Ontology-Driven Knowledge Graph Construction from Text Using Large Language Models: A Nuclear Domain Case Study

Bokyeong Kim [a,b], Yonggyun Yu [a,b*]
[a]*Korea Atomic Energy Research Institute, 111, Daedeok-daero 989beon-gil, Yuseong-gu, Daejeon, 34057, Korea*
[b]*Korea National University of Science & Technology, 217, Gajeong-ro, Yuseong-gu, Daejeon, 34113, Korea*
[*]*Corresponding author: ygyu@kaeri.re.kr*

***Keywords:** Knowledge Graph, Large Language Models, Ontology, Text-to-Knowledge Graph*

1. Introduction

Artificial intelligence (AI) has expanded its application across various industrial sectors. In the nuclear domain, AI-based techniques are increasingly employed for equipment diagnostics, remaining useful life (RUL) prediction, and abnormal event identification [1]. The effectiveness of such domain-specific AI systems fundamentally depends on the availability of high-quality, structured data. However, the nuclear industry continues to face persistent data challenges. Critical operational knowledge—including equipment specifications, maintenance histories, safety analysis reports, and operating procedures—is predominantly recorded in unstructured text formats. Moreover, these data are managed in heterogeneous formats across different plants, simulators, and organizations, often requiring direct assistance from original data creators for accurate interpretation.

To address these limitations, efforts have been made to structure nuclear knowledge through the development of domain ontologies such as DIAMOND [2]. An ontology is a knowledge structure that systematically defines the types of entities, their properties, and the relationships among them within a specific domain. In the nuclear field, the constructed ontology explicitly specifies what kinds of equipment and components exist and how they are related. For example, a pump is defined as being composed of bearings and an impeller, while a reactor is defined as equipment that includes control rods. Such relationships are predefined in a structured manner within the ontology.

This structure provides a common reference framework that enables data generated from different plants or systems to be interpreted and connected under consistent criteria, allowing dispersed operational knowledge to be represented within a unified structure.

A Knowledge Graph (KG) operationalizes this ontology by instantiating its concepts and relations as (subject, relation, object) triples, thereby transforming dispersed textual knowledge into machine-interpretable structured data [3]. Recent advances in Large Language Models (LLMs) have demonstrated that, under ontological constraints, they can directly extract such triples from unstructured text [5].

In this study, we utilized a domain ontology developed by nuclear subject-matter experts and constructed a small-scale benchmark by selecting nuclear-related entities and triples from Wikidata that correspond to the ontology concepts. Based on this benchmark, we evaluated whether LLMs can generate knowledge graph triples from text while adhering to the constraints defined by the nuclear ontology.

The experimental results indicate that ontology-guided extraction can achieve meaningful performance even under data-limited conditions. These findings suggest that leveraging expert-curated nuclear ontologies may provide a viable pathway toward automated structuring and integration of nuclear technical knowledge.

2. Methods and Results

2.1 Nuclear Ontology Specification

DIAMOND (Data Integration Aggregated Model and Ontology for Nuclear Deployment) is a formal domain ontology developed by Idaho National Laboratory (INL) to establish a standardized vocabulary for nuclear power plant systems and components [2]. An ontology is a structured knowledge model that hierarchically defines the types of entities within a specific domain, along with their properties and interrelationships, enabling heterogeneous data to be interpreted and connected under a shared semantic framework. Such a structure plays a critical role in ensuring data consistency and interoperability in digital environments.

DIAMOND is expressed in OWL (Web Ontology Language) and comprises 615 classes spanning the full scope of nuclear power plant infrastructure. It was designed to support digital transformation in nuclear operations by enabling interoperable data exchange across maintenance, operational, and safety analysis systems.

The ontology is organized around several top-level structures. The Equipment branch includes major plant systems such as Reactor, Steam Generator, Turbine, Generator, Pump, and Heat Exchanger, while the Component branch defines constituent elements including Bearing, Seal, Impeller, Rotor, and Control Rod. Higher-level classes such as Rotating Machinery and Heat Exchanger function as generalized abstractions that group equipment sharing common characteristics. This hierarchical organization clarifies classification among plant systems and systematically represents structural relationships between equipment and their components.

