DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

freeda: Smart Science Space with Information Retrieval in Academy Context

Ning Xia[1], **Agnes Kleinhans**[1], **Andreas Noback**[1], and **Peter F. Pelz**[1]

[1]Technical University of Darmstadt, Chair of Fluid Systems

Abstract

Academic information environments are often characterized by heterogeneous, distributed, and insufficiently maintained data sources, which makes relevant information difficult to find and reuse. In this work, we present a knowledge-graph-based information system *freeda*, designed to provide accurate, traceable, and reliable information retrieval in the university context. We combine the knowledge graph with a retrieval-augmented generation (RAG) framework and large language models (LLMs), enabling natural-language exploration of structured data. As a proof of concept, we developed a full-stack web application.

1. Introduction

In academic environments, we face the challenge that relevant information and data are often neither easy to find nor kept up to date [1]. For example, administrative information, such as working documents and lecture announcements, is usually hosted on institutional websites. Research outputs, in contrast, are often stored individually in specific repositories. One reason for this situation is the heterogeneity of the data (text, numerical values, medians, etc.) and information sources (websites, SQL, Git repositories, etc.) themselves. On the other hand, this heterogeneity not only makes data storage and integration challenging, but also indicates a wide range of potential use cases that should be supported. However, it is not feasible to implement each of these use cases as predefined queries over the databases.

To address the issues, we propose an information system *freeda* (Free Data, Free Science, Free Society) based on a knowledge graph (KG) [2] implemented using Resource Description Framework (RDF) that aims to provide accurate, traceable, and reliable information in an academic context. The scope of the information includes university administration, research-related information, and associated data. We use a retrieval-augmented generation (RAG) framework and utilize large language models (LLM) to retrieve information from the knowledge graph. This approach also allows users to explore the knowledge graph through natural language prompts.

2. Methodology

Academic knowledge graphs must handle highly heterogeneous data while remaining maintainable for users. At the same time, structured metadata and relationships should be presented in a way that supports understandable and effective information retrieval. Several approaches use machine learning to support knowledge graph exploration. RDF2Vec [3], for example, applies embedding models to capture semantic similarity

Project Report

Available Aug 01, 2026

Correspondence to
Ning Xia
ning.xia@tu-darmstadt.de

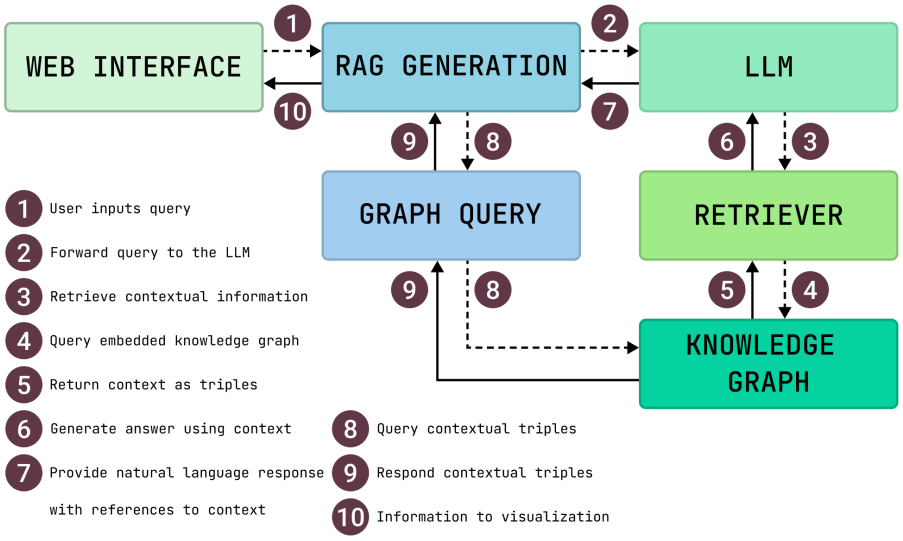

Figure 1. ***Figure 1:*** *Dataflow in* freeda *prototype as a a full-stack web application.*

between RDF triples. Text-to-SPARQL [4] uses LLMs to translate natural language into SPARQL queries.

In our project, we address these challenges by adopting a digital twin [5] approach for constructing the knowledge graph. The knowledge graph is divided into entities, each representing a real-world object. These entities contain internal relationships that describe their properties, while separate graphs represent relationships between entities. This separation simplifies metadata management across heterogeneous data types and facilitates maintainability as well as future extensions of the system.

Furthermore, this modular structure simplify the integration of RAG frameworks, which retrieve entity and its properties, and use them to answer user requests. In this way, users can interact with the knowledge graph through natural language while still benefiting from the structured representation of the underlying data. To improve usability, structured metadata and data are rendered according to their semantic and data types within a GUI. Figure 1 illustrates the system data flow.

3. Current Status and Outlook

We implemented a full-stack web application, which can be explored as a prototype at https://rmu-freeda.de, as a proof of concept and integrated about 80 heterogeneous datasets covering information on persons, scientific instruments, and collections at universities. We demonstrated how structured data can be rendered according to its data type to improve readability. We also showed the potential of a RAG-based knowledge graph explorer, through which both relevant information from the knowledge graph and the knowledge graph itself can be retrieved using natural language prompts. The demo also serves as the foundation for future development.

We are expanding the infrastructure to support the Semantic Web community and, communities using knowledge graphs for metadata management by facilitating easier workflow integration with knowledge graphs. To address our research questions, we

are finalizing our strategy for knowledge graph construction to make knowledge graph curation easier and experimenting with new methods to reduce the overhead introduced by RAG.

References

[1] M. D. Wilkinson *et al.*, "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, p. 160018, 2016, doi: 10.1038/sdata.2016.18.

[2] A. Hogan *et al.*, "Knowledge graphs," *ACM Computing Surveys*, vol. 54, pp. 1–37, 2022, doi: 10.1145/3447772.

[3] P. Ristoski, *Exploiting semantic web knowledge graphs in data mining*, vol. 38. in Studies on the Semantic Web, vol. 38. IOS Press, 2019.

[4] A. Perevalov and A. Both, "Text-to-sparql goes beyond english: Multilingual question answering over knowledge graphs through human-inspired reasoning," 2025. [Online]. Available: https://arxiv.org/abs/2507.16971

[5] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds., Springer International Publishing, 2017, pp. 85–113. doi: 10.1007/978-3-319-38756-7_4.