

# DAFOE : une plateforme multiméthodes et multimodèles pour construire des ontologies de domaine\*

S. Szulman<sup>1</sup>  
Nathalie Hernandez<sup>4</sup>  
Valery Teguiak<sup>6</sup>

J. Charlet<sup>2,3</sup>  
Nadia Nadah<sup>5</sup>

N. Aussenac-Gilles<sup>4</sup>  
Éric Sardet<sup>6</sup>

A. Nazarenko<sup>1</sup>  
Jean Delahousse<sup>7</sup>

<sup>1</sup> LIPN - UMR 7030, Université Paris 13 - CNRS - France

<sup>2</sup> INSERM UMR\_S 872, Eq. 20, Paris

<sup>3</sup> Université Pierre et Marie Curie ; AP-HP, Paris

<sup>4</sup> CNRS/IRIT et Université de Toulouse

<sup>5</sup> Heudiasyc CNRS/UMR 6599, Université de Technologie de Compiègne

<sup>6</sup> LISI-ENSMA et CRITT-Informatique, Poitiers

<sup>7</sup> MONDECA, Paris

Sylvie.Szulman@lipn.univ-paris13.fr, Jean.Charlet@spim.jussieu.fr

## Résumé

*La construction d'ontologie à partir de textes fait l'objet d'études depuis plusieurs années dans le domaine de l'ingénierie des ontologies. Un cadre méthodologique en quatre étapes (constitution d'un corpus de documents, analyse linguistique du corpus, conceptualisation, opérationnalisation de l'ontologie) est commun à la plupart des méthodes de construction d'ontologies à partir de textes. S'il existe plusieurs plateformes de traitement automatique de la langue (TAL) permettant d'analyser automatiquement les corpus et de les annoter tant du point de vue syntaxique que statistique, il n'existe actuellement aucune procédure généralement acceptée, ni a fortiori aucun ensemble cohérent d'outils supports, permettant de concevoir de façon progressive, explicite et traçable une ontologie de domaine à partir d'un ensemble de ressources informationnelles relevant de ce domaine. Le but de cet article est de présenter les propositions développées, au sein du projet ANR DaFOE 4app, pour favoriser l'émergence d'un tel ensemble d'outils.*

## Mots Clef

Ontologie, Ingénierie des connaissances, métamodélisation, TAL.

## Abstract

*The concept of ontologies, appeared in the nineties, constitute a key point to represent and share the meaning carried*

\*DaFOE4App <<http://dafoe4app.fr>> est un projet ANR TLOG 010 qui vise à construire une plateforme de construction d'ontologie, DaFOE.

*out by formal symbols. Thus, the building of such an ontology is quite difficult. A way to do so is to use preexistent elements (textual corpus, taxonomies, norms or other ontologies) and operate them as a basis to define the ontology field. However, there is neither accepted process nor set of tools to progressively built ontologies from the available resources in a traceable and explicit way. We report in this paper several propositions developed within the framework of the ANR DaFOE4App project to support emergence of such tools. Taking into account the different needs, methods and models, our proposal rely on 3 key points : 1) defining a general methodological framework to integrate the accomplishment of several design scenarios ; 2) defining a modeling structure allowing various templates ; 3) specifying and developing a platform able to integrate various kind of tools currently used in an autonomous way within a modeling structure. Moreover, the platform should insure persistence and traceability and allow the building of formal ontologies from the analysis of textual corpora using NLP (Natural Language Processing) annotated texts.*

## Keywords

Ontology, Knowledge Engineering, NLP, metamodelization

## 1 La plateforme DAFOE

Depuis son émergence, au début des années 1990, dans les recherches en modélisation de connaissances, la notion d'ontologie s'est rapidement diffusée dans un grand nombre de domaines de recherche en informatique. Compte tenu du caractère très prometteur de cette notion, de nombreux travaux ont visé à permettre son utilisation

dans des domaines aussi divers que le traitement automatique de la langue naturelle, la recherche d'information, le commerce électronique, le web sémantique, la spécification des composants logiciels et l'intégration de système d'information.

