

Thesis Offer

Research Unit: Distributed Knowledge and Artificial Intelligence Laboratory (CIAD) - UR7533

Location: University Bourgogne Europe – Dijon

Contract Duration: 36 months

Salary: Based on profile and experience

Start Date: September 2025

Application Deadline: March 28, 2025

Scientific Contacts: Ouassila Labbani Narsis – ouassila.narsis@u-bourgogne.fr

Aurélie Bertaux – aurelie.beraux@u-bourgogne.fr

Ontology-Guided RAG Approach for Improving Semantic Precision of LLMs

Context

Large language models (LLMs) have demonstrated impressive capabilities in content generation and query processing. However, they still face limitations in terms of semantic precision, particularly when handling specific, structured, and highly contextualized information.

Retrieval-Augmented Generation (RAG) complements large-scale generative models with information from external sources to enhance the quality and relevance of generated responses. However, simple retrieval of relevant external documents to complement queries may lack contextualization or semantic coherence, especially in specific domains requiring domain expertise and contextual constraints.

This thesis proposes a novel RAG approach guided by the semantics and specific constraints of the domain, aiming to generate responses that are both relevant and semantically valid. The goal is to use ontology to semantically structure knowledge and generate new knowledge through inference processes. This will guide both information retrieval and response generation within RAG architecture.

An ontology is a formal representation of knowledge of domain, organizing concepts and their relationships. It brings a richer semantic dimension to document retrieval and response generation, ensuring greater coherence and relevance. By integrating ontology at the core of the RAG pipeline, this approach aims to overcome current LLM limitations, particularly in specialized fields requiring nuanced and contextualized understanding. The combination of ontology with advanced AI models, such as RAG and LLMs, allows smarter AI systems capable of better understanding context and semantics, delivering more accurate, coherent, and informed responses. This approach can be applied to various fields, such as education, document research (filtering and labeling documents based on semantic relevance), and professions requiring fine-grained understanding of context, business requirements and domain constraints.

Thesis Objectives

- **Literature Review:** Analyze existing approaches and propose scientific positioning.
- **Ontology Development:** Create or adapt an ontology based on predefined use cases in collaboration with domain experts to structure knowledge and facilitate semantically relevant information retrieval.
- **Ontology Integration in RAG Process:** Propose an innovative architecture combining RAG and ontological reasoning to filter and/or enrich retrieved information based on semantic relevance.
- **Semantic Document Chunking:** Segment available documents into semantically coherent chunks focused on a limited number of concepts. These chunks will be labeled according to the concepts they contain and integrated into a knowledge graph.
- **Chunk Identification:** Queries will be labeled similarly to chunks, and the search for chunks with semantic labels closely matching the query will enhance it.
- **LLM Integration:** Use the proposed architecture (ontology + RAG) to constrain the LLM's response generation, ensuring responses are not only relevant but also coherent with the semantic relationships and context defined in the ontology.
- **Performance Evaluation:** Apply the solution to predefined use cases and compare the results of ontology-guided RAG systems with classic RAG systems and LLMs, evaluating semantic precision, coherence, and relevance of responses.

Keywords

Semantic modeling, Ontology, RAG, LLM, semantic chunking, knowledge graph, labeling, semantic distance.

Required Skills

We are seeking a candidate:

- Holding a research master's degree or equivalent in computer science
- With expertise in ontology and LLMs
- Accustomed to collaborative research work
- With a strong command of English (reading, writing, and speaking)

Interested candidates should send their application (CV, cover letter, academic transcripts, and recommendation letter) to the scientific contacts by **March 28, 2025**.