

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

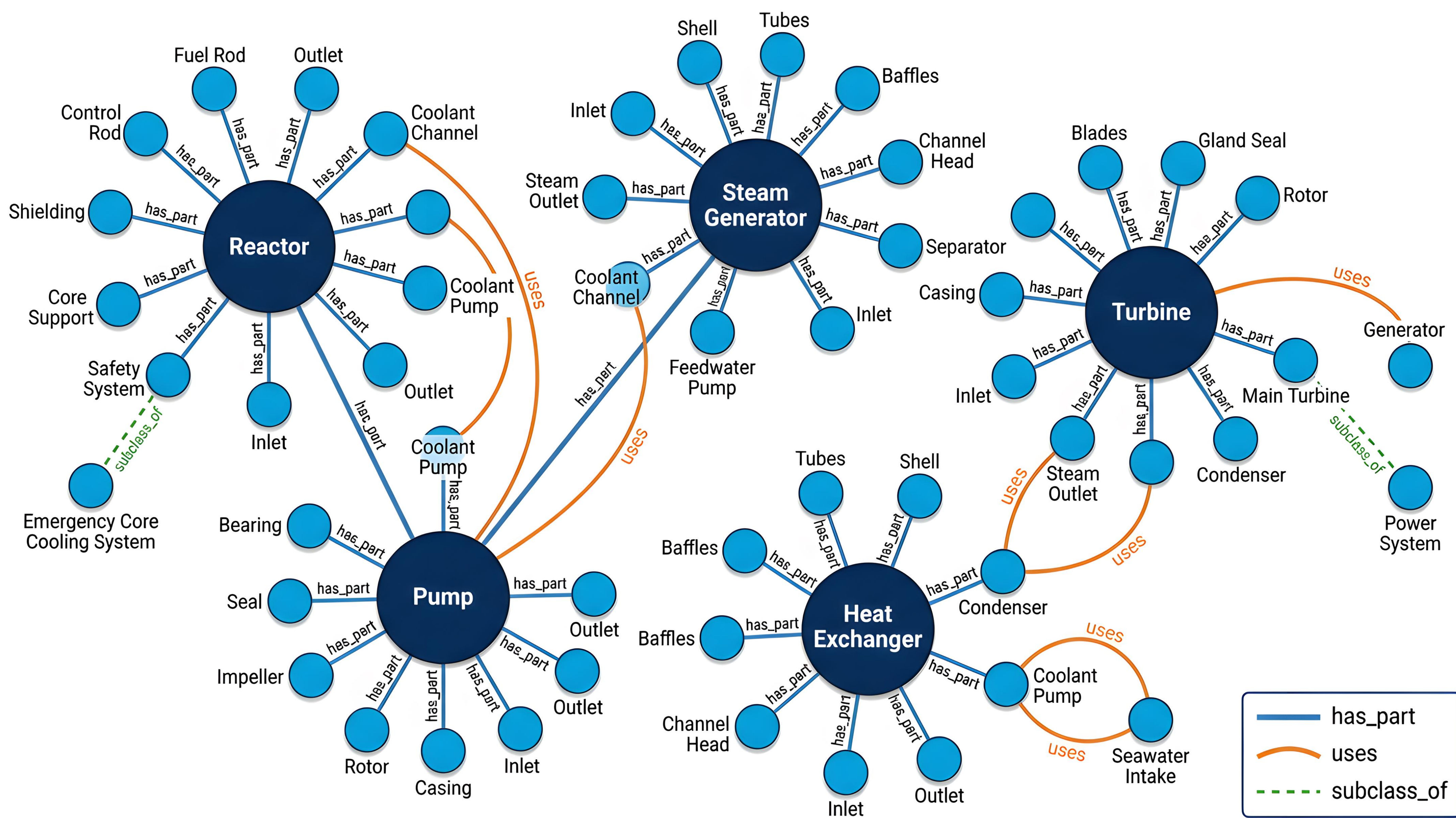
Bokyeong Kim^{a,b}, Yonggyun Yu^{a,b*}

^aKorea Atomic Energy Research Institute ^bUniversity of Science & Technology

*Corresponding author : ygyu@kaeri.re.kr

Introduction

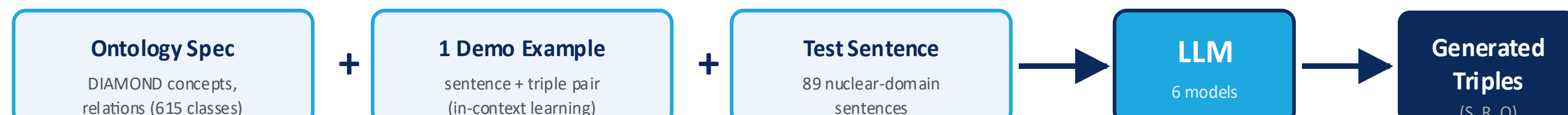
In the nuclear domain, operational knowledge such as maintenance records, safety analysis reports, and operating procedures is stored in unstructured text and distributed across plants in heterogeneous formats. This fragmentation limits AI systems that require structured, interoperable data. We address this by pairing the DIAMOND nuclear ontology with Large Language Models (LLMs) to extract (subject, relation, object) triples directly from text, producing a Knowledge Graph (KG) that conforms to expert-defined schema constraints. We evaluate six open-source LLMs (8B to 32B parameters) on a benchmark of 332 sentence-triple pairs derived from Wikidata and TekGen, under a 1-shot prompting setting without fine-tuning.



- We adopt DIAMOND, an OWL ontology with 615 classes covering nuclear systems (Reactor, Steam Generator, Turbine, Generator, Pump, Heat Exchanger) and components (Bearing, Seal, Impeller, Rotor, Control Rod).
- We align DIAMOND-conformant Wikidata triples with TekGen sentences, yielding 332 evaluation pairs (adapting Text2KGBench to the nuclear domain).
- Four ontology-aware metrics assess outputs: Ontology Conformance (OC) and Relation Hallucination (RH) for relation-level adherence; Subject and Object Hallucination (SH, OH) for entity grounding.

Experimental Setup

1-Shot Prompting Structure



