Grounding problem in Formal Concept Analysis by means of Large Language Models

Aranda-Corral, Gonzalo A.[1], Bundy, Alan[2], Borrego-Díaz, Joaquín[3], and Chan, Pak Yin[2]

[1] University of Huelva, Spain
gonzalo.aranda@dti.uhu.es
[2] University of Edinburgh, United Kingdom
alan.bundy@ed.ac.uk,histo.chanpy@outlook.com
[3] University of Sevilla, Spain
jborrego@us.es

Abstract. Formal Concept Analysis[2] (FCA) is a mathematical theory of data analysis and knowledge representation that aims to capture concept definitions through formal contexts (object-attribute pairs) and concept lattices. These lattices provide a hierarchical structure of concepts within the data, allowing for reasoning with super-sub concept relationships. However, during the transformation from formal contexts to concept lattices, some concepts emerge without distinguishing attributes, which we term 'anonymous concepts'. Naming these anonymous concepts presents a significant challenge, akin to the Grounding Problem in Cognitive Science.

The Grounding Problem[3] in CogSci refers to the challenge of connecting abstract symbols and representations to their real-world meanings and referents. This fundamental issue addresses how machines can develop genuine semantic understanding, rather than merely manipulating symbols based on a knowledge base. It encompasses various aspects, including language comprehension and concept formation. Despite advances in CogSci, the Grounding Problem remains a significant obstacle in developing systems where interactions between humans and machines are key.

Large Language Models[1] (LLMs) have emerged as powerful tools in natural language processing and artificial intelligence. These deep learning models, trained on vast corpora of text data, demonstrate remarkable capabilities in understanding and generating human-like text across diverse domains. These characteristics make LLMs suitable to face the Grounding Problem and, specifically, to enhance formal knowledge representation methods like Formal Concept Analysis.

In this paper, we present a novel approach to addressing the Grounding Problem within the context of FCA using LLMs. Our methodology consists of two main components: using the Concept Lattice to detect what we call anonymous concepts that require naming, and, developing a systematic prompting strategy for LLMs to generate meaningful names for these anonymous concepts.

Anonymous concepts are concepts that have no own-attributes and therefore no distinguishing label. By the properties of the lattice, these con-

cepts are the intersection of at least 2 superconcepts of it. Using the names of these superconcepts and the extensions of them, we construct a definition that can characterise our anonymous concept and the LLM can give us the word that represents it. This is a key idea for the construction of our question for the LLM to generate names that are not only meaningful but also align with the underlying attributes and context of each concept.

Keywords: Formal Concept Analysis 1 · Grounding Problem 2 · Large Language Models 3

References

1. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners (2020), https://arxiv.org/abs/2005.14165
2. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Berlin/Heidelberg (1999)
3. Harnad, S.: The symbol grounding problem. Physica D: Nonlinear Phenomena **42**(1), 335–346 (1990). https://doi.org/https://doi.org/10.1016/0167-2789(90)90087-6