Ontology Corpora for LLM-based Knowledge Engineering Research

Guntur Budi Herwanto*, Stefani Tsaneva and Marta Sabou

Institute of Data, Process and Knowledge Management, Vienna University of Economics and Business, Austria

Abstract

Generative AI (GenAI) solutions are likely to have a profound impact on the Knowledge Engineering (KE) field. Considerable research is needed to understand the extent to which various KE tasks can be performed with GenAI, how the performance of these tasks compares to human baselines, and how to effectively adapt KE workflows to make the best use of GenAI methods. To conduct such research, there is a need of collections of corpora of ontologies with a range of diverse characteristics to support systematic experimentation covering a broad variety of ontology types. We propose collecting such corpora and describe our ongoing efforts to collect ontologies created by students, as representative for the work of junior ontology engineers (beginners level knowledge engineering skills). We also create an ontology analysis workflow to extract key metadata from ontologies and associated reports, which we share with the community.

Keywords

ontology corpus, ontology engineering, student authored ontologies, ontology processing workflow

1. Context and Research Need

Recent reports of the success of Generative AI techniques, and in particular, large language models (LLMs) on several natural language processing and other tasks [1, 2] have inspired research on exploring how these techniques could impact Knowledge Engineering (KE). Indeed, LLMs have been explored for a variety of knowledge engineering task including: (i) the creation or completion of semantic resources such as ontologies and knowledge graphs, e.g., [3, 4, 5, 6, 7, 8]; (ii) the evaluation of semantic resources in terms of error detection [9, 10, 11, 8] and error correction [7]; (iii) ontology alignment [12], as well as (iv) competency question generation and testing [13, 14, 15].

For KE as a field, several exciting research questions await answering, just to name a few:

- *To what extent and with which performance level can LLMs perform KE tasks?* As apparent from the very brief literature overview in the introductory paragraph above, investigation of this question has been pioneered for several KE tasks already. The community would further benefit from: (i) identifying the right granularity of KE tasks when investigating the use of LLMs; and subsequently (ii) agreeing on a catalogue of such tasks that can then be systematically investigated.
- *How does LLM performance compare to human performance?* In KE, there are significant differences between the characteristics of information artifacts created by engineers with diverse levels of expertise. Therefore, interesting insights could be won by comparing against knowledge artifacts created by contributors with various levels of expertise starting from junior experts (such as students) to experienced ontology engineers.
- *How can LLMs be best included in KE workflows?* This requires answering the previous research questions, i.e., understanding how well LLMs perform on individual KE tasks, which of these tasks can be fully covered by them and which still require human involvement.

ISWC 2024 Special Session on Harmonising Generative AI and Semantic Web Technologies, November 13, 2024, Baltimore, Maryland

*Corresponding author.
guntur.budi.herwanto@wu.ac.at (G. B. Herwanto); stefani.tsaneva@wu.ac.at (S. Tsaneva); marta.sabou@wu.ac.at (M. Sabou)
0000-0003-0250-6884 (G. B. Herwanto); 0000-0002-0895-6379 (S. Tsaneva); 0000-0001-9301-8418 (M. Sabou)

- *How to support KE education with LLMs?* Could LLMs give reliable feedback on modeling errors typical for junior ontology engineers? How can they be integrated in the process of teaching ontology engineering?
- *What are cognitive differences between humans and LLMs during KE?* Which human cognitive processes occurring during KE tasks are best reproducible with LLMs? And in which case do LLMs function fundamentally differently from human agents?

Research Need. A pre-requisite for answering research questions such as the ones above in a systematic manner is the availability of corpora of ontologies that vary across several key dimensions in order to offer a broad coverage of possible ontology types. For example, ontologies should differ in terms of the domains they cover, their size, their logical complexity, as well as the expertise level of their authors.

2. Proposed Approach: collecting Ontologies created by Junior Ontology Engineers

To partially address the research need described above we propose to collect ontologies created by junior ontology engineers, for example, students, as part of their training in ontology/knowledge engineering courses. Such ontologies are valuable as examples of domain models created by junior experts. As such, they could shed light on typical modeling errors made by beginners as a basis to assess whether LLMs make similar mistakes as junior experts, as well as to what extent they can identify these mistakes.

Therefore, our proposal is to collect corpora of such ontologies. We next describe our ongoing work in this direction, primarily focusing on a corpus of ontologies created by junior ontology engineers.

