

DaFOE: une plateforme pour construire des ontologies à partir de textes et de thésaurus

Jean Charlet^{*,**}, Sylvie Szulman^{***}, Nathalie Aussenac-Gilles^{****}, Adeline Nazarenko^{***}, Nathalie Hernandez^{****}, Nadia Nadah[‡], Éric Sardet^{‡‡}, Jean Delahousse^{‡‡‡}, Valery Teguiak^{‡‡}, Audrey Baneyx^{*}

*INSERM UMR_S 872, Eq. 20, Paris ;

jean.charlet@upmc.fr

** Assistance Publique – Hôpitaux de Paris, Paris ;

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

**** CNRS/IRIT et Université de Toulouse ;

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

‡‡ LISI-ENSMA et CRITT-Informatique, Poitiers ;

‡‡‡ MONDECA, Paris ;

Mots-clef. Ontologie, construction d'ontologie, bases de données à bases ontologiques, TALN.

1 La plateforme DAFOE

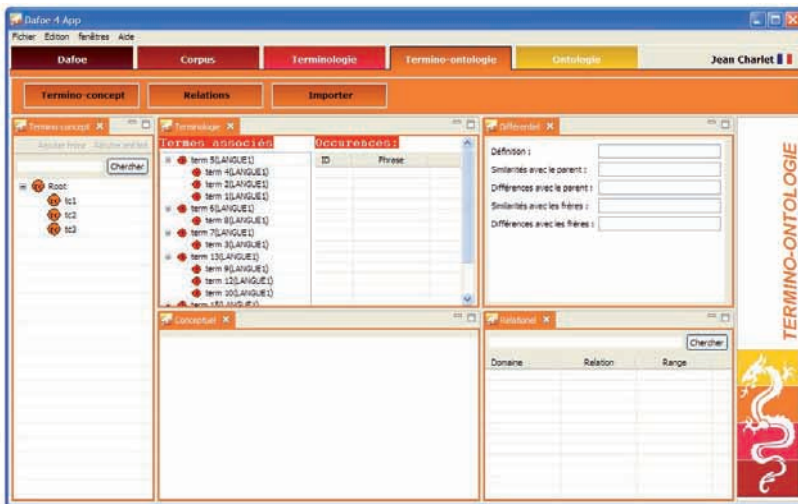
Depuis son émergence, au début des années 1990, dans les recherches en modélisation de connaissances, la notion d'ontologie s'est rapidement diffusée dans un grand nombre de domaines de recherche en informatique.

L'efficacité des applications utilisant les ontologies suppose évidemment l'existence d'une ontologie de domaine susceptible 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, Text2Onto). Ces méthodes sont implémentées dans des outils qui se distinguent par leur approche de la phase de conceptualisation plus ou moins automatique. 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

DaFOE: une plateforme pour construire des ontologies

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 greffons 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 comme 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.



Summary

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.