An Interactive System for SPARQL Query Refinement Using Natural Language Explanations

Xiangru Jian, Zhengyuan Dong, M. Tamer Özsu

University of Waterloo
{xiangru.jian, zhengyuan.dong, tamer.ozsu}@uwaterloo.ca

1 Introduction

Querying semantic web data using SPARQL remains challenging for non-expert users due to complex syntax and intricate RDF structures [4, 3, 2]. Natural language to SPARQL (NL2SPARQL) systems offer convenient interfaces but suffer from two critical limitations: low accuracy and lack of transparency, preventing users from understanding or correcting errors. While interactive tools exist [1, 5], they support limited SPARQL features and lack systematic explanations.

We present InteracSPARQL, a training-free add-on system designed to enhance any SPARQL generation tool by improving both accuracy and transparency. Unlike systems requiring intensive model training or dataset-specific tuning, InteracSPARQL works as a post-generation refinement layer that seamlessly integrates with existing NL2SPARQL approaches. The system addresses the core challenge that most query generation errors stem from incorrect entity or property IRIs rather than syntactic incompetence. Our approach combines two key innovations: (1) a two-stage natural language explanation framework transforming SPARQL queries into human-readable descriptions through rule-based AST analysis and LLM refinement, and (2) an interactive refinement loop enabling iterative improvements with dynamic entity/property search, resolving domain-specific ambiguities on demand. The system supports full SPARQL 1.1 across any knowledge graph without requiring training or adaptation. Evaluation on QALD benchmarks demonstrates substantial improvements, with human studies confirming enhanced explanation quality. By serving as a universal, training-free add-on, InteracSPARQL provides a practical path to improving both reliability and interpretability of SPARQL generation across diverse applications.

2 Talk Content

This presentation explores InteracSPARQL as a universal, training-free add-on for SPARQL generation systems. We establish the motivation by demonstrating that most generation failures stem from incorrect entity/property IRIs rather than syntactic errors, positioning our approach as a post-generation enhancement layer requiring no retraining. We cover the two-stage NLE framework, which combines rule-based AST analysis with LLM refinement, and includes an interactive refinement loop with dynamic tool-assisted entity/property search. This distinguishes our work from systems relying solely on model knowledge. We present experimental validation on QALD benchmarks showing consistent accuracy improvements across multiple LLMs, with ablation studies confirming the necessity of both explanation quality and dynamic tools. Human evaluation results validate explanation effectiveness. We conclude by discussing deployment as an add-on to existing systems and practical implications for training-free query refinement approaches.

References

[1] Yael Amsterdamer and Yehuda Callen. Sparqlit: Interactive sparql query refinement. In *Proc. 37th IEEE Int. Conf. on Data Engineering*, pages 2649–2652, 2021. doi:10.1109/ICDE51399.2021.00295.

[2] Mario Arias, Javier D. Fernández, Miguel A. Martínez-Prieto, and Pablo de la Fuente. An empirical study of real-world SPARQL queries. *CoRR*, abs/1103.5043, 2011.

[3] Angela Bonifati, Wim Martens, and Thomas Timm. An analytical study of large SPARQL query logs. *Proc. VLDB Endowment*, 11(2):149–161, 2017.

[4] Angela Bonifati, Wim Martens, and Thomas Timm. Navigating the maze of wikidata query logs. In *Proc. 28th Int. World Wide Web Conf.*, pages 127–138, 2019. doi:10.1145/3308558.3313472.

[5] Peter Ochieng. Parot: Translating natural language to sparql. *Expert Systems with Applications: X*, 5:100024, 2020. URL: https://www.sciencedirect.com/science/article/pii/S2590188520300032, doi:10.1016/j.eswax.2020.100024.