Editorial for the Special Issue on Knowledge Engineering

Groth, Paul; Blomqvist, Eva; Sequeda, Juan F.

DOI
10.1016/j.websem.2024.100840
Publication date
2024
Document Version
Final published version
Published in
Journal of Web Semantics

Web Semantics: Science, Services and Agents on the World Wide Web 83 (2024) 100840

Contents lists available at ScienceDirect

Journal of Web Semantics

journal homepage: www.elsevier.com/locate/websem

Editorial

Editorial for the Special Issue on Knowledge Engineering

In 2021, when we first proposed to do a special issue on knowledge engineering, we had seen how new methods and techniques were changing knowledge engineering practices. For example, knowledge graphs are central to a variety of intelligent applications including semantic search, recommendation, conversational agents and data analytics. Their construction and maintenance is facilitated by complex combinations of machine (e.g. information extraction, schema alignment) and human (crowdsourcing, curation) components. However, there was a need to collect and highlight work that not just changed tools (e.g. knowledge graph construction) but focused on these new practices. Our hunch that knowledge engineering was in a place to be radically updated turned out to be more than true, especially with advances in generative AI. These new technologies are causing an even deeper rethink of the theory and practice of knowledge engineering.

We hope that this special issue provides important input into the rethink that is going on within the broader knowledge engineering community. In particular, large knowledge systems are produced by a combination of humans and machines and hence give rise to new questions about knowledge modelling and knowledge acquisition, including the balance between human and machine intelligence and engineering effort, and the ability to maintain knowledge and data at scale.

The five articles in this special issue address the broad range of methods for knowledge engineering, from analyses of human practices and discussions in knowledge engineering practices, via tool support for reverse engineering, ontology alignment, and maintenance of knowledge graphs, to automation through exploiting and improving knowledge graph embeddings.

As knowledge graphs have become essential assets within applications the need to maintain them efficiently becomes critical. In their work, Belhajjame and Mejri [1] present new provenance collection approaches for knowledge graphs that are focused on assisting developers in the analysis and debugging of knowledge graphs. The introduced provenance collection approach is specifically tailored for incremental maintenance to support these interactive maintenance use cases.

The fact that knowledge assets are now regularly maintained over long periods of time (e.g. Wikidata [2] has been available since 2012) means that methods can be developed for knowledge engineering over such long-term resources. A novel example of this is in the paper by Ciroku et al. [3], which turns the classic notion of competency question production on its head. Instead of coming up with competency questions up front, their approach derives new competency questions from data. Specifically, it uses large language models to generate questions based on the verbalisation of triples in the knowledge graph. The result is a new dataset for Wikidata that contributes a set of competency questions for 20 domains. In this new approach, users can be informed of which questions a knowledge graph can support in its current state.

Ontology alignment is another knowledge engineering task that is of increased importance, as the number of available knowledge assets, e.g. ontologies, increase continuously. This is also visible in the popularity of the Ontology Alignment Evaluation Initiative (OAEI), setting baselines and providing gold standards for alignment evaluation. The paper by Hao et al. [4] introduces a new method based on ontology embeddings, enhancing classical alignment techniques by incorporating both the semantic and structural features into the alignment process. The distance, in an embedding space, between two ontological concepts to be aligned is used as the main criterion for alignment. The paper uses OAEI datasets for comparison and evaluation, showing the importance of both semantic and syntactic structures in this process.

The prior paper showed the use of embeddings for ontology alignment. Embedding representations of knowledge graphs have become a critical input for a number of knowledge engineering tasks, for example, by proposing new links to be added to a knowledge graph or allowing for querying over knowledge that is implied by not reified in the knowledge base. Yang et al. [5] introduce a new preprocessing approach based on affine transforms that helps improve downstream embedding models especially for knowledge graphs that contain temporal information.

Lastly, Koutsiana et al. [6] take advantage of the documented discussions from Wikidata to study how the community organizes and produces this critical open source knowledge graph. Wikidata is a prime example of the kind of human-machine curated knowledge graphs that have emerged in the last decade. Interestingly, they find that these discussions focus on knowledge engineering tasks such as merging, adding or deprecating concepts and the construction and maintenance of taxonomies. This is an excellent example of the ability to use empirical observations to help inform knowledge engineering techniques and methodologies.

These papers give rise to important questions going forward including:

- How can we expand and make more efficient the analysis of human-machine knowledge engineering practice?
- How can large language models be effectively employed in knowledge engineering [7] and how does this change methodologies as seen by the example of Ciroku?
- How can long-term evolution and maintenance of knowledge assets be supported and made sustainable and scalable ?
- What is the role of structured knowledge vs more unstructured knowledge in supporting knowledge intensive tasks?

https://doi.org/10.1016/j.websem.2024.100840

Available online 1 November 2024

We hope that you find these articles inspirational and that they will provide support as the community rethinks and reflects on knowledge engineering in the future.

References

[1] Khalid Belhajjame, Mohamed-Yassine Mejri, Online Maintenance of Evolving Knowledge Graphs with RDFS-Based Saturation and Why-Provenance Support, Journal of Web Semantics 78 (Oct. 2023) 100796, https://doi.org/10.1016/j.websem.2023.100796. *DOI.org (Crossref)*.

[2] Denny Vrandečić, Markus Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57 (10) (2014) 78–85.

[3] Fiorela Ciroku, et al., RevOnt: Reverse Engineering of Competency Questions from Knowledge Graphs via Language Models, Journal of Web Semantics 82 (Oct. 2024) 100822, https://doi.org/10.1016/j.websem.2024.100822. *DOI.org (Crossref)*.

[4] Zhigang Hao, et al., Ontology Alignment with Semantic and Structural Embeddings, Journal of Web Semantics 78 (Oct. 2023) 100798, https://doi.org/10.1016/j.websem.2023.100798. *DOI.org (Crossref)*.

[5] Jinfa Yang, et al., Improving Static and Temporal Knowledge Graph Embedding Using Affine Transformations of Entities, Journal of Web Semantics 82 (Oct. 2024) 100824, https://doi.org/10.1016/j.websem.2024.100824. *DOI.org (Crossref)*.

[6] Elisavet Koutsiana, et al., An Analysis of Discussions in Collaborative Knowledge Engineering through the Lens of Wikidata, Journal of Web Semantics 78 (Oct. 2023) 100799, https://doi.org/10.1016/j.websem.2023.100799. *DOI.org (Crossref)*.

[7] Bradley P. Allen, et al., Knowledge Engineering Using Large Language Models, Transactions on Graph Data and Knowledge (TGDK) 1 (1) (2023) 3:1-3:19, https://doi.org/10.4230/TGDK.1.1.3. *DOI.org (Datacite)*.

Paul Groth[a], Eva Blomqvist[b], Juan F. Sequeda[c]

[a] *University of Amsterdam*

[b] *Linköping University*

[c] *data.world*

E-mail addresses: p.t.groth@uva.nl (P. Groth), eva.blomqvist@liu.se (E. Blomqvist), juan@data.world (J.F. Sequeda).