

DaFOE : an Ontology Building Platform From Texts or Thesauri^{*}

Jean Charlet^{1,2}, Sylvie Szulman³, Nathalie Aussenac-Gilles⁴,
Adeline Nazarenko³, Nathalie Hernandez⁴, Nadia Nadah⁵, Éric
Sardet⁶, Jean Delahousse⁷, Valery Teguiak⁶

¹ INSERM UMR_S 872, Eq. 20, Paris
jean.charlet@spim.jussieu.fr

² Assistance Publique – Hôpitaux de Paris, Paris ;

³ LIPN - UMR 7030, Université Paris 13 - CNRS ;

⁴ CNRS/IRIT et Université de Toulouse ;

⁵ Heudiasyc CNRS/UMR 6599, Université de Technologie de Compiègne ;

⁶ LISI-ENSMA et CRITT-Informatique, Poitiers ;

⁷ MONDECA, Paris ;

1 La plateforme DAFOE

La construction d'ontologies à partir de textes fait l'objet d'études depuis plusieurs années dans le domaine de l'ingénierie des ontologies (TERMINAE¹ Aussenac-Gilles *et al.* (2000), Text2Onto Cimiano & Volker (2005)). 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. C'est ce que nous proposons dans la plateforme DaFOE, en même temps que nous cherchons à favoriser l'émergence de tels outils se développant comme greffon.

^{*}ANR TLOG 06 project <<http://dafoe4app.fr>>

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

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 terminologique où les concepts sont organisés et (3) un niveau où l'ontologie est formalisée Charlet *et al.* (2008). Pour valider l'architecture de méta-modélisation proposée, un premier prototype est en cours de réalisation par le LISI. C'est ce prototype dont nous ferons la démonstration.

Mots-clés : Ontologie, construction d'ontologie, bases de données à bases ontologiques, TALN

2 DAFOE : An Ontology building platform

The concept of ontologies, appeared in the nineties, constitute a key point to represent and share the meaning carried 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 propose a platform, DaFOE, which aims at 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, Ontology Engineering, Lexical Ontologies, Ontology-Based Database, Natural Language Processing

Références

- AUSSENAC-GILLES N., BIÉBOW B. & SZULMAN S. (2000). Revisiting ontology design : a methodology based on corpus analysis. In R. DIENG & O. CORBY, Eds., *Knowledge Engineering and Knowledge Management : Methods, Models, and Tools. Proc. of the 12th International Conference, (EKAW'2000)*, LNAI 1937, p. 172–188 : Springer-Verlag.
- CHARLET J., SZULMAN S., PIERRA G., NADAH N., TEGUIAK H. V., AUSSENAC-GILLES N. & NAZARENKO A. (2008). Dafoe : A multimodel and

- multimethod platform for building domain ontologies. In D. BENSLIMANE, Ed., *2^e Journées Francophones sur les Ontologies*, Lyon, France : ACM.
- CIMIANO P. & VOLKER J. (2005). Text2onto - a framework for ontology learning and data-driven change discovery. In A. MONTOYO, R. MUNOZ & E. METAIS, Eds., *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)*, volume 3513 of *Lecture Notes in Computer Science*, p. 227–238, Alicante, Spain : Springer.