Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Doctoral Consortium Track

Ontology-Aware LLMs for Human-Usable Knowledge Graph Systems

Luiz do Valle Miranda
Department of Human-Centered Artificial Intelligence, Institute of Applied Computer Science, Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, prof. Stanisława Łojasiewicza 11, 30-348 Kraków, Poland
luiz.miranda@uj.edu.pl

Abstract

Knowledge graphs (KGs) promise machine-readable semantics together with human-readable labels and relations, yet domain experts struggle to use them due to technical barriers, especially in automatically constructed graphs with unlabeled nodes and complex connections. While LLMs enable KG interfaces such as retrieval-augmented generation, their potential to leverage ontology definitions to improve graph readability remains unexplored. This research investigates LLMs as ontology-aware verbalization interfaces that generate human-readable and usable textual information encoded in the KG structure.

1 Problem, Motivation and State of the Art

Knowledge graphs (KGs) provide explicit semantic structure, support interoperability, and enable symbolic reasoning [Hogan *et al.*, 2021]. However, their practical use by non-technical users remains limited. Understanding and exploring ontology-based RDF graphs often requires expertise in SPARQL, graph algorithms, and ontology engineering [Hogan *et al.*, 2021; DeLong *et al.*, 2025]. As a result, there is a gap between the formal expressiveness of KG and their effective usability to domain experts and end users.

This thesis is motivated by two recurring obstacles to human-centered use of KG. First, some semantically important nodes introduced during transformation pipelines do not have human-readable labels, even though they represent meaningful events or reified structures in the graph [Sequeda *et al.*, 2012]. Second, even when entities are labeled, the complete set of meaningful connections between them may remain difficult to discover, as path-based exploration requires technical query-writing skills or familiarity with graph search methods [DeLong *et al.*, 2025]. Therefore, I investigate whether large language models (LLMs) can help translate graph semantics and structure into natural language grounded in local graph context and ontology definitions.

LLMs have recently been integrated with KGs for tasks such as text-to-SPARQL translation, retrieval-augmented generation, and ontology mapping [Sequeda *et al.*, 2025]. EntitySum, e.g., combines structural triple ranking with LLM-based text generation for entity summarization [Trouli *et al.*, 2026]. Path-based reasoning methods compute paths between entities and verbalizing them for explanation or multi-hop reasoning [Ghafarollahi and Buehler, 2025]. However, most approaches treat KGs primarily as retrieval indices or focus on individual reasoning traces, while the context and generalization allowed by the usage of ontologies remains largely unexplored [Hogan *et al.*, 2021].

The addition of ontological knowledge to KG-LLM pipelines for improving KG usability is supported by two preliminary studies[1]. The first investigates node-level readability through automatic label generation for unlabeled event nodes in a CIDOC-CRM cultural heritage KG using a LLM Generator–Judge pipeline; preliminary results show viable model configurations achieve high coverage and semantic adequacy, though performance varies significantly by model and validation design. The second investigates the readability at the relation-level through OntoPathVerbalizer, which classifies paths by ontology schema patterns and generates reports of all different types of connection between entities. Together, these studies support the thesis that the combination of LLMs and ontologies can make graph semantics more legible to humans beyond standard question-answering interfaces.

2 Research Questions and Expected Contributions

The research here described is based on the following research questions:

- **RQ1**: Can LLMs generate appropriate labels for KG nodes using graph context and ontology definitions?
- **RQ2**: Can LLMs comprehensively verbalize semantically distinct connections between two KG entities by leveraging ontology-guided path classification?
- **RQ3**: To what extent can ontology-aware verbalization improve access to KGs beyond existing paradigms such as text-to-SPARQL and GraphRAG?

Fig. 1 presents a visualization of the proposed ontology-aware LLM pipeline. The expected contributions of this thesis are threefold: a method for ontology-aware generation of labels for unlabeled RDF nodes, a framework for verbalizing all semantically distinct connections between entities, and a

[1]Papers describing such studies are currently under review, upon acceptance will be included in cmera-ready version.

