Comparison of Retrieval-Augmented Generation (RAG) Systems for Extracting Test Scenarios from UN Regulations

Description

The automotive industry relies on strict standards and regulations, such as the UN regulations No. 152 and No. 131 (Advanced Emergency Braking Systems), to ensure vehicle safety and compliance. Extracting relevant test scenarios from these complex regulatory documents is a challenging yet essential task for validating autonomous driving systems. This master's thesis focuses on evaluating various Retrieval-Augmented Generation (RAG) systems for their effectiveness in extracting such scenarios from UN regulations.

RAG systems combine the capabilities of information retrieval with language generation models to generate content based on extracted knowledge. The entire RAG system consists of two core modules: the retriever and the generator, where the retriever searches for relevant information from the data store and the generator produces the required contents. Different types of retrievers exist, such as dense retrievers and sparse retrievers. Dense retrievers, such as those based on BERT **[1]**, that are good at encoding semantic information, whereas sparse retrievers like TF-IDF**[2]** rely on term frequency and inverse document frequency to identify relevant documents. Additionally, there is the possibility of using knowledge graphs for RAG, such as the GraphRAG**[3]**. The student should explore and compare the most prominent solutions for these methods, considering factors such as retrieval accuracy, computational efficiency, and suitability for extracting specific regulatory information**[4]**.

Tasks

- Survey and review various RAG systems and methods, including dense retrievers, sparse retrievers, and knowledge graph-based systems, starting with those outlined in the referenced survey paper.
- Develop a framework for evaluating the effectiveness of different RAG systems (dense, sparse, and knowledge graph-based) in extracting test scenarios from UN regulations.
- Implement and compare selected RAG methods to assess their retrieval capabilities, generation quality, and applicability to specific test scenario extraction tasks.
- Experiment with different configurations, including both open-source and commercial RAG systems, and analyze their performance with respect to extracting precise and interconnected regulatory information.

References

[1] J. Devlin, M. Chang et al., “BERT: pre-training of deep bidirectional transformers for language understanding,” in NAACL-HLT, 2019.

[2] “On relevance weights with little relevance information | Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval.” Accessed: Nov. 14, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/258525.258529

[3] D. Edge *et al.*, “From Local to Global: A Graph RAG Approach to Query-Focused Summarization,” Apr. 24, 2024, *arXiv*: arXiv:2404.16130. doi: 10.48550/arXiv.2404.16130.

[4] V. Zolfaghari, N. Petrovic, F. Pan, K. Lebioda, and A. Knoll, “Adopting RAG for LLM-Aided Future Vehicle Design,” Nov. 14, 2024, *arXiv*: arXiv:2411.09590. doi: 10.48550/arXiv.2411.09590.

Technische Universität München

TUM School of Computation, Information and Technology

Lehrstuhl für Robotik, Künstliche Intelligenz und Echtzeitsysteme

---

Supervisor: Prof. Dr.-Ing. Alois Knoll

Advisor:
Vahid Zolfaghari, M.Sc.

Research project: MANNHEIM-CeCaS

Type:
BA/MA

Research area: Retrieval Augmented Generation, automotive software testing

Programming language: Python

Required skills: Object-oriented programming with Python (must have); Experience with LLM related libraries like langchain (nice to have).

Language:
English

For more information please contact us:

E-Mail: v.zolfaghari@tum.de

Internet: www.ce.cit.tum.de/air