

Une nouvelle version de l'outil Terminae de construction de ressources termino-ontologiques

Sylvie Szulman

LIPN, UMR 7030 CNRS, Université Paris 13
99 Av. Jean Baptiste Clement 93430 Villetaneuse, France
sylvie.szulman@lipn.univ-paris13.fr

Résumé : Cette démonstration présente une nouvelle version de l'outil Terminae intégrable comme greffon dans toute application eclipse. Elle prend en compte les réflexions menées dans les projets auxquels nous avons participé.

Mots-clés : Construction de ressources termino-ontologiques à partir de textes, Application Eclipse, greffons

1. Introduction

La construction de ressources termino-ontologiques à partir de textes s'appuie sur deux principes. Le premier peut s'énoncer comme il existe des textes décrivant un domaine pour une application donnée. Le second est qu'il est possible d'accéder aux éléments de connaissances du domaine en utilisant les éléments terminologiques que des outils de traitement automatique des langues peuvent fournir. Lorsque les ressources termino-ontologiques (RTO) sont construites sur ces principes, elles peuvent être considérées comme fortement couplées à une composante linguistique. Le couplage d'une composante linguistique et d'une RTO sert dans plusieurs types d'application comme la recherche d'information. La plateforme Terminae outille cette construction. Elle a été créée pour mettre en œuvre la méthodologie de même nom décrite dans Aussenac-Gilles *et al.* (2008). La participation du LIPN au projet ANR

Dafoe4App et au projet Européen Ontorule ¹ a entraîné une refonte de l'outil. Après un rappel d'éléments méthodologiques sur lesquels est fondée la plateforme, nous en décrivons les caractéristiques techniques.

2. Elements méthodologiques

La méthode mise en œuvre reste identique à la première version. Cependant plusieurs améliorations ont été apportées. Actuellement, la méthode intervient sur plusieurs niveaux de connaissances (linguistique, terminologique, termino-conceptuelle, conceptuelle) comme proposé dans (Charlet *et al.* (2008)). L'un des éléments clés de la plateforme est le niveau termino-conceptuel. Il s'agit de passer du niveau terminologique au niveau conceptuel. On construit un artefact qui a été appelé "concept terminologique" (Biébow & Szulman (1999)) ou "termino-conceptuel" plus récemment. Pour rappel, un termino-concept est un terme désambiguïsé dont le sens est défini par son usage dans le corpus. Ce niveau fait le lien entre le texte et l'ontologie. Il permet d'associer à une ontologie une composante linguistique.

3. Caractéristiques techniques

La plateforme permet d'intervenir à chaque niveau précédemment cité. Au niveau linguistique, la plateforme permet de filtrer les résultats de l'extracteur terminologique Yatea ⁽²⁾ et de l'extracteur d'entités nommées ANNIE (plateforme Gate ³), d'ajouter des termes, des occurrences.

Au niveau terminologique, la plateforme gère un ensemble d'informations linguistiques associé à chaque terme sélectionné appelé "fiche terminologique" qui comprend les variantes, les syntagmes nominaux dans lesquels le terme est en tête ou en position de modifieur, les relations terminologiques dans lesquelles le terme intervient, les termino-concepts associés.

Au niveau termino-conceptuel, la plateforme gère un réseau de termino-concepts éventuellement organisé en une hiérarchie. A un termino-concept correspond une définition en langage naturel, des termes, des liens avec d'autres termes, des relations termino-conceptuelles dans lesquelles le termino-concept inter-

¹Ce travail a été réalisé en partie pour le projet ONTORULE (FP7 231875)

²<http://search.cpan.org/~thhamon/Lingua-YaTeA-0.5/>

³<http://gate.ac.uk/ie/>

vient. L'ensemble des termino-concepts forme un thesaurus qui peut être importé ou exporté en SKOS.

Le niveau conceptuel ou ontologique est constitué par l'éditeur NeOn ToolKit ⁴. Les termino-concepts peuvent être formalisés par des classes d'une ontologie OWL créée dans l'éditeur NeOn ToolKit. L'ontologie contient les liens vers les termino-concepts sous forme d'annotations.

L'interface utilisateur présente cinq perspectives ⁵ correspondant à chaque niveau auquel s'ajoute une perspective de gestion de projets. La plateforme est développée en java 1.6 comme une application Eclipse. Elle peut être exécutée comme une application ou intégrable comme un greffon dans toute application eclipse.

4. Conclusion

Cette nouvelle plateforme permet de travailler sur les différentes couches de connaissances, de lier des éléments textuels à des éléments ontologiques et vice-versa. Elle est extensible par greffons. Le remplacement de l'éditeur Neon Toolkit par l'éditeur protégé est à l'étude.

Références

- AUSSENAC-GILLES N., DESPRES S. & SZULMAN S. (2008). The TERMINAE Method and Platform for Ontology Engineering from texts. In P. BUITELAAR & P. CIMIANO, Eds., *Bridging the Gap between Text and Knowledge - Selected Contributions to Ontology Learning and Population from Text*, p. 199–223. IOS Press.
- BIÉBOW B. & SZULMAN S. (1999). TERMINAE : A linguistics-based tool for building of a domain ontology. In D. FENSEL & R. STUDER, Eds., *Proc. of the 11th European Workshop (EKAW'99)*, LNAI 1621, p. 49–66 : Springer-Verlag.
- CHARLET J., SZULMAN S., PIERRA G., NADAH N., TEGUIAK H., AUSSENAC GILLES N. & NAZARENKO A. (2008). DAFOE : A Multimodel and Multimethod Platform for Building Domain Ontologies. In *Actes des secondes Journées francophones sur les ontologies Deuxièmes Journées Francophones sur les Ontologies*, p. 66–78, Lyon France.

⁴<http://neon-toolkit.org/>

⁵Une perspective est constituée d'un ensemble de vues. Une vue visualise un objet (liste,texte) ...