

Un système d'aide à l'extraction de relations sémantiques pour la construction d'ontologies à partir de textes

Rim Bentebibel*, Adeline Nazarenko*
Sylvie Szulman*

* Laboratoire d'Informatique de l'université Paris-Nord (LIPN)
UMR 7030 Université Paris 13 & CNRS
99, avenue Jean-Baptiste Clément
93430 Villetaneuse

prénom.nom@lipn.univ-paris13.fr

Résumé. Cet article présente une méthode d'extraction de relations sémantiques pour la construction d'ontologies à partir de corpus de textes. Notre objectif est de proposer une méthode générique, qui soit indépendante du domaine et de la langue. Elle repose sur une analyse distributionnelle des unités sémantiques du corpus pour faire émerger des relations sémantiques candidates. Cette méthode ne fait aucune hypothèse sur les types de relations recherchées ni sur leur forme linguistique. Il s'agit de regrouper les associations de termes dans des classes qui représentent des relations sémantiques candidates. L'hypothèse sous-jacente est que les occurrences de ces associations réunies sur la base des éléments de contexte qu'elles partagent ont des chances de relever d'une même relation sémantique et que les relations candidates ainsi proposées peuvent aider le travail de conceptualisation de l'ontologue.

1 Introduction

Les textes sont des sources précieuses pour la construction d'ontologies parce qu'ils portent la trace de connaissances stabilisées et partagées et qu'ils sont souvent plus faciles d'accès que les experts. Les méthodes de construction d'ontologies à partir de textes sont aujourd'hui bien connues : pour identifier les concepts du domaine, elles s'appuient sur l'analyse terminologique pour les unes, sur l'analyse distributionnelle et les classes de mots pour les autres.

Au-delà des concepts et de leurs instances, il est important de repérer les relations conceptuelles qui structurent le domaine. Des approches distributionnelles ont été proposées, mais pour l'explicitation des relations hiérarchiques uniquement. Les approches classiques, héritées de la terminologie traditionnelle, permettent d'extraire aussi des relations transversales. Elles explorent les textes à l'aide de patrons mais ceux-ci diffèrent pour chaque relation et varient souvent d'un corpus à l'autre. Nous proposons ici une méthode générique de découverte de relations sémantiques à partir de textes. Il s'agit d'explorer les textes pour identifier les relations conceptuelles qu'ils véhiculent sans idée préconçue sur le type de relations qu'on recherche.