EPRI's Experience with Semantic Tool Application in Nuclear Decommissioning Knowledge Management

Background and Motivation

In the global context of nuclear decommissioning projects, effective knowledge management is necessary due to long project durations, stringent regulatory requirements, and the safety-critical nature of operations. At the Electric Power Research Institute (EPRI), subject matter experts draw on more than 30 years of accumulated nuclear decommissioning knowledge to provide support and advice to various utilities and organizations. However, due to staff retiring, new staff onboarding, disparate data sources and legacy documentation, the development of a knowledge management system has become a priority for organizing and accessing all the information effectively. Recent advances in high-performance computing, semantic technologies, and large language models have enabled significant progress in processing and organizing unstructured text data. Taking advantage of these new technologies, EPRI is working on developing a queryable information retrieval database. This database, connected with a chatbot for ease of use, would aid the subject matter experts in their daily work by facilitating quicker access to pertinent information.

Objectives of the Use Case

The first objective of the use case is the construction of an internal tool at EPRI that would leverage modern semantic technologies to provide immediate value in the SME's day-to-day work, empower personnel with less decommissioning-specific expertise, and onboard junior staff more quickly.

The second objective is the testing of these modern semantic technologies itself, assessing their added value, applicability to specific use cases, challenges, pitfalls, and limitations. The insights learned through this project would be shared across the institute and inform further research directions and other use cases utilizing similar technologies. Furthermore, the aim is to have the code bases and algorithms developed as part of this project be modular and re-usable.

Lastly, such efforts are closely aligned with international initiatives led by organizations like the IAEA and emphasize the adoption of standardized taxonomies and ontologies to promote interoperability across the global nuclear decommissioning community. By

collaborating across organizations, best practices and experiences in applying semantic technologies specifically in the nuclear decommissioning field can be shared.

Approach and Methodology

The approach adopted for this use case leverages semantic technologies and knowledge graph architectures to organize and retrieve nuclear decommissioning information in a structured, interoperable manner. This aligns with international best practices for knowledge management in nuclear organizations as outlined in [1].

Semantic Search

The first approach focused on developing a semantic information retrieval database, relying solely on semantic search with embedding models, but without the use of Large Language Models (LLMs). The goal of the first phase was to experiment with various text processing and search optimization techniques, and to develop user interface and gather early feedback. Furthermore, it was important to establish data governance practices aligned with the directive of EPRI's legal team. As such, only publicly available data was used on a secured on-premises, internal server.

During this phase, EPRI also experimented with automating evaluation of the retrieval accuracy by testing against a set of queries & expected retrieved relevant source documents. To speed up creation of the evaluation set, an LLM was used to prepare such questions [7] which were later reviewed by the SMEs. This approach was found to have limited success, however, and ultimately SME-curated evaluation set remains as a gold standard.

Commented [DC1]: Changed to third person to match voice within the rest of the document.

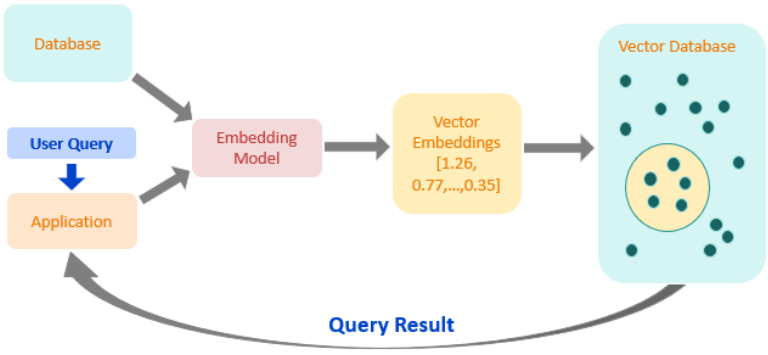
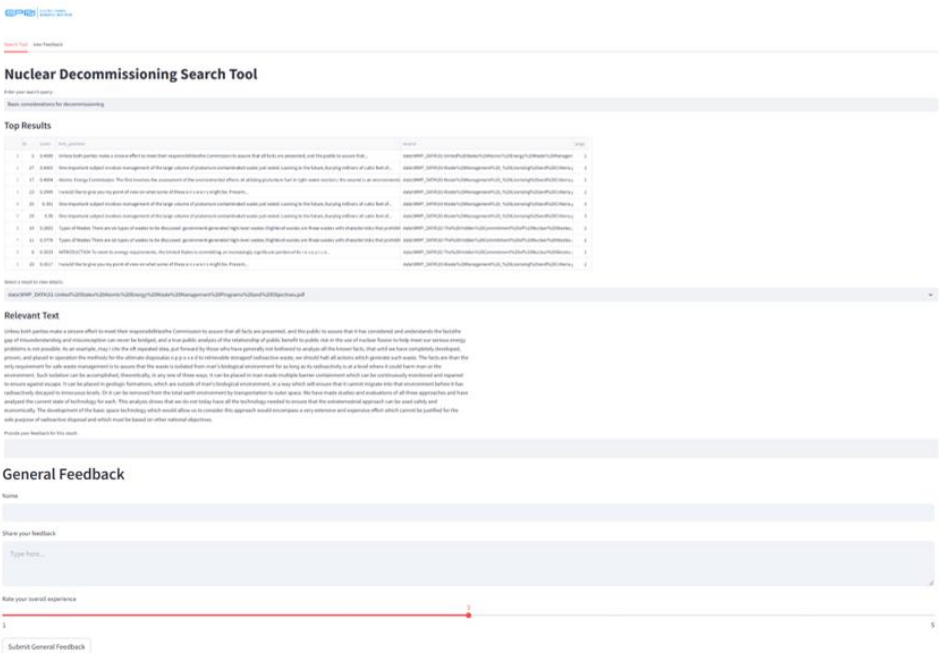


