

Automatic Ontology Term Typing by LLMs: the Impact of Prompt and Ontology Variation

Upal Bhattacharya^{1,2}, Maaike de Boer² and Sergey Sosnovsky¹

¹Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, 3584 CC, Utrecht, The Netherlands

²Department Data Science, TNO, Anna van Buerenplein 1, 2595 DA, Den Haag, The Netherlands

Abstract

Large Language Models (LLMs) have been applied to a wide variety of ontology engineering tasks. Building on initial progress, further research is needed to explore potential effects of variation over model-specific and ontology-specific factors. We perform a preliminary study on the ability of an LLM to perform term typing using only its own knowledge through concept retrieval and analyse the effect of domain contextualisation, ontology structure and popularity of ontologies on performance. Our findings suggest that LLMs are reasonably adept at identifying correct individual to concept assertions but are less capable of inferring concept hierarchies when used in a zero-shot setting. Domain contextualisation can enhance performance for structurally complex and less-popular ontologies. Our analysis furthers hints at ontology popularity improving concept retrievability while complexity in terms of structural depth and dispersion makes it difficult for LLMs to identify assertions.

Keywords

LLM, ontology evaluation, ontology learning, individual assertion, term typing

1. Introduction

Large language models (LLMs) have access to substantial factual knowledge [1, 2] and are widely used for several natural language tasks including knowledge-based tasks [3]. Their ability to "understand" complex texts and perform reasoning tasks at scale has led to the wide experimentation with LLMs for ontology enhancement [4, 5, 6, 7, 8]. Interest in leveraging LLMs for various ontology learning tasks [9, 10, 11, 12] has also been growing. However, there is still a lot to learn about the nature of the interaction between LLMs and ontologies. Whereas LLMs are trained on massive data including freely available ontologies from the Web, recent works highlight that such presence of information does not translate into strong performance on various ontology learning and knowledge tasks [13, 14].

While LLM-driven ontology enhancement has been gaining traction, more attention needs to be paid to exploring the possible effects of variability within LLMs and ontologies on the potency of such enhancement. The interplay of factors like prompt domain contextualisation, ontology structure and ontology popularity can significantly affect the ability of LLMs to perform ontology-related tasks. Understanding the effects of variation in these underlying variables and their interaction can help identify optimal strategies for LLM-supported ontology development.

This paper describes a preliminary study on the ability of an LLM to perform term typing in a zero-shot setting and presents an analysis of the observed performance depending on a small subset of potentially important LLM-specific and ontology-specific factors. Term typing is an ontology learning and enrichment task of mapping new individuals to concepts within an ontology. It requires a model to "understand" (or at least to recognize) the features of concepts within an ontology to make new individual to concept assertions. The paper addresses the following research question: **How capable are LLMs at ontology term typing through concept retrieval?** As a part of the study, we also look into the LLMs' ability to implicitly identify concept hierarchies. We investigate performance variation over prompt domain contextualisation, ontology structure, and ontology popularity on an ontology

EKAW 2024: EKAW 2024 Workshops, Tutorials, Posters and Demos, 24th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2024), November 26-28, 2024, Amsterdam, The Netherlands

✉ u.bhattacharya@uu.nl (U. Bhattacharya); maaïke.deboer@tno.nl (M. d. Boer); s.a.sosnovsky@uu.nl (S. Sosnovsky)

🆔 0000-0002-0877-7063 (U. Bhattacharya); 0000-0002-2775-8351 (M. d. Boer); 0000-0001-8023-1770 (S. Sosnovsky)



© 2024 Copyright © 2024 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

learning task. Throughout the study, we outline the importance of identifying critical factors to be able to explain variation exhibited by LLMs over different ontology development and learning tasks.

2. Related Work

Research on the application of LLMs in ontology development range from techniques to enhance existing ontologies, to approaches focused on engineering ontologies from scratch, to evaluation frameworks for assessment of performance of LLMs in various ontology learning tasks.

