A proposal to CRIS 2024:

"EMERGING TRENDS FOR INTERNATIONAL COLLABORATION IN THE CRIS DOMAIN"

Included in the following suggestion category:

«Semantic approaches, data mining and profiling in the interpretation of research information»

Title: **KNOWLEDGE GRAPHS – THE FUTURE OF INTEGRATION IN CRIS SYSTEMS FOR USES OF ASSISTANCE TO SCIENTIFIC REASONNING**

Authors: Renaud Fabre[a] & Otmane Azeroual[b]

[a] Renaud Fabre, Dionysian Economics Lab (LED), University Paris 8, 93200 Saint-Denis, France, renaud.fabre01@gmail.com

[b] German Centre for Higher Education Research and Science Studies (DZHW), Berlin, Germany, azeroual@dzhw.eu

https://cris2024.eurocris.org/

Abstract

Knowledge graphs (KG) are increasingly coming into focus as they provide a powerful method for data integration and knowledge representation. Their semantic data model, which represents knowledge in terms of entities, attributes, and relationships between those entities, applies well to descriptive encyclopedic uses, but encounters challenging limitations in scientific knowledge applications where support for contested knowledge categorizations in research is poorly applied.

In the context of the platformization of science, especially in the context of CRIS systems, we have observed that knowledge graphs can be used to combine the diverse data sources and data formats that exist in the research landscape and create unified and connected data models. It therefore enables researchers, administrators and other stakeholders to access comprehensive and consistent information relevant to their work.

The following paper examines the specific role of KG in the future of data integration in CRIS systems in supporting scientific thinking. It highlights the advantages and disadvantages of the current features of KG. Advantages and disadvantages include flexible knowledge modeling, support for semantic queries, and interoperability with other data sources and formats. Systemic limitations consist mainly in the methodological and technological expression of controversies and scientific disagreements, which significantly limits the potential of the scientific classical investigation of relatedness, identity and categorization of controversial

new ideas using knowledge graphs: this represents a serious limitation to the use of the KG of CRIS.

The paper presents advances in solutions to support scientific thinking, with various use cases and best practices for implementing KG in CRIS systems, enabling research institutions and scientific organizations to improve their data analysis and support of scientific thinking.

Concluding remarks concern ongoing work and new results; finally, we discuss the pace of challenges that open up new approaches to supporting scientific thinking that are currently opening up the interaction of large language models and the KG in new technologies.

1. Identifying challenges in supporting scientific thinking in KG

While the platformization of science expands its networks, the structural analysis of platforms' behaviors and architectures still requires a unified taxonomy, recently proposed with a nomenclature of change agents and categories (da Silva Neto & Chiarini, 2023). Considering a large-scale survey, we show that the organizational concept of data integration in CRIS systems optimizes the analytical capacities of data platforms in very large research infrastructures (Fabre, et al. 2021). In the same article, the additional future challenges of data integration and analysis on scientific platforms were examined using the example of research in bioinformatics[1].

2. Research impact assessment in KG technology

Research impact assessments have been extensively studied in line with new trends in the architecture of scientific knowledge graphs (Manghi et al. 2021) and show that representations are challenging in all areas where scientific knowledge communication needs to express disputes, disagreements and any significant contradictions build categorical knowledge allocation. New approaches are proposed to support the documentation of scientific thinking in the Open Knowledge Research Graph (ORKG) (Jaradeh et al. 2019).

3. Supporting scientific thinking with KG on science platforms

Due to the limited expressive power of knowledge graphs in scientific communication, research has been developed in recent years to better understand the drawbacks in the methodology, technologies, reasoning ability of knowledge graphs, and their message delivery mechanisms (Khemani et al. 2024). During the same period, human-AI collaboration in transparent data creation has become a hot topic, with new guidelines for “Human-AI Teaming” (Berretta et al. 2023) identifying the difficult questions surrounding and the importance of individualizing human thinking in this context final scientific validation in this context.

