

Construction d'ontologies à partir de textes : démonstration d'une approche basée sur l'analyse de graphes AMR

Aurélien Lamerrierie*, David Rouquet*

* Tétras-Libre, 8 Rue Mayencin, 38410 St Martin d'Hères
aurelien.lamerrierie@tetras-libre.fr, david.rouquet@tetras-libre.fr,
<https://www.tetras-libre.fr>

Résumé. Cet article présente le prolongement de travaux autour d'un outil permettant la construction automatique d'ontologies OWL à partir de textes exprimés en langue naturelle. La chaîne de traitement mise en oeuvre part d'énoncés non contraints pour aboutir à une structure logique encodant la connaissance extraite. Elle passe par deux phases majeures : (1) une sérialisation RDF de graphes AMR exploités comme représentation pivot des énoncés, et (2) l'extraction du contenu logique par l'analyse des représentations intermédiaires. L'implémentation est basée sur les standards du Web sémantique.

1 Introduction

Nos travaux portent sur la construction automatique d'ontologies à partir de textes en langue naturelle. Ils s'inscrivent dans la continuité du projet UNseL¹ dont l'objectif était la vérification automatisée d'exigences système, à partir d'inférences sur une ontologie extraite automatiquement. Les premiers résultats ont été présentés lors de la conférence EGC 2022 (Rouquet et al. (2022)). Cet article montre le prolongement de ces travaux avec une nouvelle démonstration traitant d'articles encyclopédiques. Initialement basé sur une sérialisation RDF de structures UNL², notre prototype exploite dorénavant et également des graphes AMR³.

Les ontologies sont des modèles de données décrivant un domaine, sous la forme d'ensembles structurés de concepts et de relations. Elles sont utiles pour de nombreuses applications, comme par exemple les systèmes de questions/réponses ou d'aide à la décision. Leur construction est un enjeu important. Ainsi, l'état de l'art⁴ présente plusieurs travaux intéressants s'appuyant sur des techniques variées. Quelques méthodes de pointe semblent déjà suffisantes pour certaines applications pratiques, telles que la classification de documents ou la recherche d'informations. Néanmoins, celles-ci visent à générer des ontologies faiblement contraintes au niveau logique, alors que de nombreuses applications basées sur du raisonnement reposent sur des axiomatisations plus complexes. Il n'existe pas de méthode éprouvée pouvant servir de référence dans ce domaine.

1. *Universal Networking system engineering Language*

2. *Universal Networking Language*, Uchida et al. (1996)

3. *Abstract Meaning Representation*, Banarescu et al. (2013)

4. Voir, par exemple, Khadir et al. (2021).