Will Wohnen

SS24, 192.116 Knowledge Graphs @ TU Wien

Yahya Jabary

Despite the vast amount of online real estate data available, traditional analysis methods often fail to uncover hidden relationships and insights, leaving investors and homebuyers with incomplete information. Knowledge graphs offer a powerful solution, enabling the integration and analysis of diverse data sources to reveal complex patterns and connections that inform smarter investment decisions.

But due to the novelty of this method in both research and industry many stakeholders remain unaware of the potential of real estate knowledge graphs. By conducting a comprehensive case study — from data scraping and preprocessing to graph analytics — we aim to provide a practical framework that showcases the technology's capabilities and initiates broader discussions on its adoption and hope to motivate further research in this area.

Solution. We plan to find the most popular Viennese real estate websites, scrape them, feed the data into Neo4J and to gain a quantitative edge in the market by applying logical inference and studying the Neo4J-GDS library.

that we have just highlighted is the real estate market. Cardorel et al.[1] propose using knowledge graphs to represent uncertain geospatial information from real estate advertisements. This approach enables the integration of geospatial data into knowledge graphs, allowing users to query and analyze location-based information more effectively. Nonetheless, the intersection of knowledge graphs and real estate remains largely unexplored, with many opportunities for further research and development.

(LO1) Understand and apply Knowledge Graph Embeddings

(LO2) Understand and apply logical knowledge in KGs

(LO4) Compare different Knowledge Graph data models from the database, semantic web, machine learning and data science communities.

(LO5) Design and implement architectures of a Knowledge Graph

(LO7) Apply a system to create a Knowledge Graph

(LO8) Apply a system to evolve a Knowledge Graph

(L11) Apply a system to provide services through a Knowledge Graph

Addendum

In this section we discuss learning objectives that involve discussion, reflection and evaluation on what has been learned rather than synthesis. These learning objectives are recognizable through the use of the verb "describe" an

(LO6) Performance Optimizations Information retrieval systems such as databases are very sensitive to query latency and throughput. As the size of the knowledge graph grows, the complexity of queries increases, leading to more expensive computations and longer execution times. To address these challenges, researchers have developed various techniques to optimize query performance in knowledge graphs. System optimizations can be broadly categorized into systems-side and logic-side optimizations. Systems-side optimizations focus on improving the underlying infrastructure, such as hardware utilization, data storage and network communication. These optimizations aim to enhance the overall system performance without changing the logical structure of queries. On the other hand, logic-side optimizations target the query itself, aiming to improve its execution plan, reduce redundant computations and optimize the logical structure of the query. The Vadalog system by Bellomarini et al.[2] is an example of a logic-side optimization technique that enhances query execution by identifying and caching subqueries. This

[1]Cadorel, L., Tettamanzi, A.G., & Gandon, F.L. (2022). Towards a representation of uncertain geospatial information in knowledge graphs. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geospatial Knowledge Graphs.

[2]Bellomarini, L., Gottlob, G., & Sallinger, E. (2018). The Vadalog system: Datalog-based reasoning for knowledge graphs. arXiv preprint arXiv:1807.08709.

technique, known as "warting", reduces the overall computation time by precomputing and storing subqueries for reuse. By combining systems-side and logic-side optimizations, researchers open up new possibilities for real-world applications.

In our case study, the most important optimizations that were applied were the use of `asyncio` co-routines for parallelizing the scraping process, and the use of `Neo4J`'s `Graph Data Science` library for optimizing the graph algorithms and queries. Using the in-house `Neo4J` database, we were able to leverage the power of in-memory graph projections and parallelized graph algorithms to massively speed up the query times and reduce the overall latency of the system to a few milliseconds per iteration. With more time and resources, we could further optimize the system enabling soft real-time querying and analysis of the real estate market in Vienna.

(LO9, L10) Real-World Applications The versatility of knowledge graphs is best captured by the survey paper by Ji et al.[3], which provides an overview of knowledge graph applications. According to this survey, knowledge graphs are used in fields including question answering systems, recommendation systems, information extraction, natural language processing, risk analysis and decision-making processes. In question answering systems, knowledge graphs help computers understand and respond to human queries more effectively. In recommendation systems, knowledge graphs provide more relevant item suggestions to users based on their preferences and behavior. In information extraction, knowledge graphs help extract important facts from unstructured text, enabling users to access structured information more easily. In natural language processing, knowledge graphs support tasks such as text classification and summarization, improving the accuracy and efficiency of language processing algorithms. One domain that stands out is the financial sector. In risk analysis, knowledge graphs help identify potential problems in complex systems by modeling relationships between different entities and events. Lastly, in decision-making processes, knowledge graphs provide structured information to guide choices and support informed decision-making.

(LO12) KGs vs. ML / AI In all these applications, knowledge graphs play a crucial role in "grounding" machine learning models (which themselves are a subset of artificial intelligence). Grounded models are better at reasoning about complex topics and provide more reliable predictions. This has most prominently been demonstrated in the field of natural language processing, where knowledge graphs have been used to improve the performance of language models and information retrieval / retrieval augmented generation systems. By connecting knowledge graphs with otherwise untrustworthy and hallucinating models (such as GPT-3), we can substantially improve the quality and trustworthiness of the predictions. This is especially important in high-stakes applications such as medical diagnosis, financial risk analysis and legal decision-making. The connection between knowledge graphs and machine learning is a key area of research that is likely to grow in importance as stakeholders increasingly rely on machine learning models for decision-making.

---

[3]Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE transactions on neural networks and learning systems, 33(2), 494-514.