Not only the What, but also the When:
Extracting Temporally Informed Relations From Wikipedia Biographies

Teresa Liberatore [1], Fina Polat [1]

[1] INDELab, University of Amsteredam, Amsterdam, The Netherlands
{t.liberatore,f.yilmazpolat}@uva.nl

Wikipedia hosts the biographies of numerous individuals who played a pivotal role in shaping today's world, presented as natural language narratives. Most of these biographies have corresponding entries in Wikidata, the structured knowledge graph of Wikipedia, but the information in Wikidata is frequently incomplete compared to the richer content of Wikipedia's texts. A major shortcoming in Wikidata is the absence of comprehensive temporal data on life events. Temporal data is essential for constructing accurate timelines and allows for more advanced queries, such as identifying other entities involved in specific events during the same time frame. Addressing this gap can provide deeper insights into the historical and social context of people' lives both as individuals and as groups.

In this work, we focus on the biographies of fashion designers. By narrowing the scope to this group, we develop a specialized ontology to systematically structure the key information present in their Wikipedia biographies. We then propose a method for completing Wikidata's knowledge graph for fashion designers by leveraging natural language biographies from Wikipedia and a fashion designer biography ontology, and employing Large Language Models for knowledge extraction from text.

Pre-trained language models have shown state-of-the-art performance on several knowledge extraction tasks [3]. The recent advances in large language models (LLM) with emergent capabilities such as in-context learning and instruction following have been shown to improve the performance of many NLP tasks. Building on the work of [1, 2], we propose to use LLMs to extract relevant facts along with their associated temporal information, guided by the underlying ontology of fashion designer biographies. This approach aims to provide richer insights into the lives and historical contexts of fashion designers and serve as a model for improving other domains within Wikidata.

References

[1] Mihindukulasooriya, N., Tiwari, S., Enguix, C. F., and Lata, K. (2023). Text2kgbench: A benchmark for ontology-driven knowledge graph generation from text. In *International Semantic Web Conference*, pp. 247–265. Springer.

[2] Polat, F., Tiddi, I., and Groth, P. (2024). Testing prompt engineering methods for knowledge extraction from text.

[3] Whitehouse, C., Vania, C., Aji, A. F., Christodoulopoulos, C., and Pierleoni, A. (2023). WebIE: Faithful and robust information extraction on the web. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pp. 7734–7755. Association for Computational Linguistics.