

Acquisition de structures lexico-sémantiques à partir de textes : un nouveau cadre de travail fondé sur une structuration prétopologique

Guillaume Cleuziou*, Gaël Dias**, Vincent Levorato*

*LIFO

Université d'Orléans - Orléans, FRANCE

nom.prenom@univ-orleans.fr

**HULTIG

Université de Beira Interior - Covilhã, PORTUGAL

ddg@di.ubi.pt

Résumé. Les structures lexico-sémantiques jouent un rôle essentiel dans les processus de fouille de textes. En codant les relations sémantiques entre concepts du discours elles apportent une connaissance stratégique pour enrichir les capacités de raisonnement. Le développement de telles structures étant fortement limité du fait des efforts nécessaires à leur construction, nous proposons un nouveau formalisme d'acquisition automatique d'ontologies terminologiques à partir de textes. Nous utilisons pour cela une formalisation prétopologique de l'espace des termes sur laquelle s'appuie un modèle générique de structuration. Nous présentons une étude empirique préliminaire rendant compte du potentiel de ce modèle en terme d'extraction de connaissances.

1 Introduction

Les structures lexico-sémantiques jouent un rôle essentiel en Recherche d'Information (RI) et en Traitement du Langage Naturel (TLN). En codant les relations sémantiques entre concepts du discours elles permettent d'enrichir les capacités de raisonnement pour des applications telles que la classification de textes (Hotho et al., 2003), l'expansion de requêtes (Bhogal et al., 2007), la recherche d'information personnalisée (Mylonas et al., 2008) pour n'en citer que quelques unes. Cependant, leur développement est fortement limité en raison des efforts nécessaires à leur construction. Ces dernières années, des études ont été menées dans le but d'apprendre de telles structures à partir des textes et ainsi réduire la quantité de travail d'ingénierie nécessaire à leur élaboration (Cimiano et al., 2009). Apprendre des structures lexico-sémantiques directement à partir des textes plutôt que de les établir manuellement présente des avantages indéniables : d'abord la possibilité d'élaborer une structure pour un domaine spécifique en ajustant le corpus de textes en conséquence ; également la réduction considérable du "coût" pour chaque entrée lexicale, permettant alors d'envisager des structures à large échelle.