

Reconnaissance d'entités nommées itérative sur une structure en dépendances syntaxiques avec l'ontologie NERD

Cédric Lopez, Melissa Mekaoui, Kevin Aubry, Jean Bort, Philippe Garnier

Emvista

<https://www.emvista.com>

Cap Oméga, Rond-point Benjamin Franklin, 34960 Montpellier

prenom.nom@emvista.com

Résumé. La reconnaissance des entités nommées (REN) consiste à repérer des éléments textuels et à les classer dans des catégories prédéfinies (noms de personnes, d'organisations, de marques, d'équipes sportives, *etc.*). La REN est souvent considérée comme l'une des briques de fondation des systèmes visant à structurer un texte tout-venant. Dans cet article, nous décrivons notre système symbolique de REN qui se caractérise par 1) l'utilisation de ressources dictionnaires limitées et 2) la prise en compte de résultats provenant d'autres briques telles que la résolution de coréférences et l'extraction de relations. Le système est basé sur la sortie d'un analyseur syntaxique en dépendances qui adopte un flot d'exécution itératif intégrant des résultats d'autres briques d'analyse. À chaque itération, des catégories candidates sont générées et sont toutes prises en compte dans les itérations suivantes. L'intérêt d'un tel système est de sélectionner définitivement le meilleur candidat uniquement à la fin du traitement afin de tenir compte de l'ensemble des éléments fournis par les différentes briques. Le système est comparé à des systèmes académiques et industriels.

1 Introduction

La reconnaissance des entités nommées (REN) dans un texte est une tâche consistant à repérer des éléments textuels et à les classer dans des catégories/types prédéfinies (*i.e.* noms de personnes, d'organisations, de marques, d'équipes sportives, *etc.*). La REN est souvent considérée comme l'une des briques de fondation des systèmes visant à structurer un texte tout-venant. Cette brique est généralement conçue de façon indépendante des autres briques du Traitement Automatique du Langage Naturel (TALN) et s'inscrit dans un flot d'exécution linéaire : les briques situées après la REN ne peuvent plus intervenir sur la reconnaissance des entités nommées, et celles situées avant ne peuvent bénéficier des résultats de la REN. Par exemple, la résolution des coréférences et l'extraction des relations sont généralement situées après la REN ce qui impose *de facto* des limites aux systèmes. Soient trois exemples :

- (1) Paris visite Paris.
- (2) Paris est triste. Elle pleure dans le salon.