

Ontology Expansion via Synthetic Dataset Generation and Transformer-Based Concept Extraction

Authors : Andrey Khalov

Abstract : The rapid proliferation of unstructured data in IT infrastructure management demands innovative approaches for extracting actionable knowledge. This paper presents a framework for ontology-based knowledge extraction that combines relational graph neural networks (R-GNN) with large language models (LLMs). The proposed method leverages the DOLCE framework as the foundational ontology, extending it with concepts from ITSMO for domain-specific applications in IT service management and outsourcing. A key component of this research is the use of transformer-based models, such as DeBERTa-v3-large, for automatic entity and relationship extraction from unstructured texts. Furthermore, the paper explores how transfer learning techniques can be applied to fine-tune large language models (LLaMA) for using to generate synthetic datasets to improve precision in BERT-based entity recognition and ontology alignment. The resulting IT Ontology (ITO) serves as a comprehensive knowledge base that integrates domain-specific insights from ITIL processes, enabling more efficient decision-making. Experimental results demonstrate significant improvements in knowledge extraction and relationship mapping, offering a cutting-edge solution for enhancing cognitive computing in IT service environments.

Keywords : ontology expansion, synthetic dataset, transformer fine-tuning, concept extraction, DOLCE, BERT, taxonomy, LLM, NER

Conference Title : ICSTO 2025 : International Conference on Semantic Technology and Ontology

Conference Location : Barcelona, Spain

Conference Dates : May 24-25, 2025