

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