4. Disputed knowledge categorization and the scientific KG

Cross-domain data extraction and knowledge graph construction for dispute analysis are generally considered an essential task in building new knowledge (Guo et al. 2023). However, the research proceeds with robust and data-neutral methods that could work in CRIS systems with the integration of rich knowledge graphs of rich and efficient services to support scientific

[1] See https://www.ebi.ac.uk/services

thinking in assessment and validation tasks. Having identified this research question in its analytical dimensions (Fabre, 2019), we introduce a conceptual work on contested knowledge categorization for research activities (Fabre et al. 2023), which must be neutral and transparent to meet the needs of scientific platforms and carry out these tasks to support scientific thinking. This knowledge graph has diverse applications in this area, as we show with various examples of big data contexts and data systems (Azeroual et al. 2023), (Azeroual, O., & Fabre, R. (2021), and there are no doubt that significant developments would take place by the time of CRIS2024.

5. New approach to supporting scientific thinking: Large Language Models (LLM) and KG

Currently, the interaction of large language models and knowledge graphs in large-scale science platforms is booming, with further unprecedented consequences for the division of tasks in the research workflow. In research and observation, much attention is paid to the interaction that graph texts have recently introduced (Zhao et al. 2023). Many consequences are currently being investigated (Wei et al. 2024).

Conclusion

Changing the nature of KG demand is the nature of scientific thinking. Technology and pluridisciplinary approaches are answers that strengthen research to meet the needs at an unprecedented pace of innovation, with research taking advantage of innovation advantages that are taking place today in applications of AI in other areas of web augmentation services. It would provide CRIS with a variety of opportunities to share experiences and develop new rules for data sharing in science.

References

1. Azeroual, O., & Fabre, R. (2021). Processing big data with apache hadoop in the current challenging era of COVID-19. *Big Data and Cognitive Computing*, *5*(1), 12.
2. Azeroual, O., Fabre, R., Störl, U., & Qi, R. (2023). Elastic Stack and GRAPHYP Knowledge Graph of Web Usage: A Win–Win Workflow for Semantic Interoperability in Decision Making. *Future Internet*, *15*(6), 190.
3. Berretta, S., Tausch, A., Ontrup, G., Gilles, B., Peifer, C., & Kluge, A. (2023). Defining human-AI teaming the human-centered way: a scoping review and network analysis. *Frontiers in Artificial Intelligence*, *6*.
4. da Silva Neto, V. J., & Chiarini, T. (2023). The Platformization of Science: Towards a Scientific Digital Platform Taxonomy. *Minerva*, *61*(1), 1-29.
5. Fabre, R. (2019). A" searchable" space with routes for querying scientific information. In *BIR 2019* (No. 2345, pp. 112-124). CEUR-WS. org.
6. Fabre, R., Egret, D., Schöpfel, J., & Azeroual, O. (2021). Evaluating the scientific impact of research infrastructures: The role of current research information systems. *Quantitative Science Studies*, 2(1), 42-64.

7. Fabre, R., Azeroual, O., Schöpfel, J., Bellot, P., & Egret, D. (2023). A Multiverse Graph to Help Scientific Reasoning from Web Usage: Interpretable Patterns of Assessor Shifts in GRAPHYP. *Future internet*, *15*(4), 147.
8. Guo, Q., Chen, X., Zhou, P., & Liao, Y. (2023). Cross-Domain Data Extraction and Knowledge Graph Construction for Dispute Analysis. In *2023 IEEE 43rd International Conference on Distributed Computing Systems (ICDCS)* (pp. 959-960). IEEE.
9. Jaradeh, M. Y., Oelen, A., Farfar, K. E., Prinz, M., D'Souza, J., Kismihók, G., ... & Auer, S. (2019). Open research knowledge graph: next generation infrastructure for semantic scholarly knowledge. In *Proceedings of the 10th International Conference on Knowledge Capture* (pp. 243-246).
10. Khemani, B., Patil, S., Kotecha, K., & Tanwar, S. (2024). A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. *Journal of Big Data*, *11*(1), 18.
11. Manghi, P., Mannocci, A., Osborne, F., Sacharidis, D., Salatino, A., & Vergoulis, T. (2021). New trends in scientific knowledge graphs and research impact assessment. *Quantitative Science Studies*, *2*(4), 1296-1300.
12. Wei, Y., Fu, S., Jiang, W., Kwok, J. T., & Zhang, Y. (2024). Rendering Graphs for Graph Reasoning in Multimodal Large Language Models. *arXiv preprint arXiv:2402.02130*.
13. Zhao, J., Zhuo, L., Shen, Y., Qu, M., Liu, K., Bronstein, M., ... & Tang, J. (2023). Graphtext: Graph reasoning in text space. *arXiv preprint arXiv:2310.01089*.