

Peuplement de base de connaissances, évaluation et système end-to-end

Maxime Prieur^{*,**}, Cédric du Mouza^{*}, Guillaume Gadek^{**}, Bruno Grilheres^{**}

^{*} Laboratoire Cédric, Conservatoire National des Arts et Métiers
maxime.prieur.auditeur@lecnam.net, dumouza@cnam.fr

^{**} Airbus Defence and Space, Élancourt France
guillaume.gadek, bruno.grilheres@airbus.com

Résumé. Les bases de connaissances (KB) sont utilisées dans de nombreux domaines, comme l'intelligence économique ou l'assistance aux utilisateurs. Elles regroupent des connaissances pouvant être exploitées par l'ordinateur. Cependant, leur création reste complexe pour un expert devant extraire et lier chaque nouvelle information. Dans cet article, nous décrivons une méthode d'évaluation de système d'enrichissement de base de connaissances. Cette évaluation est mise à l'épreuve par ELROND un système complet conçu autour d'une chaîne de traitement composée de 4 modules.

1 Introduction

Les bases de connaissances sont des structures de données régies par des ontologies pré-définies et sont très utiles pour agréger l'information dans le but de simplifier sa visualisation et son analyse. Ces bases sont utilisées dans de multiples domaines tels que les métiers de l'intelligence économique (Shue et al., 2009) afin de faciliter la prise de décision ou bien pour extraire les éléments liant des publications scientifiques (Luan et al., 2018) par exemple.

Cependant, la construction et la mise à jour manuelles de ces bases sont coûteuses puisque souvent les domaines d'utilisation exploitent des informations en constante évolution. L'alternative, un enrichissement automatique, permettrait d'extraire et d'ajouter les éléments souhaités depuis des sources d'intérêts vers la structure de données. Néanmoins, les solutions existantes sont encore limitées et il n'existe à notre connaissance aucun protocole d'évaluation couvrant l'entièreté de la tâche de peuplement (Min et al., 2018; Mesquita et al., 2019).

Dans cet article, nous tentons de combler ce manque en proposant les contributions suivantes :

- Nous formalisons une méthode d'évaluation des systèmes complets d'enrichissement de base de connaissances depuis des textes.
- Nous introduisons ELROND, un système de peuplement end-to-end en 4 étapes sur lequel nous testons notre méthode d'évaluation.
- Nous évaluons et comparons le modèle proposé en appliquant le mode d'évaluation présenté au dataset DWIE (Zaporjets et al., 2021).

La suite de cet article est structurée comme suit : après avoir présenté les travaux récents sur les systèmes d'enrichissement de base de connaissances (Knowledge Base Population, KBP)