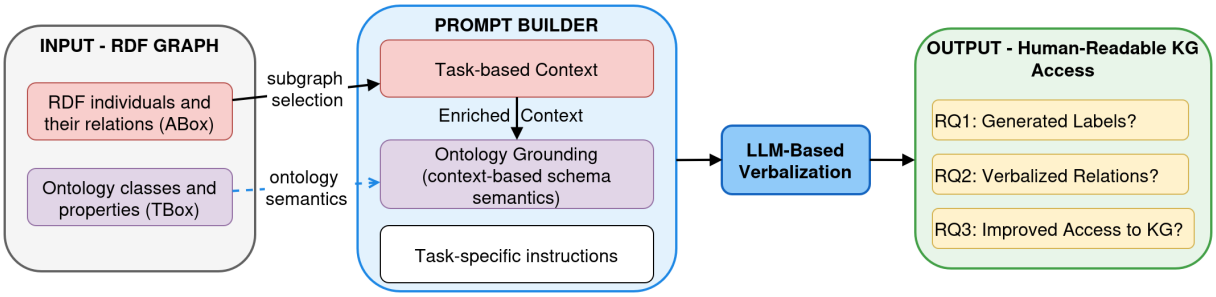

Figure 1: Ontology-aware LLM pipeline for human-readable knowledge graphs.

broader account of ontology-aware verbalization as a complementary LLM+KG paradigm that improves interpretability and access to knowledge encoded in KG structure.

3 Methodology, Impact and Limitations

The central domain of this thesis is cultural heritage, however pilot studies are planned on the fields of medicine and law. All research questions are addressed through the design and experimental evaluation of ontology-aware verbalization pipelines, followed by qualitative and quantitative assessment of their faithfulness, readability, and usefulness for improving human access to KG semantics.

Evaluation combines three forms of assessment: general users, domain experts, and LLM-based evaluation. For label generation, evaluators assess coverage, semantic adequacy, readability, and consistency, including comparison with existing labels where available. General users will assess readability and clarity, domain experts will assess semantic correctness and domain appropriateness, and LLM-based evaluation will provide an additional scalable signal of faithfulness and consistency. For path verbalization, evaluation is based on faithfulness to the graph, completeness with respect to the discovered path classes, understandability to non-technical users, and usability for exploratory analysis.

This research aims to make KGs easier to understand and use beyond technical expert communities. It may impact areas such as cultural heritage, digital humanities, biomedicine, and enterprise settings, where important knowledge is often available but difficult to access in practice. By making graph content more understandable, the work could help a wider range of users benefit from structured knowledge without requiring advanced technical expertise.

Existing KGs to be used do not contain sensitive personal information. Non-sensitive data will be collected from annotators. To address ethical concerns, the study emphasizes responsible data handling, validation mechanisms, human evaluation, and transparency in methods and code. Risks related to LLM use, such as hallucination and over-interpretation, will be mitigated through these measures.

Acknowledgments

This publication was funded by a flagship project "CHExRISH: Cultural Heritage Exploration and Retrieval with Intelligent Systems at Jagiellonian University" under the Strategic Programme Excellence Initiative at Jagiellonian University. The research for this publication has been supported by a grant from the Priority Research Area DigiWorld under the Strategic Programme Excellence Initiative at Jagiellonian University. The author used GPT-5.2 for: literature review, increase readability, and diagram generation. The authors reviewed the content and take full responsibility for it.

References

[DeLong *et al.*, 2025] Lauren Nicole DeLong, Ramon Fernández Mir, and Jacques D. Fleuriot. Neurosymbolic ai for reasoning over knowledge graphs: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 36(5):7822–7842, 2025.

[Ghafarollahi and Buehler, 2025] Alireza Ghafarollahi and Markus J. Buehler. Sciagents: Automating scientific discovery through bioinspired multi-agent intelligent graph reasoning. *Advanced Materials*, 37(22):2413523, 2025.

[Hogan *et al.*, 2021] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia D'amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge graphs. *ACM Comput. Surv.*, 54(4), July 2021.

[Sequeda *et al.*, 2012] Juan F. Sequeda, Marcelo Arenas, and Daniel P. Miranker. On directly mapping relational databases to rdf and owl. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, page 649–658, New York, NY, USA, 2012. Association for Computing Machinery.

[Sequeda *et al.*, 2025] Juan Sequeda, Dean Allemang, and Bryon Jacob. Knowledge graphs as a source of trust for llm-powered enterprise question answering. *Journal of Web Semantics*, 85:100858, 2025.

[Trouli *et al.*, 2026] Georgia Eirini Trouli, Giannis Vassiliou, Iakovos Giannakos, Nikos Papadakis, and Haridimos Kondylakis. Entitysum: Entity summarization using centrality measures. In Edward Curry, Valentina Presutti, John McCrae, Mehwish Alam, Pieter Colpaert, Josiane Xavier Parreira, Diego Collarana, Marta Sabou, Andreas Harth, and Pasquale Lisena, editors, *The Semantic Web: ESWC 2025 Satellite Events*, pages 29–33, Cham, 2026. Springer Nature Switzerland.