

KONDA: An LLM-based Tool for Semantic Annotation and Knowledge Graph Creation Using Ontologies for Research Data

Soo-Yon Kim¹ , Martin Görz¹ , and
Sandra Geisler¹ 

¹RWTH Aachen University, Aachen, Germany

*Correspondence: Soo-Yon Kim, kim@dbis.rwth-aachen.de

Abstract

Achieving semantic interoperability of research data is key to enabling cross-domain data integration, reuse, and knowledge discovery [1]. While the need to align heterogeneous datasets using shared vocabularies and ontologies is widely recognized, doing so remains a considerable challenge in practice [2], [3].

Researchers face several challenges:

- **(C1) Lack of expertise in ontologies:** Many researchers are unfamiliar with ontology engineering and semantic annotation.
- **(C2) Absence of established domain ontologies:** While some domains, such as medicine, have well-established vocabularies, other domains, such as production engineering, may lack suitable or widely adopted ontologies, making it difficult to identify reusable options.
- **(C3) Technical barriers:** The knowledge required to work with technologies such as RDF or mapping tools often presents an entry barrier.
- **(C4) Tool heterogeneity:** Working with multiple disconnected tools adds cognitive and technical overhead.
- **(C5) Limited resources:** Researchers typically face time constraints, making it difficult to invest in familiarizing themselves with complex tools or processes.
- **(C6) Proprietary solutions:** Many semantic mapping tools (e.g., Talend [4]) are proprietary and not suitable for scientific work.

To address these challenges, we present KONDA, an LLM-based tool that supports semantic enrichment of research datasets and the construction of explorable knowledge graphs within a single integrated workflow.

The KONDA workflow is as follows:

- An interface prompts the user to upload their research dataset, along with optional supplementary documents (e.g., protocols, DMPs, README files) to provide the tool with context.
- The user is supported in the selection of suitable ontologies via a direct integration with the TIB Terminology Service [5], with the option to add custom ontologies.

- The tool performs automated LLM-based semantic annotation of the dataset using the provided context and selected ontologies. A feedback screen enables the user to review and correct annotations.
- The annotated data is provided in RDF format with an immediate visualization as a knowledge graph.

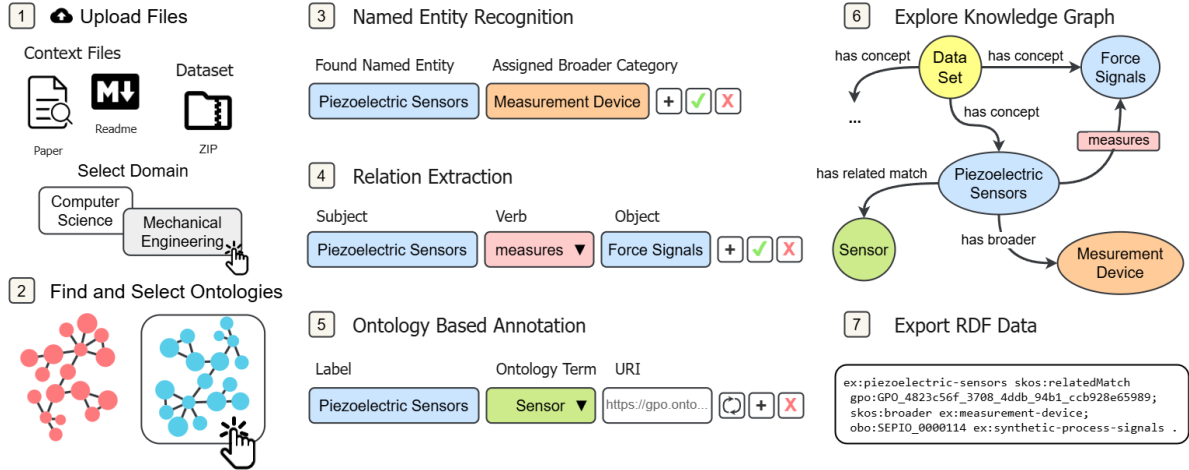


Figure 1. A simplified overview of the workflow of KONDA.

KONDA's architecture comprises a user interface, a server backend managing sessions and data processing, and an API layer that connects the tool to an LLM, where the semantic enrichment is conducted with techniques such as named entity recognition, relation extraction, and ontology-based annotation.

Through KONDA, a guided, interactive tool is provided in which users receive LLM-assisted suggestions and the opportunity to intuitively explore their enriched data directly through automated knowledge graph creation, thus reducing required technical or formal training in semantic technologies (C1, C3). The discovery of reusable ontologies is enabled through the integration of terminology services (C2). KONDA unifies the pipeline within a single, cohesive environment (C4). The tool's semi-automated workflow provides fast and visually supported results with minimal manual effort (C5) while retaining opportunities for human feedback to ensure output quality. Finally, KONDA's modular backend supports the deployment of both proprietary and open LLMs (C6).

KONDA empowers researchers to semantically enrich their datasets with minimal effort, offering an integrated and adaptable solution. Future development will focus on persistent graph storage, automated ontology recommendations, and evaluation in real-world settings. By leveraging LLMs and emphasizing usability, KONDA provides a robust foundation for advancing data interoperability across disciplines.

Resources

- KONDA: An AI RDM Tool. https://git.rwth-aachen.de/dsma/projects/iop_internal/rdm/ai-rdm-tool (restricted due to ongoing research, access possible upon request). "KONDA is an LLM-based tool for semi-automatic semantic annotation and knowledge graph creation using concepts from ontologies for research data." Last updated 04/14/2025.

Author contributions

Conceptualization: Soo-Yon Kim, Martin Görz; Data curation: Martin Görz; Funding acquisition: Sandra Geisler; Investigation: Martin Görz; Methodology: Soo-Yon Kim, Martin Görz; Project administration: Soo-Yon Kim; Software: Martin Görz; Resources: Soo-Yon Kim; Supervision: Sandra Geisler; Validation: Soo-Yon Kim, Martin Görz; Visualization: Martin Görz; Writing - original draft: Soo-Yon Kim; Writing - review & editing: Martin Görz, Sandra Geisler.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612.

References

- [1] H. P. Sustkova, K. M. Hettne, P. Wittenburg, *et al.*, “Fair convergence matrix: Optimizing the reuse of existing fair-related resources,” *Data Intelligence*, vol. 2, no. 1-2, pp. 158–170, Jan. 2020. DOI: [10.1162/dint_a_00038](https://doi.org/10.1162/dint_a_00038).
- [2] H. Liyanage, P. Krause, and S. de Lusignan, “Using ontologies to improve semantic interoperability in health data,” *Journal of innovation in health informatics*, vol. 22, no. 2, pp. 309–315, 2015. DOI: [10.14236/jhi.v22i2.159](https://doi.org/10.14236/jhi.v22i2.159).
- [3] I. M. Putrama and P. Martinek, “Heterogeneous data integration: Challenges and opportunities,” *Data in Brief*, vol. 56, p. 110853, Oct. 2024. DOI: <https://doi.org/10.1016/j.dib.2024.110853>.
- [4] Q. I. AB. “Talend.” (first published 2006), [Online]. Available: <https://www.talend.com> (visited on 04/16/2025).
- [5] L. I. C. for Science and T. U. Library. “Tib terminology service - ontologies.” (first published 2023), [Online]. Available: <https://terminology.tib.eu/ts/ontologies> (visited on 04/15/2025).