Knowledge Graphs in the Era of Large Language Models: Opportunities and Challenges

Danka Jokić, Ranka Stanković, Jelena Jaćimović

Дигитални репозиторијум Рударско-геолошког факултета Универзитета у Београду

[ДР РГФ]

Knowledge Graphs in the Era of Large Language Models: Opportunities and Challenges | Danka Jokić, Ranka Stanković, Jelena Jaćimović | South Slavic Languages in the Digital Environment JuDig Book of Abstracts, University of Belgrade - Faculty of Philology, Serbia, November 21-23, 2024. | 2024 | |

http://dr.rgf.bg.ac.rs/s/repo/item/0009159

Danka Jokić, Ranka Stanković, Jelena Jaćimović

Društvo za jezičke resurse i tehnologije JeRTeh

E-mail: danka.jokic@afrodita.rcub.bg.ac.rs; ranka.stankovic@rgf.bg.ac.rs;

jelena.jacimovic@stomf.bg.ac.rs

Knowledge Graphs in the Era of Large Language Models: Opportunities and Challenges

The emergence of large language models (LLMs) has significantly impacted the field of artificial intelligence (AI) by excelling at language processing and generation tasks. However, a critical limitation of LLMs lies in their lack of structured knowledge and reasoning capabilities, hindering their effectiveness in real-world applications that demand factual accuracy and context-aware reasoning. Knowledge graphs (KGs), on the other hand, offer a compelling solution. By representing entities and their relationships in a machine-readable format, KGs provide a rich source of structured knowledge. This convergence presents a unique opportunity to explore their symbiotic relationship: leveraging KGs to empower LLMs for the development of next-generation AI applications.

This synergy has the potential to significantly enhance LLM capabilities in several key areas. First, grounding LLMs in the structured knowledge of KGs can substantially improve their ability to comprehend factual information and generate more accurate and reliable responses to complex questions. In addition, KGs also offer the necessary context and relationships to enable LLMs to perform more sophisticated reasoning tasks and incorporate commonsense knowledge into their reasoning processes, leading to more nuanced and human-like understanding. Furthermore, the explicit relationships within KGs can be harnessed to explain the reasoning behind LLM outputs, directly addressing a critical challenge in interpretable AI.

While the potential is undeniable, challenges require attention. Developing effective methods for LLMs to jointly learn from text data and knowledge graphs is crucial for successful integration. Additionally, ensuring the consistency and quality of knowledge graph data is essential, as incomplete or inaccurate information can lead to biased or erroneous LLM results.

Despite these challenges, the integration of LLMs and KGs holds immense potential to revolutionize various AI applications. LLMs empowered by KGs can provide more accurate and comprehensive answers in question-answering systems. Intelligent assistants integrated with KGs can understand and respond to user queries with greater context and factual grounding. Additionally, the combination of LLMs and KGs can lead to the development of more factually accurate and contextually relevant natural language generation, and the creation of more sophisticated and dynamic knowledge representation systems.

Exploring this synergistic relationship between LLMs and KGs is crucial for advancing AI. By addressing the challenges and pursuing effective integration methods, we can pave the way for the development of next-generation AI applications characterized by enhanced understanding, reasoning, and knowledge representation capabilities. In this paper we will present how knowledge graphs can be used to enhance reasoning capabilities of large language models concerning online safety and moderation of harmful textual content in Serbian.

Keywords: *Knowledge Graphs (KGs), Large Language Models (LLMs), Natural Language Processing (NLP), Structured Knowledge, Data Quality, Explainable Artificial Intelligence (EAI), Online Content Safety*

References

[1] Agrawal, G., Kumarage, T., Alghamdi, Z., & Liu, H. (2024). Can Knowledge Graphs Reduce Hallucinations in LLMs? : A Survey. In K. Duh, H. Gomez, & S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 3947–3960). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.naacl-long.219.

[2] Kau, A., He, X., Nambissan, A., Astudillo, A., Yin, H., & Aryani, A. (2024). Combining Knowledge Graphs and Large Language Models (arXiv:2407.06564). arXiv. https://doi.org/10.48550/arXiv.2407.06564.

[3] Lin, J. (2022). Leveraging World Knowledge in Implicit Hate Speech Detection. In L. Biester, D. Demszky, Z. Jin, M. Sachan, J. Tetreault, S. Wilson, L. Xiao, & J. Zhao (Eds.), Proceedings of the Second Workshop on NLP for Positive Impact (NLP4PI) (pp. 31–39). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.nlp4pi-1.4.

[4] Pan, J., Razniewski, S., Kalo, J. C., Singhania, S., Chen, J., Dietze, S., ... & Graux, D. (2023). Large Language Models and Knowledge Graphs: Opportunities and Challenges. Transactions on Graph Data and Knowledge.

[5] Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying Large Language Models and Knowledge Graphs: A Roadmap. IEEE Transactions on Knowledge and Data Engineering, 36(7), 3580–3599. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/TKDE.2024.3352100.

Nikola Janković

Faculty of Philology, University of Belgrade
E-mail: nikolajankovickv@gmail.com

Jovana Ivaniš

Institute for Serbian Language SANU
E-mail: jovana.ivanis@gmail.com

Usage of the Whisper Large V3 Sr Model for the Transcription of Serbian Spoken Language in Python Programming Language on the Google Colab Platform

This paper presents a *Python* script on the *Google Colab* platform, which uses a fine-tuned model for the transcription of speech in Serbian, *Whisper Large v3 Sr* (https://huggingface.co/Sagicc/whisper-large-v3-sr-cmb), which enables free, high-quality and simple transcription of Serbian speech into text. The motivation for creating this script came from the lack of available tools which would allow researchers to use this model in a straightforward way, without the need for advanced technical knowledge and significant computing resources. This script provides a simple method for uploading audio files, transcribing them using the *Whisper Large V3 Sr model*, and downloading the transcription of the files in a textual format. Firstly, the paper briefly describes the aforementioned model used by the script, followed by a detailed description of how the script functions, along with a user manual. The paper will also present problems which we needed to overcome in order to successfully implement this approach, such as the need for automatic audio file segmentation,