordons la notion de flexibilité au sein de la plate-forme DaFOE sur deux axes : (1) la flexibilité des modèles vise à proposer une architecture capable de supporter l'évolution des modèles au sein de la plate-forme tant au niveau persistance des données qu'au niveau des programmes d'accès à ces données, ainsi que la diversité des modèles d'ontologies, (2) la flexibilité fonctionnelle visant à proposer une architecture logicielle facilement extensible. Pour résoudre les problèmes sous-jacents au premier axe, nous utilisons l'approche de l'Ingénierie Dirigée par les Modèles (également appelée gestion des modèles (Bernstein 03)) qui utilise des architectures analogues à celle du MOF (Meta Object Facility) pour la programmation au niveau du modèle et la génération de code. L'architecture du MOF est caractérisée par l'existence de trois niveaux de modélisation : M_1 pour les modèles, M_2 pour les méta-modèles et M_3 pour le méta-

méta-modèle (M_0 , pour les instances). Chaque niveau est représenté comme instance d'une structure définie par le niveau de modélisation supérieur, le M_3 étant réflexif.

Les premiers travaux que nous avons réalisés ont porté sur la construction d'ontologies à partir de textes (Aussenac-Gilles *et al.* 00) qui s'intéresse à la possibilité d'utiliser des corpus de textes d'un domaine, en extraire, à l'aide d'outils de Traitement Automatique de la Langue (TAL), des termes ainsi que les relations entre ces termes et construire, in fine, une ontologie de domaine à partir de ces termes. Un tel processus de construction passe par 3 phases (Charlet *et al.* 08) : (1) une phase d'analyse terminologique qui permet l'extraction automatique des termes et relations des textes, (2) une phase d'analyse termino-ontologique, nécessaire pour d'éventuelles désambiguïsations des termes et relations et enfin, (3) une phase d'analyse ontologique permettant de créer des concepts formels de l'ontologie.

Compte tenu d'une part de la diversité des informations extraites par les outils de TAL, et d'autre part de la volonté de semi-automatiser ce processus de construction, nous avons conçu une architecture de (méta)modélisation qui, d'une part, permet de représenter séparément les modèles correspondants à chacune des phases ci-dessus, et qui, d'autre part, pour rendre flexible le modèle de chacune des phases, le représente comme une architecture à trois niveaux de modélisation. Celle-ci s'inspire de l'architecture de base de données à base ontologique OntoDB (Hondjack *et al.* 07).

Par ailleurs, afin d'offrir la possibilité de réutiliser notre système pour la construction d'ontologies à partir d'autres sources d'information tels que les thésaurus, les schémas de base de données ou des ontologies existantes, nous avons introduit des possibilités d'entrées sorties pour chacune des phases de modélisation. De plus, nous permettons de définir des morphismes entre modèles pour générer, à partir des instances du modèle d'une étape, une proposition des instances du modèle de l'étape suivante et inversement. Comme solution au problème de flexibilité fonctionnelle, nous avons proposé une architecture logicielle constituée d'une *application noyau* et extensible à base de *greffons (plug-ins)*.

Bibliographie

- N Aussenac-Gilles, B. Biebow, S. Szulman "Revisiting ontology design : a methodology based on corpus analysis", 12th *International Conference, EKAW 2000 Proceedings*, October 2000; p. 27-66.
- Philip A. Berstein., Applying model management to classical meta data problems, *CIDR SIGMOD*, 2003, p 209-220.
- J. Charlet, S. Szulman, G. Pierra, N. Nadiyah, H.V. Téguiak, N Aussenac-Gilles, A. Nazarenko, "Dafoe : A multimodel and multimethod platform for building domain ontologies", *JFO2008 Lyon*, décembre 2008, p. 74-85.
- H. Dehainsala, G. Pierra, L. Bellatreche, "OntoDB : An ontology-based database for intensive applications", *DASFAA*, 2007, p. 497-508.