Figure 1 Standard semantic search approach, figure based on https://www.qwak.com/post/utilizing-llms-with-embedding-stores



Knowledge Graph

The second phase of the project centered on building a knowledge graph [2] pipeline. A Knowledge Graph (KG) enables representation of entities and relationships in a machine-readable format. This approach supports semantic interoperability across taxonomies and ontologies. We followed a standard approach as described in [1, 2], and used the IAEA taxonomy document [3] as a starting point. To construct the knowledge graph and design an iterative pipeline, we followed steps as outlined in [4-6]:

- **Text Preprocessing and Normalization**
  Publicly available NLP libraries such as **SpaCy** [https://spacy.io] and **NLTK** [https://www.nltk.org] were used for tokenization, lemmatization, and stop-word removal. Text preprocessing remains as a challenging part of the process and needs further refinements.
- **Entity and Relationship Extraction**

  An LLM-based keyword extraction and entity definition was used as a starting point. The entities were compared to the IAEA taxonomy document. Entities (nodes) and relationships (edges) were then compared, deduplicated and reviewed to further help

define the Graph Schema. This approach was helpful in developing a first pass at the KG with minimal SME involvement, and in preparing a pipeline with SME-in-the-loop to help develop the KG. We also developed a pipeline to train a RoBERTa transformer model in order to provide more reliable output, but this work is still in progress as more labeled training data is needed.

- **Schema Design and Refinement**

  The Graph Schema was developed iteratively, and further refinements will follow based on the developed pipelines, relying more on the SME and possible IAEA collaboration inputs.

- **Graph Pruning Cleaning and Merging**
  An on-premises instance of a Graph Database was used for graph storage and schema merging, enabling efficient deduplication and pruning.

A knowledge graph was built for each ingested document and the entire database will be combined in the future once the Graph Schema is finalized.

A GraphRAG approach (based on [8]) was developed to provide answers to the queries by synthesizing the semantic search results and the results from knowledge graph queries. The resulting knowledge graph was benchmarked against an SME-curated dataset which showed promising preliminary results.

Lessons Learned and Future Directions

The modern semantic technologies, including AI tools like LLMs, significantly simplify construction and utilization of knowledge management systems. However, as LLMs may produce hallucinations, the process must have a SME in the loop and be benchmarked. As such, an international collaboration is necessary to enable easier sharing of best practices but also the decommissioning-specific SME expertise necessary to benchmark the results.

The research undertaken in the project will continue through 2026 with emphasis on Graph Schema refinement, and benchmarking of the results. EPRI also plans to compare the results to existing Nuclear-specific chatbots and assess usefulness of semantic technologies, and transferability of the research undertaken here to other use cases within the nuclear domain.

References

[1] International Atomic Energy Agency (IAEA). (2021). *Exploring Semantic Technologies and Their Application to Nuclear Knowledge Management*. Vienna: IAEA. Retrieved from https://www.iaea.org/publications/13469/exploring-semantic-technologies-and-their-application-to-nuclear-knowledge-management

[2] https://neo4j.com/blog/knowledge-graph/what-is-knowledge-graph/

[3] International Atomic Energy Agency (IAEA). (2023). *A Taxonomy for the Decommissioning of Nuclear Facilities*. Vienna: IAEA. Retrieved from https://www.iaea.org/publications/15456/a-taxonomy-for-the-decommissioning-of-nuclear-facilities

[4] https://www.falkordb.com/blog/how-to-build-a-knowledge-graph/

[5] https://rel.relational.ai/rel/concepts/relational-knowledge-graphs/schema

[6] https://neo4j.com/blog/knowledge-graph/how-to-build-knowledge-graph/

[7] https://www.llamaindex.ai/blog/building-and-evaluating-a-qa-system-with-llamaindex-3f02e9d87ce1

[8] https://neo4j.com/blog/genai/what-is-graphrag/

[9] https://github.com/microsoft/graphrag