Ontology-Based Prompt Engineering for Text-to-SPARQL: A Case Study on Tuberculosis in Wikidata

Fariz Darari

Faculty of Computer Science, Universitas Indonesia, Depok 16424, Indonesia

Abstract
This study explores whether a few-shot, ontology-based prompt engineering approach can improve text-to-SPARQL generation in the health domain. Using tuberculosis (TB) in Wikidata as a case study, we analyze how TB knowledge is represented from an ontological perspective and embed this knowledge into prompts as ontological terms, few-shot examples, and query templates. We evaluate this ontology-guided strategy against a naive prompting method using small language models on a curated set of natural language questions about TB. Our results show that the ontology-based approach significantly improves answer precision. A qualitative analysis offers key insights and lessons learned from this approach. These preliminary findings highlight the potential of ontology-based prompt engineering and motivate further evaluation across different domains.

1. Introduction

Tuberculosis (TB) remains one of the world's leading infectious diseases, primarily affecting the lungs and spreading through airborne transmission. While TB is both preventable and curable, it continues to represent a major global health burden, with about a quarter of the world's population estimated to be infected with TB bacteria. Of those infected, 5—10% will eventually develop active disease and show clinical symptoms [1]. The distribution of TB is also highly uneven, with five countries accounting for more than half of global cases in 2022: India (26%), Indonesia (10%), China (6.8%), the Philippines (6.8%), and Pakistan (6.3%) [2].

The situation is exacerbated by poor patient knowledge about various aspects of TB. One study demonstrated that a structured health-education programme significantly improved patient understanding of TB, reinforcing the critical role of healthcare providers in disease management, control, and prevention [3]. These findings highlight the persistent challenges of TB awareness and the continuing need for innovative approaches to information delivery.

Meanwhile, large language models (LLMs) are becoming transformative tools in digital health. Recent work shows that LLMs, such as GPT-4, substantially outperform conventional virtual assistants (VAs) in delivering accurate and clinically relevant information, even in complex domains like mental health [4]. Our focus, however, is on the complementary role of small

16th International Conference on Biological and Biomedical Ontology (ICBO 2025) — ICBO EAST, November 1, 2025, Nara, Japan
fariz@ui.ac.id (F. Darari)
0000-0001-6025-609X (F. Darari)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-04-17

language models (SLMs), which, despite having more limited capacity, offer key advantages of low cost, lower computational requirements, and broader accessibility across diverse hardware and user groups [5]. Importantly, wider adoption of SLMs can advance equity in global health by enabling researchers, clinicians, and patients in low- and middle-income countries to access advanced digital health tools without requiring high-end infrastructure. This positions SLMs as an inclusive and practical option for integrating knowledge-driven conversational systems into healthcare, particularly in resource-constrained settings.

In this context, Wikidata[1] offers a promising foundation for structured biomedical knowledge integration. As an open, multilingual knowledge graph that supports Wikipedia and other Wikimedia projects, Wikidata provides standardized, linked, and freely reusable data [6]. Its structured format allows for both human- and machine-readable access, enabling interoperability with external datasets and serving as a knowledge base for downstream applications.

Building on these developments, Wikidata can also serve as a structured knowledge graph underlying SLMs to support knowledge-driven question answering about TB. This paper aims to:

1. Explore how tuberculosis is represented in Wikidata from an ontological perspective.
2. Assess how this ontological knowledge can support text-to-SPARQL question answering tasks related to TB.

Prior work on text-to-SPARQL and knowledge-driven question answering commonly leverages LLMs augmented with explicit knowledge graph context, such as RDF subgraphs, selected graph fragments, example queries, or ontology elements, to guide query generation and reduce hallucination [7, 8, 9]. These approaches consistently emphasize the importance of supplying structured KG-based context either through prompt augmentation or retrieval pipelines to improve the accuracy and interpretability of SPARQL generation. In contrast to these general frameworks, this paper focuses on the ontological representation of TB in Wikidata and analyzes its role in supporting text-to-SPARQL question answering within the TB domain.

2. Tuberculosis Representation in Wikidata

This section summarizes how tuberculosis (TB) is modeled in Wikidata, highlights the core ontological properties used to represent diseases, and demonstrates how these properties enable concise SPARQL queries for knowledge-driven question answering.