We evaluate six open-source LLMs (Qwen3-14B, Qwen3-32B, Qwen2.5-14B, Llama3.1-8B, DeepSeek-R1-32B, Phi-4) ranging from 8B to 32B parameters. Each model receives a single prompt containing the ontology specification, one sentence-triple demonstration, and the test sentence. No fine-tuning is performed. Evaluation uses 89 held-out sentences.

Conclusions

- Constructed 332-pair nuclear KG benchmark via DIAMOND × Wikidata × TekGen.
- Six LLMs achieved OC > 0.93 in a 1-shot setting; one reached 1.000.
- Entity-level hallucination did not track relation conformance, leaving entity grounding as an open problem.
- Fewer than 10% of retrieved triples aligned to TekGen sentences, identifying corpus expansion as the main bottleneck.

References

- [1] Y. Fayyaz et al., "NLP in the Nuclear Industry," Nuclear Technology, 2025.
- [2] A. Al Rashdan et al., "DIAMOND: Data Integration Aggregated Model and Ontology for Nuclear Deployment," INL, 2020.
- [3] A. Hogan et al., "Knowledge Graphs," ACM CSUR, vol. 54, no. 4, 2021.
- [4] N. Mihindukulasooriya et al., "Text2KGBench," ISWC, 2023.
- [5] S. Pan et al., "Unifying LLMs and Knowledge Graphs," IEEE TKDE, 2024.

Acknowledgements

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

Results

Table 1. Triple extraction performance across six open-source LLMs (1-shot, 89 test sentences).

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

- **All OC > 0.93:** relation-level constraints reliably respected
- **Qwen3-32B:** perfect OC, yet SH = 0.138, OH = 0.093 (entity grounding is a separate axis)
- **< 10%** of retrieved nuclear triples align with TekGen, identifying corpus scarcity as the bottleneck

Subject vs Object Hallucination by Model