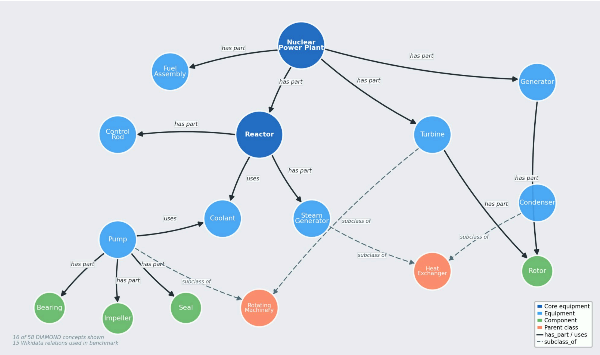

Fig. 1. Partial DIAMOND ontology hierarchy mapped to Wikidata. 16 of 58 selected concepts are shown with representative has_part, uses, and subclass_of relations used in the benchmark.

In this study, we constructed a benchmark by selecting only those DIAMOND ontology concepts that also exist as entities in Wikidata and are supported by sentence-level evidence in the TekGen corpus. This selection enabled evaluation of LLM-based triple extraction under conditions comparable to those of general-domain benchmarks.

2.2 Benchmark Construction Pipeline

Nuclear-related triples were first collected from Wikidata and filtered to retain only those whose concepts and relations are defined in the DIAMOND ontology. The retained triples were treated as ground-truth triples for evaluation. For each ground-truth triple, a supporting sentence was retrieved from the TekGen corpus to construct evaluation instances. Each instance consists of the ontology specification, an input sentence, and the corresponding ground-truth triple. Triples containing entities or relations not defined in the ontology were excluded to ensure that the evaluation strictly operates within the predefined ontological constraint space.

The benchmark was constructed by adapting the prompt-based Text-to-Knowledge-Graph generation setting introduced in Text2KGBench [4] to the nuclear domain. Specifically, the evaluation framework retains the core structure of aligning sentences with ontology-constrained triples while restricting generation to predefined concepts and relations, but replaces the general-domain ontology with the DIAMOND nuclear ontology.

During inference, the model is provided only with (1) a description of the ontology, including the permitted concepts and relations, (2) a small number of in-context sentence–triple examples, and (3) the test sentence. The model has no access to external knowledge graphs or structured databases. It must generate triples solely based on the input sentence while conforming to the ontology-defined concepts, relations, and structural constraints.
The predicted triples are compared against the ground-truth triples, and performance is evaluated using ontology-aware metrics, including ontology conformance and hallucination rates.

2.3 Experimental Setup

Six open-source LLMs were evaluated under identical inference settings. The models ranged from 8B to 32B parameters. For each model, the prompt comprised the nuclear ontology context—specifically, the list of permitted concepts and relations—along with one demonstration example in the form of a sentence–triple pair and the test sentence. No fine-tuning was performed. Evaluation was conducted on 89 test sentences.

To assess ontology compliance and hallucination behavior, we employed ontology-aware evaluation metrics. Ontology Conformance (OC) is defined as the proportion of generated triples whose relations belong to the predefined ontology relation set. A triple is considered ontology-conforming if its relation matches one of the canonical relations specified in the ontology.

Relation Hallucination (RH) is defined as the proportion of generated triples whose relations are not included in the ontology relation set. Under this formulation, relation hallucination is complementary to ontology conformance.

Subject Hallucination (SH) and Object Hallucination (OH) quantify the proportion of generated triples whose subject or object, respectively, does not appear in either the input sentence or the ontology concept set. A subject or object is regarded as hallucinated if it cannot be identified in the input sentence or among the ontology-defined concepts.

All metrics are computed at the sentence level and macro-averaged over the test set. If no triples are generated for a given sentence, the metric value for that instance is set to zero.

2.4 Results

Table I summarizes the performance of all six models.

Table I: Triple extraction performance comparison

Model	Params	OC	RH	SH	OH
Qwen3-14B	14B	0.998	0.002	0.094	0.060
Llama3.1-8B	8B	0.934	0.067	0.032	0.071
Qwen3-32B	32B	1.000	0.000	0.138	0.093
DeepSeek-R1-32B	32B	0.988	0.012	0.067	0.069
Qwen2.5-14B	14B	0.942	0.058	0.292	0.294
Phi-4	14B	0.951	0.050	0.117	0.134

Several models achieved high Ontology Conformance scores. All models recorded OC values above 0.93, and Qwen3-32B achieved a perfect OC score of 1.000. These results indicate that most models generated triples using only relations defined in the ontology.