2.1. Overview and Key Entities

In Wikidata, tuberculosis (Q12204) is a disease that primarily affects the human lung (Q2640512) or lung (Q7886). A related condition is latent tuberculosis (Q4254929), in which the infection remains inactive in the body. The disease is caused by the bacterium *Mycobacterium tuberculosis* (Q130971) and is most commonly spread through airborne transmission (Q11986959).

The diagnosis of tuberculosis often involves imaging techniques such as the chest radiograph (Q1283318). A frequent clinical manifestation of the disease is cough (Q35805). The medical

[1] http://www.wikidata.org

specialty primarily responsible for its treatment and management is pulmonology (Q203337). Treatment typically includes the use of drugs such as isoniazid (Q423169), while prevention is supported by the BCG vaccine (Q798309).

Globally, tuberculosis represented a significant epidemiological burden with an estimated 10,400,000 cases in 2016 worldwide. An important historical figure who died from tuberculosis is the Indonesian military leader General Soedirman (Q2367329). Tuberculosis has also been the subject of scientific literature, including the article The tuberculosis challenge in a rural South African HIV programme (Q33530558).

Efforts to control and eliminate tuberculosis are supported by international and national initiatives, such as the Stop TB Partnership (Q7619759), the National TB Elimination Program (Q7277279), and the annual observance of World Tuberculosis Day (Q518978). Finally, tuberculosis is indexed in external classification systems, notably the ICD-10 (Q45127), where it is assigned the codes A15–A19.

2.2. Core Ontological Properties

As described above, a disease (e.g.,) can be represented in Wikidata using the following properties (among others):

- **instance of (P31):** type to which this subject corresponds/belongs
- **subclass of (P279):** this item is a subclass (subset) of that item
- **has cause (P828):** underlying cause; entity that ultimately resulted in this effect
- **health specialty (P1995):** main specialty relevant to this medical condition or treatment
- **symptoms and signs (P780):** possible symptoms or signs of a medical condition
- **medical examination (P923):** examinations that might be used to diagnose and/or prognose the condition
- **drug or therapy used for treatment (P2176):** drug, procedure, or therapy used to treat the condition
- **disease transmission process (P1060):** process by which a pathogen is transmitted
- **number of cases (P1603):** cumulative number of confirmed, probable, and suspected occurrences
- **ICD-10 ID (P494):** identifier in the ICD Terminology of Diseases — Version 10

2.3. Illustrative SPARQL Queries

These ontological properties enable the construction of targeted SPARQL queries. Below are several examples related to tuberculosis.

SPARQL Query

```
# What is the cause of tuberculosis?

SELECT ?causeLabel
WHERE {
  wd:Q12204 wdt:P828 ?cause .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

SPARQL Query

```
# How is tuberculosis transmitted?

SELECT ?transmissionLabel
WHERE {
  wd:Q12204 wdt:P1060 ?transmission .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

SPARQL Query

```
# What is the cause of death of General Soedirman?

SELECT ?causeLabel
WHERE {
  wd:Q2367329 wdt:P509 ?cause .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en" }
}
```

3. Ontology-Based Prompt Engineering

Prompt engineering is an emerging field focused on designing and optimizing prompts for language models (LMs). It helps researchers improve model performance on tasks such as question answering and reasoning, while enabling developers to build reliable interaction techniques [10].

An ontology provides a shared vocabulary and machine-interpretable definitions of concepts and their relationships within a domain. It supports knowledge sharing, reuse, and explicit representation of assumptions, while separating domain knowledge from operational details. Ontology engineering follows key principles: there is no single correct model, development is iterative, and concepts should align closely with real-world objects and relationships [11].

In this section, we combine the two, creating ontology-based prompt engineering.

3.1. Prompt

The Ontology-Based Text-to-SPARQL (Wikidata) prompt as shown below[2] is designed to translate natural-language questions into valid SPARQL queries. The translation process is explicitly grounded in the ontological properties (PIDs) and entity mappings (QIDs) available in Wikidata, ensuring that the generated queries are syntactically correct and semantically aligned with the structure of the knowledge graph. By relying on standardized ontological relations, the prompt enables consistent and reproducible knowledge retrieval for biomedical and related domains.

The prompt makes use of four key parameters: the natural-language question (question), property to PID mappings (property_mappings) the human-readable entity name (entity_name), and its corresponding Wikidata identifier (qid). Together, these inputs provide the necessary

[2]The full version is available at: https://github.com/fadirra/icbo25-text-to-sparql/blob/main/text-to-sparql.prompt

context for query construction. The entity name and QID anchor the query to the relevant Wikidata item, while the natural-language question guides the choice of ontological property that defines the relation of interest.

