

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

**Date de début :** Mars 2026

**Durée :** 4 - 5 mois

**Mots clés :** SPARQL, Web sémantique, Graphes de connaissances, LLMs, Correction de requêtes.

## Contexte

Les graphes de connaissances constituent une infrastructure centrale pour la représentation et l'interrogation de données structurées dans de nombreux domaines (Web de données, bases de connaissances ouvertes, systèmes d'aide à la décision). Le langage SPARQL permet d'interroger ces graphes de manière précise, mais sa maîtrise reste complexe, tant pour les utilisateurs humains que pour les systèmes automatisés.

Dans les systèmes récents de traduction du langage naturel vers SPARQL (Text-to-SPARQL), les modèles de langage (LLMs) sont de plus en plus utilisés. Toutefois, ces modèles génèrent fréquemment des requêtes SPARQL erronées, soit sur le plan syntaxique, soit plus souvent sur le plan sémantique (mauvais prédicats/propriétés, incohérences avec l'ontologie).

La correction automatique de requêtes SPARQL constitue donc un enjeu majeur pour améliorer la fiabilité, la robustesse et l'explicabilité des LLMs en les combinant avec des schémas ontologiques.

## Objectifs du stage

L'objectif principal du stage est de concevoir, implémenter et évaluer un système de correction automatique de requêtes SPARQL en s'appuyant sur les capacités des modèles de langage, tout en intégrant des connaissances symboliques issues du schéma du graphe (ontologie, contraintes).

Plus précisément, le stage vise à :

- Analyser les erreurs fréquentes dans les requêtes SPARQL.
- Exploiter les LLMs pour proposer des corrections de requêtes erronées.
- Guider la correction par des informations structurées (schéma RDF/OWL, contraintes, exemples valides).
- Comparer différentes stratégies de correction :
  - LLM seul.
  - LLM guidé par ontologie, approche hybride de type RAG symbolique.
- Évaluer l'impact de la correction sur l'exécution effective des requêtes et la qualité des réponses.

## 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 : SPARQL, RDF, OWL, Protégé.
- LLMs : prompts, fine-tuning.

### Scientifiques :

- Compréhension des modèles LLM.
- Capacité d'analyse, rigueur scientifique.

## Livrables

- Un prototype fonctionnel de correction automatique de requêtes SPARQL.
- Rapport de stage (état de l'art, méthodologie, expériences, perspectives).
- Une contribution potentielle à une publication pour une conférence.

## Candidature

Envoyer CV, lettre de motivation, lettre(s) de recommandation et relevés de notes à :

- Thamer MECHARNIA\*, [t.mecharnia@iut.univ-paris8.fr](mailto:t.mecharnia@iut.univ-paris8.fr)
- Sihem BELABBES\*, [s.belabbes@iut.univ-paris8.fr](mailto:s.belabbes@iut.univ-paris8.fr)

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

## Références

- [Almendros-Jiménez et al., 2017] Almendros-Jiménez, J. M., Becerra-Terón, A., and Cuzzocrea, A. (2017). Detecting and diagnosing syntactic and semantic errors in sparql queries. In *EDBT/ICDT Workshops*.
- [Jian et al., 2025] Jian, X., Dong, Z., and Özsü, M. T. (2025). Interacsparql : An interactive system for sparql query refinement using natural language explanations. *ArXiv*, abs/2511.02002.
- [Qi et al., 2024] Qi, J., Su, C., Guo, Z., Wu, L., Shen, Z., Fu, L., Wang, X., and Zhou, C. (2024). Enhancing sparql query generation for knowledge base question answering systems by learning to correct triplets. *Applied Sciences*.
- [Rademaker et al., 2024] Rademaker, A., Lima, G., Fiorini, S. R., and da Silva, V. T. (2024). Deductive verification of llm generated sparql queries. In *DLNLD*.
- [Reddy and Kumar, 2010] Reddy, B. R. K. and Kumar, P. (2010). Efficient approximate sparql querying of web of linked data. In *Uncertainty Reasoning for the Semantic Web*.
- [Vossebeld and Wang, 2025] Vossebeld, F. and Wang, S. (2025). Learning to refine : An agentic rl approach for iterative sparql query construction. *ArXiv*, abs/2511.11770.