This abstract is based on two papers

- Neuhaus, Fabian and Hastings, Janna. 'Ontology Development Is Consensus Creation, Not (merely) Representation'. 1 Jan. 2022 : 495 – 513. DOI: 10.3233/AO-220273
- Neuhaus, Fabian. 'Ontologies in the Era of Large Language Models – a Perspective'. 1 Jan. 2023 : 399 – 407. DOI: 10.3233/AO-230072

This paper is not an extension of an earlier conference of workshop paper.

The Illusionary Goal of Automating Ontology Development – with or without Large Language Models

Fabian Neuhaus[1], Janna Hastings[2,3]

[1]*Institute of Intelligent Cooperating Systems, Otto von Guericke University Magdeburg, Germany*
[2]*Faculty of Medicine, Institute for Implementation Science in Health Care, University of Zurich, Switzerland*
[3]*School of Medicine, University of St. Gallen, Switzerland*

Abstract

This abstract is based on two papers, namely 'Ontology Development Is Consensus Creation, Not (merely) Representation' [1] and 'Ontologies in the Era of Large Language Models – a Perspective' [2]. We argue that the goal of fully automating ontology development is based on a misunderstanding of what ontology engineers actually do, namely creating a consensual view on the domain. And that large language models, in spite of their many possible benefits, cannot replace an open, participatory, and transparent ontology development process.

A naïve view of ontology development is the following: to develop an ontology, all that one needs to do is to gather the relevant knowledge about a given domain (through studying textual resources or by talking to an expert in that domain) and then formalise it in some appropriate logical language, such as the Web Ontology Language (OWL) or variants of first-order logic (FOL). From this conception, it is often thought to follow that the whole process of ontology development should be amenable to automation. After all, the formalisation of the content of natural language text in a formal language is akin to a translation task from one natural language to another. Because widely available automated natural language translation tools translate between natural languages, it seems as though the automation of ontology development should be possible.

Since large language models (LLM) became available, – according to the same naïve view – even the task of gathering the relevant knowledge is superfluous. Large language models learn the relevant knowledge when they are trained on text corpora. Thus, they can be used to automate the knowledge acquisition phase. Hence, ontology development does not require the painfully slow process of talking to domain experts but instead can be replaced by some clever prompt engineering.

This naïve view of ontology development is based on a misunderstanding of what ontologies are, what they contain, and how they are built. In particular, it misconstrues the task of the ontologist as a mere translator, who is tasked with specifying a shared conceptualisation by formalising it in a logical language. It also misconstrues the capabilities of LLMs.

In [3] ontologies are defined as *formal, explicit specifications of some shared conceptualisation*, where '[...] an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group'. There are several reasons why an ontology ought to represent a *consensus* by experts in the domain: (a) if the ontology represents only one point of view on a controversial subject, it reduces the likelihood of adoption by the relevant community, (b) a divergent interpretation of the vocabulary of the ontology by the users will lead to interoperability issues, and (c) if conflicting views are represented unresolved in the same ontology, the ontology will likely be incoherent and, thus, hard to understand for users.

The naïve view of ontology development is based on the misconception that such a consensus exists before the ontology engineering process starts. This may be the case in rare situations (e.g. if one is tasked to develop an ontology that represents the content of a standard document), but in the vast majority of ontology development projects, there is no pre-existing consensus. Because in any non-trivial domain, there are divergent views on the domain, different opinions, and competing conceptualisations. This situation will be reflected in the literature (or other texts) that are written by experts on the domain, although, often the differences of opinion will only be implicit. Thus, a major task during ontology development is to create a consensus, which is afterward represented in the ontology. 'Consensus' in this sense does not imply that all issues are settled and that arguments have ceased. Rather, it denotes a working agreement about the types of entities to which the discourses in a given domain are referring and their most important relationships. This consensus is reflected by the meaning postulates in the ontology that determines both the logical and the intended semantics of its vocabulary.

Consensus creation is a social process, which cannot be replaced by LLMs. Because if the available literature does not contain a consensus on a given domain, LLMs are not able to learn one. The reason is that LLMs are not designed to resolve divergent points of view and ambiguities, rather they learn to navigate them during their training by adjusting a mapping from texts to probability distributions of tokens appropriately. For this reason, even minor variants of the same prompt or, depending on the configuration, even the same prompt may lead to the generation of contradicting statements. Thus, simple ontological questions like "Is an X a kind of Y?" may be answered by the same LLM both positively and negatively, if both possible views are supported by the literature. Given how LLMs work, we should not expect the output of an LLM to reflect a logically consistent and ontologically sound view of any given domain. Let alone some views that reflect a consensus that does not exist.

Consensus creation for ontology development is a complex social process, which involves two distinct steps: reaching agreement between domain experts that are actively engaged in ontology development (micro-level consensus), and the wider community that the ontology is intended to serve (macro-level consensus). Consensus creation requires an open, participatory, and transparent ontology development process that engages the stakeholders of the larger community. Since, by definition, a participatory process cannot be fully automated, it is not possible to fully automate ontology engineering – with or without LLMs.

References

[1] F. Neuhaus, J. Hastings, Ontology development is consensus creation, not (merely) representation, Applied ontology 17 (2022) 495–513. URL: https://arxiv.org/abs/2210.12026. doi:10.3233/AO-220273.

[2] F. Neuhaus, Ontologies in the era of large language models–a perspective, Applied Ontology 18 (2023) 399–407. doi:10.3233/AO-230072.

[3] R. Studer, R. Benjamins, D. Fensel, Knowledge engineering: Principles and methods, Data & knowledge engineering 25 (1998) 161–197.