In contrast, greater variation was observed in hallucination metrics. Llama3.1-8B exhibited relatively low subject and object hallucination rates, with SH equal to 0.032 and OH equal to 0.071. By comparison, Qwen2.5-14B showed the highest values, with SH equal to 0.292 and OH equal to 0.294. Although Qwen3-32B achieved perfect ontology conformance at the relation level, it still recorded SH of 0.138 and OH of 0.093. This indicates that even when relations strictly follow the ontology, models may still generate entities that are not grounded in the input sentence. In other words, adherence to relation-level constraints and reduction of entity-level hallucination represent different aspects of model behavior.

Another limitation in the nuclear domain is data sparsity. Although several thousand nuclear-related triples were retrieved from Wikidata, fewer than 10 percent could be aligned with sentence-level evidence in the TekGen corpus. This limited availability of aligned textual data directly constrains the evaluation scale and highlights the need for domain-specific corpus development in future work. Consequently, the OC, RH, SH, and OH values reported in this study are based on only 89 test sentences, and caution is needed when generalizing these results to broader ontology coverage or different sentence distributions.

3. Conclusions

This study applies the Text2KGBench benchmark to the nuclear domain using the DIAMOND ontology. DIAMOND concepts were mapped to corresponding Wikidata entities to collect nuclear-related triples, and supporting sentences were identified from the TekGen benchmark corpus. Through this mapping and alignment process, a dataset of 332 sentence–triple pairs was constructed.

Six open-source LLMs were evaluated under a 1-shot prompting setting. All models achieved Ontology Conformance scores above 0.93, and one model reached perfect conformance. This indicates that the generated relations were largely restricted to those defined in the ontology. However, subject and object hallucination rates varied across models. Even models with perfect relation-level conformance still produced entity-level hallucinations, suggesting that relation constraint adherence and entity grounding are distinct aspects of model behavior.

Data sparsity remains a key limitation in the nuclear domain. Fewer than 10 percent of the retrieved nuclear-related triples could be aligned with sentence-level evidence in TekGen. This limited alignment restricts evaluation scale and indicates the need for expanded domain-specific corpora in future work. As the final evaluation relies on only 89 sentences, the generalizability of the OC, RH, SH, and OH results to a wider range of ontology concepts and relation types remains to be validated.

Acknowledgement

This study was supported by the Substantiation Support Program through the Korea Innovation Foundation, funded by the Ministry of Science and ICT (No. 76170-26).

REFERENCES

[1] Y. Fayyaz, W. Elouataoui, Y. Gahi, K. El-Khatib, G. Harvel, and K. Sankaranarayanan, “Natural Language Processing in the Nuclear Industry: Opportunities and Challenges,” Nuclear Technology, 2025, doi:10.1080/00295450.2025.2481358.

[2] A. Al Rashdan, J. Browning, and C. Ritter, “DIAMOND: Data Integration Aggregated Model and Ontology for Nuclear Deployment,” Idaho National Laboratory, 2020.

[3] A. Hogan, E. Blomqvist, M. Cochez, C. D’Amato, G. Decker, C. Gutierrez, S. Kirrane, J. E. Labra Gayo, R. Navigli, S. Neumaier, A.-C. Ngonga Ngomo, A. Polleres, S. Rashid, A. Rula, and A. Schmelzeisen, “Knowledge Graphs,” ACM Computing Surveys, Vol. 54, No. 4, pp. 1-37, 2021.

[4] N. Mihindukulasooriya, M. T. Alshaikh, and A. T. Abdelrahman, “Text2KGBench: A Benchmark for Ontology-Driven Knowledge Graph Generation from Text,” Proc. ISWC 2023, pp. 247-265, 2023.

[5] S. Pan, L. He, Y. Yu, J. Chen, and H. Liu, “Unifying Large Language Models and Knowledge Graphs: A Roadmap,” IEEE Transactions on Knowledge and Data Engineering, Vol. 36, No. 7, pp. 3580-3599, 2024.