Several studies focused on using LLMs for automated ontology development to reduce human intervention [15]. Dong et al. [8] approached the task of concept placement with a three-step strategy of edge search, formation and selection using various models. Their results highlight that fine-tuned BERT-based [16] models outperform larger decoder-only models used in a zero-shot setting. Focusing on decoder-only LLMs, Giglou et al. [12] modeled ontology alignment as a paired-sentence classification task by prompting LLMs to answer questions of equivalence of concepts given their immediate parents or children. He et al. [7] evaluated the ability of LLMs to perform concept matching as a binary classification task to find that zero-shot decoder-only models do not perform as well as their encoder-only counterparts. Snijder et al. [17] compared different methods for ontology alignment in the application domain of the labour market. Chen et al. [10] addressed the task of subsumption inference as a paired-sentence classification task by fine-tuning a BERT model on an ontology entity-based paired-sentence dataset.

The ability of LLMs to perform complex tasks and provide formatted outputs prompted their use in the development of ontologies from the ground up using natural language guidelines or ancillary tasks. Kommineni et al. [18] used LLMs to generate ontologies and knowledge graphs by tasking the model to generate competency questions, an ontology and, subsequently, a knowledge graph from the human-verified questions. Using an LLM to parse natural language sentences into OWL syntax, Mateiu and Groza [6] were able to create a tool that is capable of generating and populating simple ontologies from basic sentences. However, the applicability of such a tool for more complex requirements has not been tested. Funk et al. [5] leveraged natural language as well, adopting a repeated prompt strategy to identify child concepts and their optimal placements to create a hierarchy. Domain ontology development requires significant knowledge about the domain itself thereby posing a more nuanced problem for LLMs. Doumanas et al. [19] investigated the use of LLMs to develop domain ontologies for the Search and Rescue domain. Their findings highlight impressive capabilities of LLMs to fit new factual information into an ontology framework and affirm LLMs as capable ontology engineers.

The lack of clarity behind the mechanisms responsible for the observed performance of language models necessitates development of evaluation frameworks for various ontology-related tasks. Investigating LLM hallucinations over simple information about well-known ontologies, Bombieri et al. [13], probed models for concept labels from the Gene Ontology [20] and the Uberon Ontology [21] using only their IDs. Their findings show low rate of hallucinations but poor performance in general highlighting some degree of memorisation proportional to the popularity of the ontologies on the Web. He et al. [11] provided a framework for evaluation of LLMs for subsumption inference. Modelled as a natural language inference task, their study found decoder-only models to be quite adept at identifying simple and complex subsumptions. Looking into term typing, taxonomy discovery and relation extraction performance, Babaei Giglou et al. [22] evaluated models in a zero-shot settings on nine different datasets. Mai et al. [23] gauged reasoning and learning capabilities of language models to find prediction inconsistencies suggesting that such models tend to fall back to their pre-learned lexical senses as opposed to using the provided semantic meanings of concepts in ontologies.

Despite the growing interest and the considerable progress made over the last few years in application of LLMs for ontology development and evaluation methods for these techniques, there is one important aspect that has not been sufficiently addressed in literature. We need to obtain a better understanding of performance variability arising from the underlying model-specific and ontology-specific factors. The interplay of these factors can be quite significant as suggested by variation in reported performance across studies.

3. Variation Analysis

While there are many LLM-specific and ontology-specific factors that can contribute to performance variability on an ontology learning task, we focus on a specific subset for our study on term typing. We briefly outline some of these factors contributing to the form and analysis of our experimentation.

LLM-based Variability: When used without fine-tuning, LLM-based variability (apart from the choice of different LLMs) is primarily driven by **prompt variability**. We consider the following elements of a prompt as variables of interest for probing LLMs for our present study:

- **Nature of the task:** An ontology learning task can be structured in different ways e.g. classification, summarization, retrieval, etc. Prompting an LLM to perform an ontology learning task using different output strategies allows for analysis of the suitability of a particular task to each performance objective.
- **Prompting strategy:** Various prompting strategies like zero-shot, few-shot and retrieval-augmented-generation allow assessment of the amount of example data required by an LLM to perform an ontology learning task. It indicates the relevancy of the pre-learned knowledge of an LLM to perform the task.
- **Domain contextualisation:** Defining the role of an LLM as an ‘assistant’ or ‘expert’ on a particular domain or topic in addition to the type of task to be performed e.g. classification, retrieval, etc. is a powerful tool that encourages LLMs to derive the correct context from the user input data. In the space of an ontology learning task, this domain in itself is multi-faceted and can take on any of the following forms:
 - **Generic:** The LLM is not given any role other than that based on the task to be performed.
 - **General domain of ontologies:** The LLM is defined as an expert in ontologies.
 - **Topic of an ontology:** The LLM is defined as an expert in the topic of an ontology e.g. ‘You are a wine expert’ for an LLM as a wine expert for the Wines Ontology [24]
 - **Combination of a general ontology and a topic:** The LLM is defined as an expert in ontologies and an expert in the topic of a particular ontology.

The degree of contextualisation provided by the domain specification in the prompt helps assess the optimal level of specificity for an LLM to perform an ontology learning task.

Data Variability: The heterogeneity of ontologies modelling different domains introduces variability based on their *structure* and *content*. We investigate variability along the following variables:

- **Ontology Structure:** Ontology structure is driven by the nature of the underlying topic. Metrics for measuring the structural complexity such as depth, breadth, dispersion and tangledness [25] can help categorise ontologies and reason over similarities and differences in performance of LLMs across ontologies and ontology learning tasks based on their structure.
- **Popularity:** LLMs are capable of memorizing their training data [26] and their performance on ontology learning tasks is affected by the popularity of the ontology on the Web [13]. Analysing the effect of the **popularity of the topic of an ontology** provides insight into the ability of LLMs to leverage pre-learned information and their capability to override it as required.

4. Term Typing Ranked Retrieval

We present a preliminary study on term typing by LLMs as a ranked retrieval problem and analyse the effect of variation based on the following three variables: **domain contextualisation**, **ontology structure** and **ontology popularity** on the task.

Following a similar investigation of term typing in Giglou et al. [12], we model the task as a retrieval problem and prompt OpenAI’s GPT-4o [27] model in a zero-shot setting with a modified task of generating a ranked list of concepts of length up to the depth [28] of the ontology. The choice of

zero-shot prompting forces the model to utilise only its own knowledge of the concepts and individuals that it is prompted with (using only their labels) in order to make the correct assertions. This choice highlights the relevancy of an LLM’s world knowledge to a basic ontology learning task. Modelled as a retrieval task with a retrieval length greater than one provides insight into the LLM’s ability to accurately identify and infer concept hierarchies on its own when given only the concept labels.

We compute the standard information retrieval (IR) metrics: R-Precision, Mean Average Precision (mAP) and Normalized Discounted Cumulative Gain (nDCG) with mAP and nDCG computed at depth k of the ontology (i.e. mAP@ k and nDCG@ k). R-Precision and mAP provide insight into the general ability of the model to identify relevant concept assertions and transitive ancestor relations with the latter laying greater emphasis on the order of retrieval based on the hierarchy. nDCG lays greater emphasis on retrieving the relevant concepts and parents in the correct hierarchy and acts as an indicator of the understanding LLMs have of concept hierarchies from just their labels. We define the relevance of an ancestor concept A for an individual I according to Equation 1 where C_I is the directly asserted concept of the individual I and $d(\cdot, \cdot)$ is the edge distance between A and C_I . For all non-ancestor concepts, we set the relevance to 0.

$$\text{Relevance}(A, I) = \frac{1}{1 + d(A, C_I)} \quad (1)$$

We conduct experiments over two ontologies of varying size, complexity and popularity. The Wines Ontology [24] is a well-known ontology that is relatively small and structurally-simple, and enjoys significant popularity by itself and in terms of the domain it represents. The CASE Ontology [29] is a larger, more complex and newer ontology focused on accurately capturing the lifecycle of digital evidence. We inject individuals into the CASE Ontology using the Owl Trafficking example provided on the CASE Ontology website¹ and also include concepts from the closely related UCO Ontology [29]. Hereafter, we refer to this composite constructed ontology as the CASE Ontology itself. Table 1 highlights the relevant structural metrics of the two ontologies to highlight their differences.