L'efficacité de toutes ces approches présuppose néanmoins l'existence d'une ontologie de domaine susceptible d'être développée, ou d'être mise en œuvre, au sein de l'application cible. 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, normes ou fragments d'ontologie préexistants, et de les exploiter comme base pour définir progressivement l'ontologie du domaine. La construction d'ontologie à partir de textes fait l'objet d'études depuis plusieurs années dans le domaine de l'ingénierie des ontologies. Un cadre méthodologique en quatre étapes (constitution d'un corpus de documents, analyse linguistique du corpus, conceptualisation, opérationnalisation de l'ontologie) est commun à la plupart des méthodes de construction d'ontologies à partir de textes (TERMINAE<sup>1</sup> [1], Text2Onto [3]). Ces méthodes sont implémentées dans des outils qui se distinguent par leur approche de la phase de conceptualisation plus ou moins automatique [4]. Cependant s'il existe des outils largement utilisés, tels que Protégé, pour représenter formellement une ontologie supposée déjà conçue, et s'il existe également plusieurs plateformes de traitement automatique de la langue (TAL) permettant d'analyser automatiquement les corpus et de les annoter tant du point de vue syntaxique que statistique, il n'existe actuellement aucune procédure généralement acceptée, ni a fortiori aucun ensemble cohérent d'outils supports, permettant de concevoir de façon progressive, explicite et traçable une ontologie de domaine à partir d'un ensemble de ressources informationnelles relevant de ce domaine. C'est ce que nous proposons dans la plateforme DaFOE pour laquelle nous espérons que de nombreux grefons viendront l'enrichir.

Un cadre méthodologique a été élaboré durant la définition de la plateforme. Il a été utilisé de deux façons, à savoir comme cadre permettant d'avoir une description commune des processus mis en jeu en même temps que modèle évoluant pour être à même de tenir compte des desiderata de tous les partenaires. Ainsi, la plateforme a différents niveaux d'entrées, correspondant aux différentes ressources, et différents niveaux de sortie correspondant à des produits de plus en plus élaborés (1) des réseaux terminologiques s'organisant durant l'analyse des données, (2) un niveau termino-conceptuel où les concepts sont organisés et (3) un niveau où l'ontologie est formalisée [2] (cf. fig. 1).

Pour valider l'architecture de méta-modélisation proposée, un premier prototype est en cours de réalisation par le LISI.

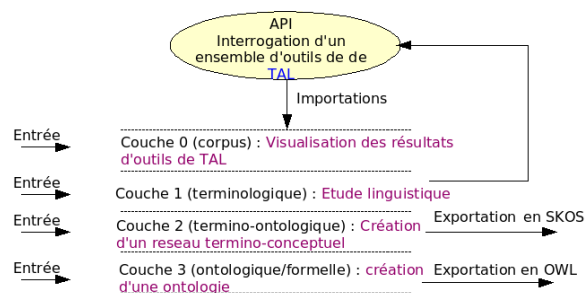


FIG. 1 – Les couches du modèle de données.

C'est ce prototype dont nous ferons la démonstration.

## 2 Remerciements

Les auteurs remercient l'ensemble des membres du projet DaFOE 4app pour le travail réalisé qui a pu être synthétisé dans cet article.

## Références

- [1] N. Aussenac-Gilles, B. Biébow, and S. Szulman. Revisiting ontology design : a methodology based on corpus analysis. In R. Dieng and O. Corby, editors, *Knowledge Engineering and Knowledge Management : Methods, Models, and Tools. Proc. of the 12th International Conference, (EKAW'2000)*, LNAI 1937, pages 172–188. Springer-Verlag, 2000.
- [2] J. Charlet, S. Szulman, G. Pierra, N. Nadah, H. V. Tequiak, N. Aussenac-Gilles, and A. Nazarenko. Dafoe : A multimodel and multimethod platform for building domain ontologies. In D. Benslimane, editor, *2<sup>e</sup> Journées Francophones sur les Ontologies*, Lyon, France, novembre 2008. ACM.
- [3] P. Cimiano and J. Volker. Text2onto - a framework for ontology learning and data-driven change discovery. In A. Montoyo, R. Munoz, and E. Metais, editors, *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)*, volume 3513 of *Lecture Notes in Computer Science*, pages 227–238, Alicante, Spain, JUN 2005. Springer.
- [4] T. Mondary, S. Despres, A. Nazarenko, and S. Szulman. Construction d'ontologies à partir de textes : la phase de conceptualisation. In Y. Prié, editor, *19<sup>es</sup> Journées Francophones d'Ingénierie des Connaissances (IC)*, pages 87–98, 18-20 Juin 2008.

<sup>1</sup><http://www-lipn.univ-paris13.fr/~szulman/logi/index.html>