Natural Language Processing for Knowledge Graph Extraction and Evaluation

Roos M. Bakker[1,2][0000−0002−1760−2740]

[1] TNO Netherlands Organisation for Applied Scientific Research, Kampweg 55, 3769ZG Soesterberg, The Netherlands.
[2] Leiden University Centre for Linguistics, Leiden University, 2311 BE Leiden, Reuvensplaats 3-4, the Netherlands roos.bakker@tno.nl

Keywords: Knowledge Graph Extraction · Ontology Learning · Natural Language Processing · Large Language Models · Ontology Evaluation

Knowledge graphs, ontologies, and other semantic models are powerful solutions to structure data and express linguistic entities and relations between them. However, creating these models requires significant manual effort from domain experts and developers. Natural Language Processing (NLP) offers an opportunity to automate this task by extracting structured information from texts. With the vast amount of textual data online, semantic models can be created and maintained in various ways. This ongoing research explores various methods for automating the extraction and evaluation of semantic models across domains such as law and safety. We present current findings, and future directions.

In the legal domain, we investigated techniques for extracting knowledge graphs—structured representations of domain-specific knowledge. Knowledge graphs were introduced in the 1980s as a tool for organising knowledge [10] and have gained widespread popularity since the introduction of the Google knowledge graph [16]. They consist of entities and relations between them, called triples. One way to populate knowledge graphs with domain-specific facts is by extracting subgraphs. For the legal domain, we explored the extraction of these subgraphs using a domain-specific format: FLINT [18]. The FLINT ontology defines frames for legal actions, with roles such as the *actor*, *object*, and the *recipient* of an *action*. We fine-tuned a BERT model [9] with a newly created dataset for this task [6, 19]. These fine-tuned models proved very effective for this task, reaching accuracy scores of 0.80 on average [6]. In recent work, we compared a collection of fine-tuned models with a rule-based approach [3], and with different prompting strategies for Large Language Models (LLMs)[3]. Although the fine-tuned models performed best, LLM approaches outperformed the rule-based approach and have the advantage that they do not need annotated data, thereby providing a solution for domains with limited resources.

Instead of creating subgraphs, extracting a full knowledge graph can provide more insight for other domains. In the safety domain, we compared various relation extraction techniques from news messages using statistical methods such

[3] This work is currently under review

as co-occurrences [4, 7], classification models leveraging embeddings [11, 15], and LLM prompting strategies. We evaluated performance against a manually created ground truth. The results showed that statistical and relation classification models struggled to capture the richness of natural text and were outperformed by LLM prompting strategies. However, LLMs still fell short of matching human-created ground truths, with F1 scores around 0.6 [4].

Ontologies, while closely related to knowledge graphs, impose more rigorous constraints on relations and entities [1]. They formalise knowledge through general facts and logical relations, such as taxonomies and properties [17]. Due to their increased complexity and generalisations, ontologies are harder to extract from text than knowledge graphs. Ontology learning focuses on this task and has been a long existing field of research, with early work using rule-based approaches [8], and advancing onto statistical approaches [12]. With the advancement of NLP techniques due to architectures such as transformers, the quality of both knowledge graph extraction and ontology learning approaches has improved [14]. In recent work, we expanded our experiments from the safety domain by adding experiments with LLMs for ontology learning. We annotated the same news dataset [4] and modelled an ontology from the annotations. We included various prompting strategies for GPT-4o and evaluated them against the ontology. F1 scores ranged from 0.2 for properties, to 0.65 for individuals, with average scores lying around 0.5, revealing that some ontology elements are easier to automatically extract than others [5].

While most of our experiments have been evaluated against a ground truth, in practice, a ground truth is often unavailable and impractical to create due to the size of the graphs. Consequently, any modifications made must be evaluated manually. To facilitate automatic evaluation, we proposed a set of evaluation metrics which measure the semantic and syntactic quality of a graph [2]. We validated the metrics through experiments that involved removing concepts and adding faulty ones. The results demonstrated that these metrics can effectively indicate positive or negative changes in the graph, alerting developers when a review of their modifications may be necessary [2].

A similar evaluation challenge arises when examining techniques used for knowledge graph extraction—LLMs. While the output in the above experiments was evaluated in different ways, open questions remain such as the consistency of results and robustness of models [4, 13]. We are currently investigating these evaluation challenges and potential solutions.