3. Ongoing Work and Preliminary Resources

Following suit from the proposed idea, we have been collecting ontologies from student assignments from knowledge engineering courses that we have taught during the last decade. Next, we describe the current status of this corpus, as well as a processing workflow that extracts form ontologies their characteristics, competency questions and potential pitfalls. This extraction workflow is made publicly available and can already be used by fellow-researchers.

3.1. A corpus of 500+ student-authored ontologies

Corpus source. Ontologies were collected from student projects in ontology development courses at two institutions: the Vienna University of Technology (TU) and the Vienna University of Economics and Business Administration (WU). For the ontology development, students were instructed to model a domain of their choice (with suggestions including Music, Movies, University, or Tourism), create an ontology with a certain minimum number of concepts and relations, and extend it with specific constraints. They were required to formulate competency questions for their envisioned application. The courses were taught over a decade (2014-2024). TU courses were at the master's level, with TU-ISS designed for beginners and TU-SAIKS for intermediate to advanced students. The WU course was at the bachelor's level. We also collected the student report associated with each ontology as a source for identifying the ontology's domain and competency questions. Each ontology and report was processed with the workflow described in Sect. 3.2 to extract metadata as exemplified in Table 2. Table 1 captures statistics about the corpus based on this metadata.

Corpus size. The corpus comprises 504 ontologies from 16 different classes across the three courses, with a significant majority originating from the TU-ISS course. The ontologies are in the range of small (avg. 26-27 classes/250-300 axioms) size for beginner courses or medium size (avg. 50 classes/2400 axioms) for the advanced course.

Domains covered by the ontologies are shown in Figure 1. TU-ISS and WU-K2 share a big focus on the movie domain, and each has a big occurrence in books, music, and fashion. In contrast, TU-SAIKS, the

Table 1
Ontology Corpus Statistics by Course: Counts of Ontologies (Onts) and Weighted (by the Onts) Averages of Competency Questions (CQs), Classes, and Axioms

Institution	Onts	Total CQs	Avg CQs	Avg Classes	Avg Axioms
TU-ISS (2020-2022)	380	2907	7.65	27.19	291.33
WU-K2 (2022-2024)	76	564	7.42	26.47	259.11
TU-SAIKS (2014-2023)	48	321	6.69	50.22	2438.12
Sum/Weighted Avg	504	3792	7.52 ± 0.28	29.27 ± 6.80	490.93 ± 631.86

Figure 1: Ontology domain coverage across three courses: TU-ISS, TU-SAIKS, and WU-K2

intermediate to advanced course, exhibits a more diverse range of domains, including energy, university management, banking, and various services. Although individual counts for these domains in TU-SAIKS are relatively low, they collectively represent a broad spectrum of interests.

Competency questions. TU-ISS and WU-K2 gathered more competency questions (CQs) due to higher enrollment in their introductory courses. Despite running the longest, TU-SAIKS, an advanced course with lower enrollment, contributed fewer CQs. However, the average number of CQs per ontology remained consistent across all courses.

Table 2
Overview of the meta-data items extracted for ontologies, exemplified with a music related ontology.

<table>
<tr><td>Domain</td><td colspan="6">Music and Song Classification</td></tr>
<tr><td>Summary</td><td colspan="6">An ontology describing various aspects of music, including genres, instruments, artists, and song properties. It focuses on classifying songs, analyzing their components, and representing relationships between different musical elements.</td></tr>
<tr><td colspan="7">Metrics (Count)</td></tr>
<tr><td>Axioms</td><td>Logical Axioms</td><td>Classes</td><td>Object Properties</td><td>Data Properties</td><td>Individuals</td><td>Triples</td></tr>
<tr><td>412</td><td>1220</td><td>62</td><td>34</td><td>11</td><td>4</td><td>412</td></tr>
<tr><td>Ontology Features</td><td colspan="6">CDEHNQRU</td></tr>
<tr><td colspan="7">Competency Questions (8 total, examples):
- Are rock songs using more percussion instruments than pop songs?
- Does a certain reggae song use a piano, a guitar and drums?
- Which Metal songs can be played, if the band has only drums and a guitar?</td></tr>
<tr><td colspan="7">OOPS Pitfalls [16] (5 total):
1. Using different naming conventions in the ontology (Minor, P22, 1 affected element)
2. Equivalent classes not explicitly declared (Important, P30, 1 affected element)
3. Inverse relationships not explicitly declared (Minor, P13, 7 affected elements)</td></tr>
<tr><td colspan="7">ProCK Heuristic Error Check [17] (3 types):
1. missing-closure, e.g., “Instrument SubClassOf isPlayedby some Instrumentalist”
2. missing-disjointness, e.g., between classes such as “Lyrics” and “Musician”
3. trivially-satisfiable-allValuesFrom, e.g., “Genre SubClassOf IsPerformedby min 1 Musician”</td></tr>
</table>