The instructions within the prompt define strict output constraints. The model is required to generate only the SPARQL query, without any explanatory text, and to always include the Wikidata label service to ensure human-readable results. Concise variable names (e.g., ?causeLabel, ?symptomLabel) are preferred to improve readability. The query structure is expected to follow provided examples, which demonstrate how different natural-language formulations map onto specific Wikidata properties.

In practice, the model identifies the entity's QID, selects the relevant property that matches the relation implied in the question, and composes a minimal triple pattern that captures the semantics of the query. Examples in the prompt reinforce this grounding. For instance, the question "What is the cause of X?" is linked to the property has cause (P828), while "How is X transmitted?" corresponds to disease transmission process (P1060). In both cases, qid substitutes for the entity in the final query.

Overall, this ontology-based prompt provides a structured and controlled framework for text-to-SPARQL generation. By explicitly linking natural-language inputs to QIDs and PIDs, it reduces ambiguity, ensures syntactic validity, and supports the systematic construction of queries for knowledge-driven question answering over Wikidata.

3.2. Implementation Steps

The text-to-SPARQL pipeline was implemented in several modular steps. First, an *entity extraction* component identifies the main entity of interest from the question using a large language model guided by a JSON-based extraction prompt. The output provides a semantic scaffold for aligning the question to Wikidata's ontology and supplies the entity name for the next step.

Second, an *entity linking* function queries Wikidata's `wbsearchentities` API to map the extracted entity name to its corresponding Wikidata identifier (QID). This step anchors questions to specific items in the knowledge graph, ensuring precision in downstream query generation.

Third, we set up a *vector database of Wikidata properties* using ChromaDB. Property labels and descriptions were embedded with a sentence-transformer model and indexed. At query time, the system retrieves the top-k most relevant properties to the input question via vector similarity search. This step grounds property selection in both linguistic similarity and ontological alignment.

Finally, the *question answering module* integrates these components to produce executable SPARQL queries. Given a question, it: (i) extracts the entity name, (ii) resolves it to a QID, (iii) retrieves candidate properties via the vector database, and (iv) instantiates an ontology-based SPARQL prompt with the question, entity name, QID, and candidate properties. The language model then generates a syntactically valid SPARQL query, which is executed against the Wikidata endpoint using the `SPARQLWrapper` library. Results are returned as lists of answers, enabling direct comparison with ground-truth outputs.

Prompt Example

```
PROMPT TITLE: Ontology-Based Text-to-SPARQL (Wikidata)

ROLE
You translate natural-language questions into syntactically valid SPARQL queries
for Wikidata, grounded in an explicit ontology (properties/relations)
and entity mappings.

INPUTS
# Natural language question: {question}

# QID mapping (Wikidata entity of interest)
- `{entity_name}`: {qid}

# PID mapping (Wikidata properties or relations)
- `has cause`: P828
- `disease transmission process`: P1060
- `cause of death`: P509
# META: Other properties are redacted due to space reasons
{property_mappings}

OUTPUT CONSTRAINTS
- Output ONLY a valid SPARQL query (no prose, no explanations).
- Always include the label service:
  SERVICE wikibase:label {{ bd:serviceParam wikibase:language "en" }}
- Prefer concise SELECT variables (?causeLabel, ?symptomLabel, etc.).
- The output SPARQL query must be inspired from the query EXAMPLES below,
both for the structure, ontological terms, and form.

PROCEDURE (high level)
- The translation typically relies on entities (QID) and ontological relations or
properties (PID) in the questions.
- Select an appropriate query pattern: the triple patterns of
subject, predicate, and object; directions; and variables.
- Compose a minimal SPARQL query with label service;
avoid extraneous clauses unless required.
- Ensure syntactic validity.

EXAMPLES (with ontological terms to reinforce grounding), assume that QIDX is
the QID of the entity X.
# Natural language: What is the cause of X?
   Ontological terms involved: has cause (P828)
   SPARQL:
   SELECT ?causeLabel WHERE {{
     wd:QIDX wdt:P828 ?cause .
     SERVICE wikibase:label {{ bd:serviceParam wikibase:language "en" }}
   }}

# Natural language: How is X transmitted?
   Ontological terms involved: disease transmission process (P1060)
   SPARQL:
   SELECT ?transmissionLabel WHERE {{
     wd:QIDX wdt:P1060 ?transmission .
     SERVICE wikibase:label {{ bd:serviceParam wikibase:language "en" }}
   }}

# META: Other examples are redacted due to space reasons

FINAL INSTRUCTION
Now produce ONLY the SPARQL query for:
Q: {question}
A:
```

Table 1
Jaccard similarity results (30 questions) and qualitative analysis for text-to-SPARQL generation without vs. with ontology-based prompting.

