

Extraction d'instances de relations n-aires issues d'articles scientifiques guidée par une ontologie: comparaison de méthodes structurelles, fréquentistes et sémantiques.

Lentschat^{1,2}, M.; Buche¹, P.; Dibia-Barthelemy³, J.; Menut, L.; Roche², M.

1-UMR IATE Montpellier, 2-UMR TETIS Montpellier, 3-UMR MIA Paris

Résumé :

L'extraction d'instances de relations n-aires depuis des articles scientifiques est une question complexe nécessitant de s'adapter aux spécificités de la tâche et du domaine. Nous proposons une approche guidée par une ontologie fondée sur l'utilisation des informations contenues dans les tableaux des documents, extraits sous forme d'instances de relations n-aires partielles, et visons à les compléter à l'aide d'instances d'arguments reconnues dans le texte des documents. Trois méthodes sont évaluées et comparées : (1) une méthode utilisant la structure des documents ; (2) une méthode fréquentiste qui s'appuie sur l'analyse des cooccurrences entre instances d'arguments du texte et arguments déjà renseignés dans les relations partielles issues des tableaux ; (3) une méthode sémantique utilisant des modèles de langage *word embedding* pour déterminer la similarité entre les instances d'arguments du texte et les instances d'arguments renseignés dans les relations partielles.