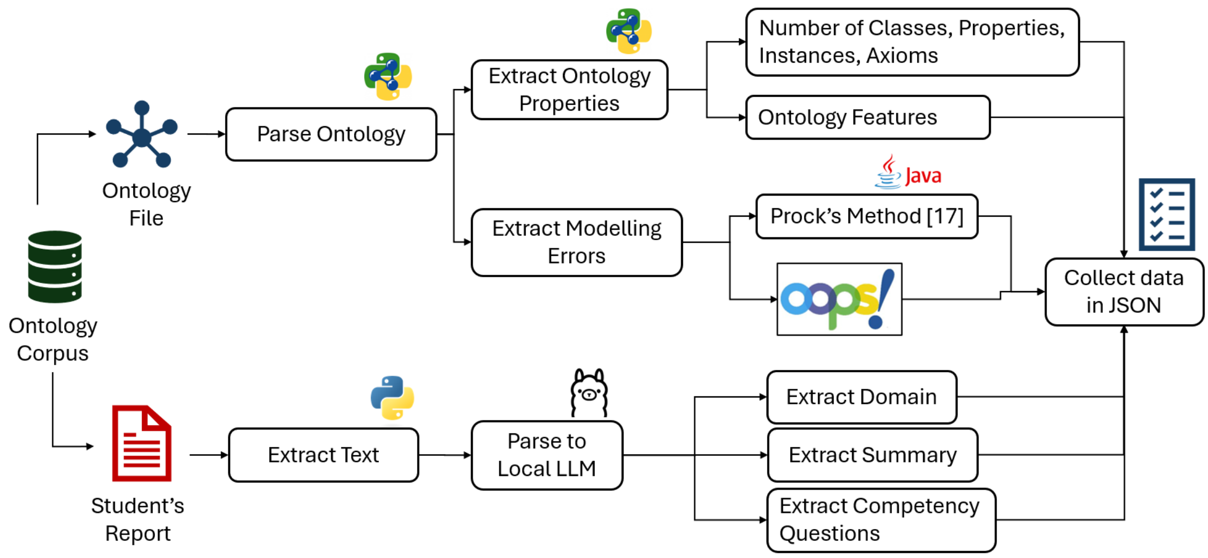

Figure 2: Workflow for extracting metadata from ontologies and their associated report.

3.2. A workflow for extracting ontology metadata

Figure 2 illustrates the workflow for extracting ontology metadata from ontology files and associated students' reports. The ontology analysis utilizes established Python libraries, such as RDFLib, to process ontology files and extract key ontological properties, such as the number of classes, properties, instances, and axioms, while also extracting the ontology's expressivity. Additionally, the analysis incorporates the identification of structures typically prone to errors using Prock's Method [17] and the OntOlogy Pitfall Scanner (OOPS!) [16].

The student report analysis begins by extracting text from the reports using Python scripts. The extracted content is then parsed using a local language model (LLM), specifically Gemma 2[1], deployed using Ollama on university internal infrastructure to ensure data privacy. This LLM-driven process extracts three key components from each report: the domain context, a summary of the report's content, and relevant competency questions that were written by students as part of their reports, with the LLM's role being to identify and extract these pre-existing questions rather than generate new ones.

Finally, the workflow aggregates the data from the ontology analysis and student report evaluation branches into a JSON format, with a format similar to that of Table 2. The source code for this workflow is available online[2] to be utilized by others to process their own ontlogy collections.

4. Outlook

To further expand research on LLM-enabled KE we argue for the need of collecting corpora of ontologies exhibiting variety in their key properties. Future work for our community could include:

- *Collecting more ontology corpora.* We focus on collecting student-authored ontologies. We plan to expand our ontology corpus by processing the remaining ontologies from existing courses and collaborating with other institutions teaching ontology-related classes. Our ontology-metadata extraction workflow could enable fellow-researchers that collect ontologies as a by-product of their work to create their own ontology corpus either for internal experimentation or for sharing those with the community.
- *Extending ontology analysis workflows.* The metadata collected about the ontologies is important for being able to select suitable ontologies for experimentation. In our work, we will enhance

[1]https://ollama.com/library/gemma2
[2]Workflow for metadata extraction from ontologies and reports: https://github.com/wu-semsys/ontology-analysis

the current analysis workflow by integrating additional evaluation tools and developing an anonymization pipeline to facilitate ontology sharing while protecting privacy.