Metric	Wines Ontology	CASE Ontology
Classes (#)	76	434
Individuals (#)	131	161
Depth (#) [28]	4	8
Breadth (#) [28]	62	228
Dispersion (#) (max.) [28]	3	118

Table 1
Ontology Metrics for Wines Ontology and CASE Ontology

We design prompts for the both ontologies based on the four types of domain contextualisation. Domain contextualisation is achieved by specifying the role of the LLM as an expert in a domain following one of the four types outlined in Section 3. We also specify the task of generating a ranked list of the most relevant concepts of length equal to the depth of each ontology. For each type of prompt, we provide a flat list of all the concept labels from the ontology from which the LLM is to generate its responses².

4.1. Results and Discussion

Table 2 provides the retrieval-based metrics for both ontologies. We observe better performance on the Wines Ontology across all IR metrics. The simpler structure of the ontology, its general popularity, and the popularity of its domain could be major contributors to the LLM’s ability to retrieve the correct concepts. This line of reasoning is supported by significantly lower performance on the CASE Ontology. A smaller and more shallow ontology could contribute to the reduction in difficulty of retrieving concepts in a correct hierarchy resulting in higher nDCG scores on the Wines Ontology. Consequently, lower nDCG values for the ontology of greater depth reaffirms this notion.

¹<https://caseontology.org/examples/>

²The exact prompts used can be found at GitHub.

Domain Context	Wines Ontology				CASE Ontology			
	R-Precision	mAP@4	nDCG@4	ρ	R-Precision	mAP@8	nDCG@8	ρ
Generic	0.7873	0.8192	0.7736	<i>0.4418</i>	0.2743	0.2611	0.5609	0.5496
Ontology	0.7491	0.7855	0.7401	<i>0.4431</i>	0.3169	0.3155	0.5970	<i>0.3095</i>
Topic	0.7196	0.7603	0.7116	<i>0.5004</i>	0.2417	0.2326	0.5222	<i>0.0375</i>
Ontology and Topic	0.7681	0.7998	0.7552	<i>0.5742</i>	0.3001	0.3113	0.5096	<i>0.4474</i>

Table 2

IR Metrics and Pearson Correlation (ρ) between retrievability and dispersion [28] for Wines Ontology and CASE Ontology. Correlation values in italics are significant at a p-value of 0.05. Values in bold indicate the maximum values for that particular metric.

We observe that a general ontology contextualisation of an LLM results in better performance than contextualising LLM as a topic expert. Optimal prompt engineering of this contextualisation may improve performance but simple prompts highlight that a topic-based domain contextualisation is not improving the results of term typing (at least for GPT 4.0 as the chosen LLM). The generic prompt performs the best on the smaller and more popular Wines Ontology. The simpler structure and popularity of the ontology coupled with the best performance suggests that popularity can dominate other variables of interest thus marginalising the need for carefully considered domain contextualisation in prompt engineering.

We measure the Pearson correlation between the retrievability of a concept and its dispersion [28] to analyse the effect of ontology structure on performance. Dispersion is a measure of ontological structural complexity defined for a concept as the number of child concepts it has. We define the retrievability of a concept as the number of times it is predicted as a relevant concept by the LLM, across all queries. Our observations suggest that domain contextualisation influences structural considerations during retrieval. Column ρ in Table 2 outlines the correlation between dispersion and retrievability. The domain contextualisations involving ontologies show moderate correlation between dispersion and retrievability. Concepts with high dispersion have several child concepts and represent conceptually ‘broader’ formalisations. Such concepts are possibly semantically wider in their scope and thus could be easier for LLMs to retrieve using their own knowledge when contextualised to consider hierarchies. Similar correlation between all contextualisations for the Wines ontology (max. dispersion: 3) indicates that in simpler ontologies, where dispersion is not well-pronounced, performance is not affected. Future studies with more ontologies of varying degrees of dispersion would help corroborate this better. We do not observe statistically significant correlation between the depth of a concept and its retrievability.