Model/Setting	Jaccard	Correct	Qualitative insights
Naive (no ontology-based prompt)	0%	0/30	Failed entirely without structured ontology guidance.
llama3.2:3b (ontology-based)	40%	12/30	Frequent and diverse errors reflecting capacity limits.
mistral:7b (ontology-based)	43%	13/30	Slightly better overall compared to llama3.2:3b and llama3.1:8b.
llama3.1:8b (ontology-based)	36%	11/30	Despite its larger size, performance remains modest.
gpt-oss:20b (ontology-based)	80%	24/30	Highest accuracy; remaining failures are dominated by entity recognition/linking errors and property misalignment.

4. Results and Discussion

4.1. Experimental Setup

We evaluate small language models (SLMs) with at most 20 billion parameters to assess the impact of ontology-based prompt engineering. Specifically, we test four locally hosted models via Ollama:[3] `gpt-oss:20b`, `llama3.1:8b`, `llama3.2:3b`, and `mistral:7b`. All experiments run on a laptop equipped with an Intel Iris Xe graphics adapter (128 MB), 40 GB RAM, and an Intel Core i7-1355U processor (1.7 GHz).

The evaluation uses 30 natural-language questions derived from 10 tuberculosis seed questions. For each question, we compare predicted answer sets with gold answers using Jaccard similarity, where 1.0 denotes a perfect match.

Our code, prompt, and dataset are available at: https://github.com/fadirra/icbo25-text-to-sparql

4.2. Experiment Results

We summarize the overall performance of the evaluated models (see Table 1). The naive approach without ontology-based prompt engineering failed entirely, achieving a Jaccard similarity of 0% (0 out of 30). Incorporating ontology-based prompting substantially improved performance across models. The smallest model, `llama3.2:3b`, reached 40% similarity (12/30), while `mistral:7b` slightly outperformed it with 43% (13/30). Interestingly, `llama3.1:8b`, despite its larger size, reached only 36% similarity (11/30). The best results were obtained with `gpt-oss:20b`, which achieved 80% similarity (24/30), demonstrating that larger SLMs can significantly benefit from ontology-based guidance.

Despite strong overall performance, qualitative inspection reveals characteristic error patterns across models. For `gpt-oss:20b`, errors were mainly due to entity recognition and property

[3] https://ollama.com/

misalignment. For example, the question *"Which condition led to General Soedirman's death?"* failed because the entity linking step did not correctly resolve the subject. Similarly, *"The Mantoux test is used to diagnose which disease?"* was misinterpreted: instead of selecting the correct property *medical examination (P923)*, the model incorrectly used *positive diagnostic predictor for (P3356)*. In addition, *"Which event is held in commemoration of tuberculosis?"* failed because the entity search returned the event *World Tuberculosis Day* directly, bypassing the tuberculosis entity that should have been queried.

For `llama3.2:3b`, errors were more frequent and diverse, underscoring the capacity limits of very small models. Several questions failed due to overly elaborate query construction, such as *"In what ways is tuberculosis spread?"* and *"How does tuberculosis commonly present?"*, where the model produced unnecessarily multi-triple patterns far beyond what was required. Other questions suffered from property misalignment, e.g., *"What did General Soedirman die of?"*, where the model incorrectly mapped to *has cause (P828)* instead of the correct *cause of death (P509)*. Entity recognition errors were also common: in *"The Mantoux test is used to diagnose which disease?"*, the entity search returned the disease *tuberculosis* directly, while in *"Which event is held in commemoration of tuberculosis?"*, the model should have linked to *tuberculosis*, not *World Health Day* (as output by the system). Finally, in *"Which scholars are active in research on tuberculosis?"* and *"The painting of X depicts what disease?"*, the generated queries failed due to excessive complexity or malformed `SERVICE` clauses.

These results indicate that ontology-based prompt engineering significantly boosts the reliability of text-to-SPARQL generation, even for relatively small models. Larger models (though still considered small), such as `gpt-oss:20b`, achieve much higher accuracy but remain vulnerable to entity linking and property selection errors. Smaller models, while benefiting from ontology-based guidance, still tend to over-generate or misalign properties, underscoring the importance of combining structured prompts with robust entity linking strategies.

