

KGNode: Path-Aware Subgraph Extraction for Training-Free Question Answering over RDF Knowledge Graphs

Authors:

A F M Mohimenul Joaa, Faerber Michael

Question answering over large RDF knowledge graphs requires selecting a compact, question-relevant part of a graph before a reliable SPARQL query can be generated. Existing approaches either rely on KG- or benchmark-specific training, such as learned entity linkers, retrievers, or KG embeddings, or explore the graph with limited question awareness, for example, through fixed-depth BFS or topology-driven random walks. We present KGNode, a training-free framework for path-aware subgraph extraction in knowledge-graph question answering. KGNode combines hybrid seed discovery, which fuses semantic and keyword search with type-aware filtering, with path-history-guided priority traversal, which ranks candidate graph extensions by comparing complete accumulated paths to a question-derived traversal template. The resulting subgraph serves as inspectable evidence for downstream SPARQL generation.

We evaluate KGNode on DBLP-QuAD over DBLP and QALD-10 over Wikidata. Without KG-specific fine-tuning or task-specific training data, KGNode achieves 70%-76% end-to-end accuracy on DBLP-QuAD across four LLM backends (gpt-4o-mini, gpt-4.1, gpt-5.4-mini, gpt-5-mini) and 69.0%-79.2% on QALD-10 using gpt-4o-mini, gpt-5.4-mini, and gpt-4.1. Component-level analysis shows robust seed discovery; remaining errors stem mainly from SPARQL generation on DBLP-QuAD, with relation retrieval coverage an additional bottleneck on QALD-10. These results indicate that training-free, path-aware subgraph extraction is a viable retrieval layer for portable question answering over heterogeneous RDF knowledge graphs.