Ontoterminological lexicography: Testing Chat GPT in terminology generation, defining, translation, and ontology creation in German, English and Polish

Barbara Lewandowska-Tomaszczyk (University of Applied Sciences in Konin) and Grzegorz Pawłowski (University of Applied Sciences in Konin)

Abstract

The present study focuses on research investigating the quality of automatic terminology generation and ontology creation in three languages to serve as tools towards building multilingual ontoterminological dictionaries/thesauri of restricted domains, particularly as manually-curated background knowledge for specialist fields, is a rather scarce resource. The major task was defined as a comparison of a manual ontoterminological analysis of electrotechnological data in a manufacturing instruction directory and those generated by Chats GPT's comparative responses from available sources of the same domain. The first steps, based on previous research, involved a *manual analysis* of a German manufacturing instruction (*Fertigungsvorschrift)* terminology for the assembly and functioning of thermal switches. Next steps present testing of ChatsGPT functions for the *terminological translation capacity* and *ontology building* for relevant dictionary/thesaurus preparation. The tests were performed on two tools independently – an earlier version of LLMs – AI ChatGPTPro, and on recent ChatGPT4 version. They included two larger batches: prompts referring to the manufacturing instructions, and prompts asking to scrutinize all data available to the systems. The following steps had 6 to 8 prompts (i) generating sets of relevant terminology in German, Polish, and English, (ii) providing definitions of the terms (words and phrases) in German, (iii), translating them into Polish and English, followed by tasks (iv) to generate, and (v) to visualise the ontoterminological data in German, Polish, and English, and (vi) provide the ontology description and interpretation. In the first batch, with prompts based on the instructions, ChatsGPT were asked (vii) to additionally provide – in the three languages – *informal* professional equivalents of all the terms generated in the previous testing steps. The ontology and the visualisation of the generated ontological architecture reflect the relationships and hierarchy of terms and concepts in electrical engineering for the devices. In the explanation of the Components, Processes, Tools and Testing Guidelines, detailed information on the functioning of the components and relationships was provided by ChatGPT4.

In particular, the evaluation of ChatGPT4 lexicographic and terminological functionalities is generally positive: as presented in the responses, ChatGPT4 satisfactorily passed the requested tests, surpassing in some respect manual analyses. The detailed results will be demonstrated in the presentation. The definitions, equivalents and ontologies, fully automatically generated, will be discussed for possible uses in ontoterminological lexicography. On the other hand, all generated data and explanations need evaluation by domain specialists, while the ontology *visualisation* and results concerning the professional *informal* data require both more reliable visualisation techniques and further combined linguistic investigation and subject-matter specialist verification.

Background References

Funk, M., Hosemann, S., Jung, J., C., and C. Lutz. 2023. Towards Ontology Construction with Language Models https://arxiv.org/pdf/2309.09898.pdf

Hallig, R, and Walther von, W. 1963, *Begriffssystem als Grundlage fur die Lexikographie. Versuch eines Ordnungsschemas.* Berlin: Akademie-Verlag (2nd ed.).

Kless, D., Jansen,. L., Lindenthal, J., and J.,Wiebensohn. 2012. A method for re-engineering

a thesaurus into an ontology. In: M. Donnelly and G. Guizzardi (eds.), *Formal Ontology in Information Systems*. Proceedings of the Seventh International Conference (FOIS 2012), IOS Press, 2012, 133-146.

Konysa, A. and Z. Drążek. 2020. Ontology Learning Approaches to Provide Domain-Specific Knowledge Base. *Procedia Computer Science* 176 (2020) 3324–3334.

Author (2023a)

Author (2023b)

Petroni, F., Rocktäschel, T., Riedel, T., Lewis, P.S.H., Bakhtin, A., Wu, Y., and A. H. Miller. 2019. Language models as knowledge bases? *Proceedings of EMNLP-IJCNLP*, Association for Computational Linguistics. 2463–2473.

Roche, C. 2005. Terminologie et ontologie *Revue Langages* 39.157:48-62

Rees, G. P. and Lew, R. 2023. The Effectiveness of OpenAI GPT-Generated Definitions Versus Definitions from an English Learners' Dictionary in a Lexically Orientated Reading Task. *International Journal of Lexicography*. 1-25. https://doi.org/10.1093/ijl/ecad030

Roche, C. 2007. Le terme et le concept : fondements d'une ontoterminologie TOTh 2007, *Terminologie & Ontologie : Théories et applications*, Annecy. 1-22

Roche, C. 2012a. Ontoterminology: How to unify terminology and ontology into a single paradigm *LREC 2012, Eighth international conference on Language Resources and Evaluation*, Istanbul (Turkey). 2626-2630

Roche, C. 2012b Ontologie: entre terminologie et connaissance de spécialité. GLAT – Terminologies: textes, discours et accès aux savoirs spécialisés, Aula Meridiana Facoltà di Giurisprudenza del'Università di Genova, May 2012, Genova, Italie. hal-01180282

Roget, P. M. 1852. *Thesaurus of English words and phrases. classified and arranged so as to facilitate the expression of ideas and assist in literary composition.* London etc.: Longmans, Green & Co.

--- 1984. *Roget's II. The new thesaurus.* New York: Berkley (Condensed version of the homonymous publication by Houghton Mifflin, 1980).

Schryver, de G.-M. 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography*, 2023, **36**, 355–387 https://doi.org/10.1093/ijl/ecad021

Sierra, G., 2000. The onomasiological dictionary: a gap in lexicography. In: Heid, U., Evert, S., Lehmann, E., and C. Rohrer (Eds.), *Proceedings of EURALEX 2000* . Stuttgart University. 223-235. https://euralex.org/wp-content/themes/euralex/proceedings/

Veseli, B., Singhania, S., Razniewski, S. and G. Weikum. 2023. Evaluating language models for knowledge base completion, In: Proceedings of ESWC, vol. 13870 of LNCS, Springer. 227–243. https://doi.org/10.1007/978-3-031-33455-9_14. doi:10.1007/978-3-031-33455-9_14.

Vrolijk, J., Reklos, I., Vafaie, M., Massari, A., M. Mohammadi, M. and S. Rudolph. 2022.Toward a comparison framework for interactive ontology enrichment methodologies, in: Proc. Of VOILA@ISWC, vol. 3253 of *CEUR Workshop Proceedings*, CEUR-WS. 41–50.