5. Conclusions

This paper investigated ontology-based prompt engineering for text-to-SPARQL generation in the health domain, using tuberculosis (TB) in Wikidata as a case study. We first analyzed how TB is represented from an ontological perspective, identifying key entities and properties relevant for biomedical question answering. Building on this representation, we designed a structured prompt that explicitly incorporates QIDs and PIDs, and implemented a pipeline combining entity linking, property retrieval via vector search, and SPARQL query construction.

Our experiments compared ontology-based prompts with a naive prompting approach across four small language models (`llama3.2:3b`, `llama3.1:8b`, `mistral:7b`, and `gpt-oss:20b`). Results showed that ontology-based prompting yields substantial gains in accuracy: the naive approach failed completely, while ontology-guided models achieved up to 80% Jaccard similarity. Qualitative analysis revealed that larger models benefit most from the structured guidance, though common errors persist in entity recognition and property alignment. Smaller models, while improved by ontology-based prompts, often generated overly complex or malformed queries.

Overall, this study demonstrates that explicitly grounding prompts in ontological knowledge significantly improves the reliability of text-to-SPARQL generation. The findings highlight both the promise and the limitations of small language models for biomedical question answering. Future work should scale evaluation to larger, more diverse datasets and richer, multi-hop question sets, extend to additional domains, integrate more robust entity-linking systems, and explore hybrid strategies that combine symbolic ontologies with language-model reasoning.

Declaration on Generative AI

ChatGPT was used primarily for initial drafting and brainstorming. All content was reviewed and edited by the author, who takes full responsibility for the final manuscript.

References

[1] WHO, Tuberculosis, Online, 2025. URL: https://www.who.int/news-room/fact-sheets/detail/tuberculosis, accessed: 2025-09-05.

[2] WHO, Global tuberculosis report, Online, 2024. URL: https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports, accessed: 2025-09-05.

[3] K. Yousif, M. Ei Maki, R. K. Babikir, H. Abuaisha, The effect of an educational intervention on awareness of various aspects of pulmonary tuberculosis in patients with the disease, Eastern Mediterranean Health Journal 27 (2021) 287–292. doi:10.26719/emhj.20.102.

[4] E. Sezgin, Redefining virtual assistants in health care: The future with large language models, Journal of Medical Internet Research 26 (2024) e53225. URL: https://www.jmir.org/2024/1/e53225. doi:10.2196/53225.

[5] Q. Zhang, Z. Liu, S. Pan, The rise of small language models, IEEE Intelligent Systems 40 (2025) 30–37. doi:10.1109/MIS.2024.3517792.

[6] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Commun. ACM 57 (2014) 78–85. doi:10.1145/2629489.

[7] L. Kovriguina, R. Teucher, D. Radyush, D. Mouromtsev, SPARQLGEN: One-Shot Prompt-based Approach for SPARQL Query Generation, in: N. Keshan, S. Neumaier, A. L. Gentile, S. Vahdati (Eds.), SEMANTiCS (Posters & Demos), volume 3526 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3526/paper-08.pdf.

[8] C. V. S. Avila, V. M. Vidal, W. Franco, M. A. Casanova, Experiments with text-to-SPARQL based on ChatGPT, in: 2024 IEEE 18th International Conference on Semantic Computing (ICSC), 2024, pp. 277–284. doi:10.1109/ICSC59802.2024.00050.

[9] J. G. Ongris, E. Tjitrahardja, F. Darari, F. J. Ekaputra, FrOG: Framework of Open GraphRAG, in: S. Tiwari, N. Mihindukulasooriya, J. D'Souza, F. Osborne, E. Marx (Eds.), Joint Proceedings of the Fourth International Workshop on LLM-Integrated Knowledge Graph Generation From Text (TEXT2KG) and the Second International BiKE Challenge, volume 4020 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2025, pp. 116–134. URL: https://ceur-ws.org/Vol-4020/paper-07.pdf.

[10] V. Pawar, M. Gawande, A. Kollu, A. S. Bile, Exploring the potential of prompt engineering: A comprehensive analysis of interacting with large language models, in: 2024 8th Interna-

tional Conference on Computing, Communication, Control and Automation (ICCUBEA), 2024, pp. 1–9. doi:10.1109/ICCUBEA61740.2024.10775016.

[11] N. Noy, D. McGuinness, et al., Ontology development 101: A guide to creating your first ontology, 2001. URL: http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html.