- *Establishing systematic evaluation campaigns.* As ontology corpora become available, they could enable the community to establish benchmarks for ontology engineering tasks and launch evaluation campaigns to systematically study the impact of GenAI methods on knowledge engineering.

Acknowledgments

This work was funded by the Austrian Science Fund (FWF) Bilateral AI (Grant Nr. 10.55776/COE12) and HOnEst (V 745) projects.

References

[1] C.-H. Chiang, H.-y. Lee, Can Large Language Models Be an Alternative to Human Evaluations?, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023. doi:10.18653/v1/2023.acl-long.870.

[2] M. Sallam, K. Al-Salahat, H. Eid, J. Egger, B. Puladi, Human versus Artificial Intelligence: ChatGPT-4 Outperforming Bing, Bard, ChatGPT-3.5, and Humans in Clinical Chemistry Multiple-Choice Questions, medRxiv (2024). doi:10.1101/2024.01.08.24300995.

[3] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen, N. Zhang, LLMs for Knowledge Graph Construction and Reasoning: Recent Capabilities and Future Opportunities, World Wide Web 27 (2024). doi:10.1007/s11280-024-01297-w.

[4] M. Trajanoska, R. Stojanov, D. Trajanov, Enhancing Knowledge Graph Construction Using Large Language Models, 2023. arXiv:2305.04676.

[5] S. Carta, A. Giuliani, L. Piano, A. S. Podda, L. Pompianu, S. G. Tiddia, Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction, arXiv preprint arXiv:2307.01128 (2023).

[6] B. Zhang, I. Reklos, N. Jain, A. M. Peñuela, E. Simperl, Using Large Language Models for Knowledge Engineering (LLMKE): A Case Study on Wikidata, arXiv preprint arXiv:2309.08491 (2023).

[7] N. Fathallah, A. Das, S. De Giorgis, A. Poltronieri, P. Haase, L. Kovriguina, NeOn-GPT: A Large Language Model-Powered Pipeline for Ontology Learning, in: The Semantic Web: ESWC Satellite Events, 2024.

[8] H. Khorashadizadeh, N. Mihindukulasooriya, S. Tiwari, J. Groppe, S. Groppe, Exploring In-Context Learning Capabilities of Foundation Models for Generating Knowledge Graphs from Text, 2023. arXiv:2305.08804.

[9] B. P. Allen, P. T. Groth, Evaluating Class Membership Relations in Knowledge Graphs using Large Language Models, in: The Semantic Web: ESWC Satellite Events, 2024.

[10] S. Tsaneva, S. Vasic, M. Sabou, LLM-driven ontology evaluation: Verifying ontology restrictions with ChatGPT, in: The Semantic Web: ESWC Satellite Events, 2024.

[11] S. Tsaneva, D. Dessì, F. Osborne, M. Sabou, Enhancing Scientific Knowledge Graph Generation Pipelines with LLMs and Human-in-the-Loop, in: 4th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment, co-located with The 23rd International Semantic Web Conference, ISWC, 2024. Accepted for publication.

[12] H. B. Giglou, J. D'Souza, F. Engel, S. Auer, LLMs4OM: Matching Ontologies with Large Language Models, in: The Semantic Web: ESWC Satellite Events, 2024.

[13] B. Zhang, V. A. Carriero, K. Schreiberhuber, S. Tsaneva, L. S. González, J. Kim, J. de Berardinis, OntoChat: a Framework for Conversational Ontology Engineering using Language Models, in: The Semantic Web: ESWC Satellite Events, 2024.

[14] N. Tufek, A. S. S. Thuluva, T. Bandyopadhyay, V. P. Just, M. Sabou, F. J. Ekaputra, A. Hanbury, Validating Semantic Artefacts With Large Language Models, in: The Semantic Web: ESWC Satellite Events, 2024.

[15] Y. Rebboud, L. Tailhardat, P. Lisena, R. Troncy, Can LLMs Generate Competency Questions?, in: The Semantic Web: ESWC Satellite Events, 2024.

[16] M. Poveda-Villalón, A. Gómez-Pérez, M. C. Suárez-Figueroa, Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation, International Journal on Semantic Web and Information Systems (IJSWIS) 10 (2014) 7–34.

[17] A. Prock, Hybrid Human-Machine ontology verification: Identifying common errors in ontologies by integrating human computation with ontology reasoners, Master's thesis, TU Vienna, 2021.