

AlvisAE, un éditeur d’annotation sémantique structurée guidée par une ontologie évolutive

Robert Bossy, Frédéric Papazian, Claire Nédellec

Unité MIG (Mathématique, Informatique et Génome,
Domaine de Vilvert, 78352 Jouy-en-Josas
{prénom.nom}@jouy.inra.fr

Résumé : De nombreux éditeurs d’annotation sémantique ont été proposés ces dernières années, qui permettent d’associer à des fragments de texte en langage naturel, des catégories ou concepts et des relations préalablement définis dans un schéma formel. Les textes ainsi annotés sont essentiellement utilisés pour nourrir des bases de connaissances formelles ou comme données d’entraînement en extraction d’information (EI). L’éditeur *Cadix* en est un exemple [1]. Le développement de l’annotation sémantique automatique à l’aide d’ontologies a créé un besoin d’éditeurs plus riches, avec des interfaces graphiques intuitives, permettant des annotations collectives et concurrentes, proposant des fonctions de résolution de conflits et de calcul d’accord inter-annotateurs [2, 3, 4, 5]. *AlvisAE* (*Alvis Annotation Editor*) se positionne dans une nouvelle génération d’éditeur d’annotation qui accroît le potentiel applicatif en permettant conjointement l’annotation de texte *et* la révision collective de l’ontologie. Ce type d’outil s’inscrit dans la double perspective d’annotation sémantique et de population d’ontologie. Le texte est vu comme à la fois comme une source de nouveaux concepts pour l’ontologie et l’objet de l’interprétation formelle guidée par l’ontologie. L’utilisation d’AlvisAE dans de nombreux projets d’annotation sémantique dans le domaine des Sciences et la Vie a montré la faisabilité d’applications où l’ontologie est complétée collectivement au fur et à mesure que le texte à annoter révèle de nouveaux concepts. Par exemple, l’annotation des 131 documents de la tâche *Bacteria Biotope* de la *BioNLP Shared Task* en 2013 par l’ontologie *OntoBiotope* (> 1700 concepts) a été réalisée par 7 annotateurs. L’architecture d’application Web permet l’utilisation d’AlvisAE par un navigateur et le partage instantané et collectif des évolutions. Son ergonomie intuitive pour des utilisateurs non informaticiens et la pré-annotation automatique immédiatement visualisable et modifiable ont apporté un gain de temps considérable par rapport à des versions précédentes d’éditeurs. Outre de nouvelles perspectives applicatives, cette proposition ouvre de nouvelles pistes de recherche sur le maintien de la cohérence des annotations et de l’ontologie, d’un point de vue à la fois formel et social.

Mots-clés : *Editeur d’annotations collectives, Annotation sémantique de connaissances, Evolution des ontologies.*

Références

- [1] Cadix. <http://caderige.imag.fr/>
- [2] Widlöcher Antoine et Mathet Yann J. (2009). La plate-forme Glozz : environnement d'annotation et d'exploration de corpus. *TALN*, Senlis, France.
- [3] Ogren Philip V. (2006). Knowtator: A Protégé plug-in for annotated corpus construction. In Proceedings of *HLT-NAACL*.
- [4] Stenetorp Pontus, Pyysalo Sampo, Topić Goran, Ohta Tomoko, Ananiadou Sophia, et Tsujii Jun'ichi. (2012). BRAT: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL '12)*. Association for Computational Linguistics, Stroudsburg, PA, USA, p. 102-107.
- [5] Papazian Frédéric, Bossy Robert et Nédellec Claire, (2012). AlvisAE: a collaborative Web text annotation editor for knowledge acquisition. Proceedings of *The 6th Linguistic Annotation Workshop (The LAW VI)*, Jeju, Korea.