Empowering Knowledge Graphs with 🤖: A Cost-Effective Workflow

*Anahita Pakiman, Christian Beil and Stefan Götz

Information Management, brox-IT Solutions, Hannover, Germany.

*Corresponding author(s). E-mail(s): apakiman@brox.de;
Contributing authors: cbeil@brox.de; sgoetz@brox.de;

Linked data is a core element of Semantic Web Technologies (SWT), facilitating large-scale data integration and reasoning. The Knowledge Graph (KG), a key component of the Semantic Web, has demonstrated significant power in various domains. The strength of KG lies in data engineering, specifically in the construction of semantics through the modelling of graph data and ontologies. Although KG has the potential to revolutionise business operations, its integration into industries such as manufacturing is still incomplete and has not yet reached its full potential [1]. Several challenges contribute to this incomplete integration, such as the difficulty of integrating KG data with existing enterprise systems, the demand for skilled developers, and the high costs associated with implementing KG solutions. In addition, the business value of KGs is not clearly defined in many sectors.

On the other hand, Large Language Model (LLM) is another approach of various technologies in knowledge engineering. The importance and growth of Language Model (LM) and especially LLM in the current decade is highlighted by a survey assessing the popularity of different aspects of knowledge engineering [2]. This survey suggests that the era of SWT is stabilising, while the use of LLM is increasing. Looking ahead, a key question for the field is how to continue to develop. Despite impressive results from LLM, uncertainties remain that are [3]:

- Output with hallucination [4] and requires ground truth
- Expensive to train and operate
- Difficult to fix and update

- Difficult to audit and explain, which is necessary for domain analysis
- Inconsistent responses
- Struggles with low-resource languages

As far as is known, the KGs do not have these problems. Therefore, using KGs to generate AI solutions for storing domain-specific knowledge still has great potential and utility.

In a synergistic combination, LLMs methods have recently contributed significantly to breakthroughs in KG construction, e.g. tasks like named entity recognition, entity typing, entity linking, coreference resolution, and relation extraction [5]. Moreover, deep knowledge representation models refine KGs by addressing issues like completing corrupt tuples, discovering new tuples within existing graphs, and merging graphs from diverse sources. Several knowledge bases, including TransOMCS, ASER, and huapu, have implemented automatic KG construction methods [5]. We also firmly believe that the construction of KGs with the support of LLMs is a promising and forward-looking approach.

Given KGs enterprise's technology readiness, our training is dedicated to presenting a cost-effective workflow developed by brox-IT Solutions. The solution emphasises the use of low-cost or open-source technologies. The workshop also covers KGs construction through the incorporation of LLM techniques to automate the process. This will not only streamline KGs construction, but also highlight Graph Machine Learning (GML) discoveries as an added value to enterprise data. The aim is to help industries take their first steps into this technology. This training is aimed specifically at organisations interested in building an Enterprise KGs, enabling them to demonstrate the value that KGs can bring to their business.

1 Targeted audiences

This training is aimed at individuals or groups within industries interested in digitisation and seeking valuable use cases for implementing KGs. It is designed for those with minimal to no previous experience of KGs and data science, but who are keen to integrate KGs into their organisation's data systems. Our primary intention is to provide a streamlined workflow that will allow for the proof of concept of the KGs within organisations. This workflow will minimise the complexity associated with exploring and establishing KGs. In order to facilitate the learning process, all training materials will be accessible via a GitHub repository, allowing participants to actively engage with the content. However, due to practical limitations in dealing with large numbers of participants, feedback sessions during the training will address general questions, while specific or in-depth questions will receive dedicated support after the training, particularly if the number of participants exceeds 15.

References

[1] Buchgeher G, Gabauer D, Martinez-Gil J, et al (2020) Knowledge graphs in manufacturing and production: A systematic literature review. IEEE Access 9:55,537–55,554. URL https://api.semanticscholar.org/CorpusID:229210861

[2] Groth P, Simperl E, van Erp M, et al (2023) Knowledge graphs and their role in the knowledge engineering of the 21st century. In: Dagstuhl Reports, Schloss Dagstuhl-Leibniz-Zentrum für Informatik

[3] Vrandečić D (2023) Keynote: The future of knowledge graphs in a world of large language models, URL https://www.knowledgegraph.tech/speakers/denny-vrandecic-2/,https://www.youtube.com/watch?v=WqYBx2gB6vA&t=1059s, the Knowledge Graph Conference

[4] Welleck S, Kulikov I, Roller S, et al (2019) Neural text generation with unlikelihood training. arXiv preprint , id: 190804319

[5] Zhong L, Wu J, Li Q, et al (2023) A comprehensive survey on automatic knowledge graph construction. ACM Computing Surveys URL https://api.semanticscholar.org/CorpusID:256808669