

Large Language Models for Ontology Engineering: From Taxonomies to Axiomatized Ontologies

Project description

Ontologies and knowledge graphs are core Semantic Web technologies for structured knowledge representation, making knowledge digital and machine-readable, and empowering applications such as search engines and scientific data platforms. They are traditionally created by Semantic Web professionals in close collaboration with domain experts, a process that is rigorous but time-consuming and difficult to scale. Recent advances in AI, particularly large language models (LLMs), have shown the capability to support several of these tasks, such as term extraction, class definition, and serialization. A growing body of research has explored this direction in recent years, yet many fundamental questions remain open.

Extensive experience in developing highly standardized and interoperable ontologies has led to a well-defined understanding of how AI can best support ontology engineering. Automating this workflow is a substantial undertaking comprising many interdependent tasks. As a first step, we have already developed an approach (LLM4BFO) that extracts terms from textual sources and organizes them into taxonomies aligned with a top-level ontology.

This master's thesis project addresses the next stage of the pipeline: transforming automatically generated taxonomies into full-fledged ontologies. The student will design, implement, and evaluate the LLM pipeline that enriches taxonomies by reusing properties defined in existing ontologies (e.g., object and data properties from top- and mid-level ontologies) and by designing logical axioms (e.g., class restrictions and disjointness) that capture domain semantics. The agent will be integrated into our existing ontology engineering pipeline, and its output will be evaluated against human-engineered ontologies with respect to property reuse, axiom correctness, and logical consistency.



Tasks

- Review state-of-the-art literature on LLM-based ontology and knowledge graph engineering.
- Design an AI agent architecture for transforming taxonomies into axiomatized ontologies. Implement the agent using LLMs and semantic web technologies.
- Evaluate the agent's output against gold-standard ontologies and quality metrics.

This thesis will be supervised by **Prof. Dr. Harald Sack, Information Service Engineering Institute AIFB, KIT, in collaboration with FIZ Karlsruhe.**

Which prerequisites should you have?

- Good programming skills in Python or Java
- Interest in Natural Language Processing
- Interest in Semantic Web technologies
- Team player with good communication skills

Contact persons:

Dr. Hossein Beygi Nasrabadi

Hossein.Beygi_Nasrabadi@fiz-Karlsruhe.de