5. Conclusion

We present a preliminary study on the ability of LLMs to perform term typing in a zero-shot setting. An LLM is tasked with retrieving a ranked list of the most relevant concepts given an individual by leveraging its own knowledge about the entities from their labels. We summarize the findings of our study as follows:

- LLMs have reasonable capability in performing term typing using their own world knowledge but this does not help them identify concept hierarchies, particularly in less popular domains.
- Popularity of domains seems to play an important role in a zero-shot setting and can override other variables of interest.
- For less popular domains, domain contextualisation can improve performance. Considering structural experts over topic experts may yield better performance.
- Concepts with greater dispersion may be semantically broader and can therefore be easier for an LLM to retrieve.

Future works will focus on investigating the task of term typing over different prompting strategies and conducting an exhaustive analysis of all relevant factors. Extension of the work to other ontology tasks will lead to the creation of a comprehensive and robust analysis system that can be utilised to ensure optimal performance for any ontology development task using LLMs.

References

- [1] F. Petroni, T. Rocktäschel, S. Riedel, P. Lewis, A. Bakhtin, Y. Wu, A. Miller, Language Models as Knowledge Bases?, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 2463–2473. doi:10.18653/v1/D19-1250.
- [2] A. Roberts, C. Raffel, N. Shazeer, How Much Knowledge Can You Pack Into the Parameters of a Language Model?, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 5418–5426. doi:10.18653/v1/2020.emnlp-main.437.
- [3] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J.-Y. Nie, J.-R. Wen, A Survey of Large Language Models, 2023. arXiv:2303.18223.
- [4] H. Liu, Y. Perl, J. Geller, Concept placement using BERT trained by transforming and summarizing biomedical ontology structure, Journal of Biomedical Informatics 112 (2020) 103607. doi:10.1016/j.jbi.2020.103607.
- [5] M. Funk, S. Hosemann, J. C. Jung, C. Lutz, Towards Ontology Construction with Language Models, in: KBC-LM @ ISWC 2023, 2023. doi:10.48550/arXiv.2309.09898.
- [6] P. Mateiu, A. Groza, Ontology engineering with Large Language Models, 2023 25th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (2023) 226–229. doi:10.1109/SYNASC61333.2023.00038.
- [7] Y. He, J. Chen, H. Dong, I. Horrocks, Exploring large language models for ontology alignment, in: I. Fundulaki, K. Kozaki, D. Garijo, J. M. Gómez-Pérez (Eds.), Proceedings of the ISWC 2023 Posters, Demos and Industry Tracks: From Novel Ideas to Industrial Practice Co-Located with 22nd International Semantic Web Conference (ISWC 2023), Athens, Greece, November 6–10, 2023, volume 3632 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023.
- [8] H. Dong, J. Chen, Y. He, Y. Gao, I. Horrocks, A Language Model Based Framework for New Concept Placement in Ontologies, in: A. Meroño Peñuela, A. Dimou, R. Troncy, O. Hartig, M. Acosta, M. Alam, H. Paulheim, P. Lisena (Eds.), The Semantic Web, Springer Nature Switzerland, Cham, 2024, pp. 79–99. doi:10.1007/978-3-031-60626-7_5.
- [9] Y. He, J. Chen, D. Antonyrajah, I. Horrocks, BERTMap: A BERT-Based Ontology Alignment System, Proceedings of the AAAI Conference on Artificial Intelligence 36 (2022) 5684–5691. doi:10.1609/aaai.v36i5.20510.
- [10] J. Chen, Y. He, Y. Geng, E. Jiménez-Ruiz, H. Dong, I. Horrocks, Contextual semantic embeddings for ontology subsumption prediction, World Wide Web 26 (2023) 2569–2591. doi:10.1007/s11280-023-01169-9.
- [11] Y. He, J. Chen, E. Jimenez-Ruiz, H. Dong, I. Horrocks, Language Model Analysis for Ontology Subsumption Inference, in: A. Rogers, J. Boyd-Graber, N. Okazaki (Eds.), Findings of the Association for Computational Linguistics: ACL 2023, Association for Computational Linguistics, Toronto, Canada, 2023, pp. 3439–3453. doi:10.18653/v1/2023.findings-acl.213.
- [12] H. B. Giglou, J. D’Souza, F. Engel, S. Auer, LLMs4OM: Matching Ontologies with Large Language Models, 2024. arXiv:2404.10317.
- [13] M. Bombieri, P. Fiorini, S. P. Ponzetto, M. Rospocher, Do LLMs Dream of Ontologies?, 2024. doi:10.48550/arXiv.2401.14931. arXiv:2401.14931.
- [14] K. Wang, G. Qi, J. Li, S. Zhai, Can Large Language Models Understand DL-Lite Ontologies? An Empirical Study, 2024. doi:10.48550/arXiv.2406.17532. arXiv:2406.17532.
- [15] R. M. Bakker, D. L. D. Scala, From Text to Knowledge Graph: Comparing Relation Extraction Methods in a Practical Context (????).
- [16] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:

- Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. doi:10.18653/v1/N19-1423.
- [17] L. L. Snijder, Q. T. S. Smit, M. H. T. de Boer, Advancing Ontology Alignment in the Labor Market: Combining Large Language Models with Domain Knowledge, *Proceedings of the AAAI Symposium Series 3* (2024) 253–262. doi:10.1609/aaais.v3i1.31208.
 - [18] V. K. Kommineni, B. König-Ries, S. Samuel, From human experts to machines: An LLM supported approach to ontology and knowledge graph construction, 2024. doi:10.48550/arXiv.2403.08345. arXiv:2403.08345.
 - [19] D. Doumanas, A. Soularidis, K. Kotis, G. Vouros, Integrating LLMs in the Engineering of a SAR Ontology, in: I. Maglogiannis, L. Iliadis, J. Macintyre, M. Avlonitis, A. Papaleonidas (Eds.), *Artificial Intelligence Applications and Innovations*, Springer Nature Switzerland, Cham, 2024, pp. 360–374. doi:10.1007/978-3-031-63223-5_27.
 - [20] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology: Tool for the unification of biology, *Nature Genetics* 25 (2000) 25–29. doi:10.1038/75556.
 - [21] C. J. Mungall, C. Torniai, G. V. Gkoutos, S. E. Lewis, M. A. Haendel, Uberon, an integrative multi-species anatomy ontology, *Genome Biology* 13 (2012) R5. doi:10.1186/gb-2012-13-1-r5.
 - [22] H. Babaei Giglou, J. D’Souza, S. Auer, LLMs4OL: Large Language Models for Ontology Learning, in: T. R. Payne, V. Presutti, G. Qi, M. Poveda-Villalón, G. Stoilos, L. Hollink, Z. Kaoudi, G. Cheng, J. Li (Eds.), *The Semantic Web – ISWC 2023*, Springer Nature Switzerland, Cham, 2023, pp. 408–427. doi:10.1007/978-3-031-47240-4_22.
 - [23] H. T. Mai, C. X. Chu, H. Paulheim, Do LLMs Really Adapt to Domains? An Ontology Learning Perspective, 2024. doi:10.48550/arXiv.2407.19998. arXiv:2407.19998.
 - [24] H. P. P. Filho, Ontology Development 101: A Guide to Creating Your First Ontology (????).
 - [25] R. Wilson, J. Goonetillake, W. Indika, A. Ginige, A conceptual model for ontology quality assessment: A systematic review, *Semantic Web* 14 (2023) 1051–1097. doi:10.3233/SW-233393.
 - [26] N. Carlini, D. Ippolito, M. Jagielski, K. Lee, F. Tramèr, C. Zhang, Quantifying Memorization Across Neural Language Models, *ArXiv* (2022).
 - [27] Hello GPT-4o, <https://openai.com/index/hello-gpt-4o/>, ????
 - [28] A. Gangemi, C. Catenacci, M. Ciaramita, J. Lehmann, Ontology evaluation and validation An integrated formal model for the quality diagnostic task, 2005.
 - [29] E. Casey, S. Barnum, R. Griffith, J. Snyder, H. Van Beek, A. Nelson, Advancing coordinated cyber-investigations and tool interoperability using a community developed specification language, *Digital Investigation* 22 (2017) 14–45. doi:10.1016/j.diin.2017.08.002.