

Laboratoire d'accueil : Équipe EID du LIASD, IUT de Montreuil, 140 Rue de la Nouvelle France, 93100 Montreuil, France

Date de début : Février / Mars 2026

Durée : 4 - 5 mois

Mots clés : Intelligence artificielle symbolique, LLM, ontologies, Raisonnement OWL, cohérence sémantique, ingénierie des Connaissances

Contexte

Les grands modèles de langage (LLM) tels que GPT, Llama ou Mistral sont capables de raisonner sur du texte, d'extraire des entités et d'établir des relations sémantiques complexes. Leur utilisation pour l'**enrichissement automatique d'ontologies** est un domaine émergent en ingénierie des connaissances.

Toutefois, les informations produites par un LLM ne respectent pas nécessairement la **cohérence logique** ni la **structure formelle** d'une ontologie exprimée en OWL. Ce stage s'inscrit dans une démarche de recherche visant à concevoir une approche **neuro-symbolique**, combinant :

- La créativité sémantique d'un LLM pour proposer de nouveaux concepts et propriétés.
- La rigueur du raisonnement logique via un moteur de raisonnement OWL permet de valider ces propositions.

L'objectif est de développer un cadre d'**enrichissement automatique** d'ontologies à partir de textes scientifiques, de notices techniques ou de graphes RDF existants.

Objectifs du stage

Le/la stagiaire contribuera à la définition et à la mise en œuvre d'une approche d'enrichissement d'ontologie combinant LLM et raisonnement symbolique. Les objectifs principaux sont :

1. Étudier l'état de l'art sur les approches d'enrichissement ontologique, les LLM et le raisonnement neuro-symbolique.
2. Concevoir un pipeline intégrant :
 - Extraction automatique de candidats à partir de textes (concepts, relations, propriétés).
 - Génération d'axiomes OWL par un LLM (domain, range, hiérarchie).
 - Vérification de la cohérence à l'aide d'un raisonneur symbolique (comme Pellet ou SHACL).
3. Implémenter un **prototype Python** intégrant ces modules.
4. Évaluer la qualité des enrichissements sur une ou plusieurs ontologies de référence (ex. domaine des matériaux, de la santé, du bâtiment, etc.).

Profil et compétences

Etudiant · e en 2e année de master ou en dernière année d'école d'ingénieur dans le domaine de l'intelligence artificielle, avec des compétences :

Techniques :

- Programmation : Python (rdflib, transformers, torch).
- Web sémantique : RDF, OWL, SPARQL, Protégé.
- NLP / LLMs : prompts, pipelines, fine-tuning (un plus).
- Raisonnement symbolique, logiques des descriptions, utilisation de reasoners.

Scientifiques :

- Compréhension des modèles LLM.
- Conception d'ontologies et validation sémantique.

Livrables

- Prototype Python fonctionnel d'enrichissement neuro-symbolique.
- Rapport de stage (méthodologie, expériences, perspectives).
- Une contribution potentielle à une publication pour une conférence.

Contact

- Thamer MECHARNIA*, t.mecharnia@iut.univ-paris8.fr
- Islam BARCHOUCH*, i.barchouch@iut.univ-paris8.fr
- Jinfeng ZHONG*, j.zhong@iut.univ-paris8.fr
- Sihem BELABBES*, s.belabbes@iut.univ-paris8.fr

* Équipe EID – laboratoire LIASD, IUT de Montreuil (Université Paris 8)

Références

- [Iyer et al., 2019] Iyer, V., Mohan, L., Bhatia, M., and Reddy, Y. R. (2019). A survey on ontology enrichment from text. In Sharma, D. M. and Bhattacharya, P., editors, *Proceedings of the 16th International Conference on Natural Language Processing*, pages 95–104, International Institute of Information Technology, Hyderabad, India. NLP Association of India.
- [Kollapally et al., 2025] Kollapally, N. M., Geller, J., Keloth, V. K., He, Z., and Xu, J. (2025). Ontology enrichment using a large language model : Applying lexical, semantic, and knowledge network-based similarity for concept placement. *Journal of Biomedical Informatics*, page 104865.
- [Li et al., 2025] Li, J., Garijo, D., and Poveda-Villalón, M. (2025). Large language models for ontology engineering : A systematic literature review.
- [Mateiu and Groza, 2023] Mateiu, P. and Groza, A. (2023). Ontology engineering with large language models. In *2023 25th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, pages 226–229. IEEE.
- [Polleres et al., 2021] Polleres, A., De Meester, B., Heyvaert, P., Arndt, D., Dimou, A., and Verborgh, R. (2021). Rdf graph validation using rule-based reasoning. *Semantic Web (1570-0844)*, 12(1).