In summary, this research highlights the potential of NLP for the automatic creation of knowledge graphs and ontologies across various domains. Our findings show the effectiveness of fine-tuned models for creating subgraphs, and the promising capabilities of LLMs for knowledge graph extraction and ontology learning. Evaluation of both knowledge graphs and LLMs remains a challenge, although we provide a step towards automatic evaluation [2]. In future efforts, we will aim to enhance extraction techniques and create more robust evaluation frameworks, further advancing the automation of knowledge graph construction.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Al-Aswadi, F.N., Chan, H.Y., Gan, K.H.: From Ontology to Knowledge Graph Trend: Ontology as Foundation Layer for Knowledge Graph. In: Iberoamerican Knowledge Graphs and Semantic Web Conference. pp. 330–340. Springer (2022)
2. Bakker, R.M., de Boer, M.H.T.: Dynamic Knowledge Graph Evaluation. TechRxiv preprint (2024), under review
3. Bakker, R.M., de Boer, M.H.T., van Drie, R.A.N., Vos, D.: Extracting Structured Knowledge from Dutch Legal Texts: A Rule-Based Approach. In: EKAW - KM4LAW 2022: International Conference on Knowledge Engineering and Knowledge Management, The Knowledge Management for Law workshop (2022)
4. Bakker, R.M., Di Scala, D.L.: From Text to Knowledge Graph: Comparing Relation Extraction Methods in a Practical Context. In: First International Workshop on Generative Neuro-Symbolic AI, co-located with ESWC 2024. Hersonissos, Crete, Greece (May 26 2024)
5. Bakker, R.M., Di Scala, D.L., de Boer, M.H.T.: Ontology Learning from Text: an Analysis on LLM Performance. In: Proceedings of the 3rd NLP4KGC International Workshop on Natural Language Processing for Knowledge Graph Creation, co-located with Semantics 2024. Amsterdam, Netherlands (September 17–19 2024)
6. Bakker, R.M., van Drie, R.A.N., de Boer, M.H.T., van Doesburg, R., van Engers, T.M.: Semantic Role Labelling for Dutch Law Texts. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference. pp. 448–457. European Language Resources Association (2022), https://aclanthology.org/2022.lrec-1.47
7. Bakker, R.M., Kalkman, G.J., Tolios, I., Blok, D., Veldhuis, G.A., Raaijmakers, S., de Boer, M.H.T.: Exploring Knowledge Extraction Techniques for System Dynamics Modelling: Comparative Analysis and Considerations. In: Proceedings of the Benelux Conference on Artificial Intelligence (BNAIC) (November 2023)
8. Buitelaar, P., Cimiano, P., Magnini, B.: Ontology learning from text: methods, evaluation and applications, vol. 123. IOS press (2005)
9. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
10. Feigenbaum, E.A.: The art of artificial intelligence: themes and case studies of knowledge engineering. In: Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 2. p. 1014–1029. IJCAI'77, Morgan Kaufmann Publishers Inc. (1977)
11. Huguet Cabot, P.L., Navigli, R.: REBEL: Relation extraction by end-to-end language generation. In: Findings of the Association for Computational Linguistics: EMNLP 2021. pp. 2370–2381. Association for Computational Linguistics (2021)
12. Khadir, A.C., Aliane, H., Guessoum, A.: Ontology learning: Grand tour and challenges. Computer Science Review **39**, 100339 (2021)
13. Loy, M., Sinha, D.A., Futrell, R.: Exploring the Sensitivity of LLMs' Decision-Making Capabilities: Insights from Prompt Variation and Hyperparameters. Findings of hte Association for Computational Linguistics (EMNLP) pp. 3711–3716 (2023)

14. Pan, J.Z., Razniewski, S., Kalo, J.C., Singhania, S., Chen, J., Dietze, S., Jabeen, H., Omeliyanenko, J., Zhang, W., Lissandrini, M., Biswas, R., de Melo, G., Bonifati, A., Vakaj, E., Dragoni, M., Graux, D.: Large language models and knowledge graphs: Opportunities and challenges (2023)
15. Rossiello, G., Chowdhury, M.F.M., Mihindukulasooriya, N., Cornec, O., Gliozzo, A.M.: KnowGL: Knowledge generation and linking from text. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 16476–16478 (2023)
16. Singhal, A.: Introducing the Knowledge Graph: things, not strings (2012), https://blog.google/products/search/introducing-knowledge-graph-things-not/
17. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: Principles and methods. Data & knowledge engineering **25**(1-2), 161–197 (1998)
18. van Doesburg, R.: A formal method for interpretation of sources of norms. Tech. rep., Leibniz Center for Law, University of Amsterdam (2017)
19. van Drie, Romy A. N. and de Boer, Maaike H. T. and Bakker, Roos M. and Tolios, Ioannis and Vos, Daan: The Dutch Law as a Semantic Role Labeling Dataset. In: Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law. p. 316–322. ICAIL '23, Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3594536.3595124