A Domain-knowledge Ontology for Knowledge Graph-poWered Research Assistant for Polymer Science

Anmol Saini[1,*], Jeffrey Ethier[2], and Cogan Shimizu[1]

[1]Wright State University, Department of Computer Science & Engineering, Dayton, Ohio, USA
[2]US Air Force Research Laboratory, Materials & Manufacturing Directorate, Dayton, Ohio, USA
[*]Corresponding author: saini.25@wright.edu.

Introduction

Although automation via computer-aided research techniques is growing more prominent in materials science, its incorporation in (specifically) polymer science remains limited [1]. The field currently relies largely on manually controlled experimentation, which can lead to bias and introduce human error [2]. This, in turn, induces difficulties in reproducibility and replicability (e.g., recently as reported by Garisto [3]). Additionally, the slower nature of manual experimentation protracts the discovery of new polymers. Furthermore, it results in polymer scientists' workload being repetitive, with experiments often needing to be executed several times sequentially with minute modifications until the desired result is achieved [4].

To address these shortcomings, we have begun work on a Knowledge Graph-poWered Research Assistant for Polymer Science (KGWRAPS), an artificial intelligence (AI) agent outfitted with large language models (LLMs) and knowledge graphs (KGs) to support domain experts in the design and execution of polymer experiments. By combining conversational capabilities with complex domain knowledge, this research assistant will expedite work in polymer science as a part of the movement towards self-driving laboratories [5] and ultimately catalyse the digital transformation of the polymer science research landscape. In this work, we focus on the domain KG providing KGWRAPS with its polymer knowledge.

Objective

The domain KG is a foundational aspect of KGWRAPS; and thus, the design of the underlying schema or ontology is critical. We account for such polymer science concepts as materials, properties, processes, conditions, and so on and capture the semantics of the domain so as to provide an AI agent with the knowledge necessary to assist scientists in designing and executing experiments. Furthermore, we intend for this KG to act as a unified repository of knowledge that incorporates diverse and expansive polymer data from existing sources and new experiments to enable more convenient access and retrieval of information. In future work, we will expand the current representation, which focuses on more concrete characteristics, to incorporate factors like equipment capability and experimental protocols (e.g., incorporation into an electronic lab notebook).

Methods

Our approach to developing this schema primarily relied on identifying existing resources for the structural modelling of polymer concepts and then selecting and arranging the concepts most suitable to our use case. The two resources we initially prioritised are the Community Resource for Innovation in Polymer Technology (CRIPT)[1] and NanoMine[2] due to their comprehensive coverage of materials concepts and particular focus on polymers. In laying the foundations, we used the main CRIPT Nodes [6] to define the primary modules of our schema. Specifically, we adapted the Modular Ontology Modeling (MOMo) methodology [7] to develop the schema modularly, so that specific components could be considered independently of each other. Then, we incorporated more granular detail through the use of NanoMine concepts [8]. In particular, we use the *struc-*

[1]`https://www.criptapp.org`
[2]`https://materialsmine.org/nm`

ture of NanoMine as a template rather than the graph itself, which allows for better flexibility. Thus, we expanded the modules we defined to contain adequate detail.

Results

In the end, we designed a modular ontology for polymer science concepts. It contains 19 modules concerning the following topics: Node, Project, Collection, Experiment, Inventory, Material, Property, Mechanical Property, Viscoelastic Property, Electrical Property, Process, Physical Process, Characterization Method, Data, Computation, Computational Process, Reference, Citation, and Software. Figure 1 provides an example of the schema diagram for the Mechanical Property module. The rest of the modules can be found in the publicly available GitHub repository.[3]

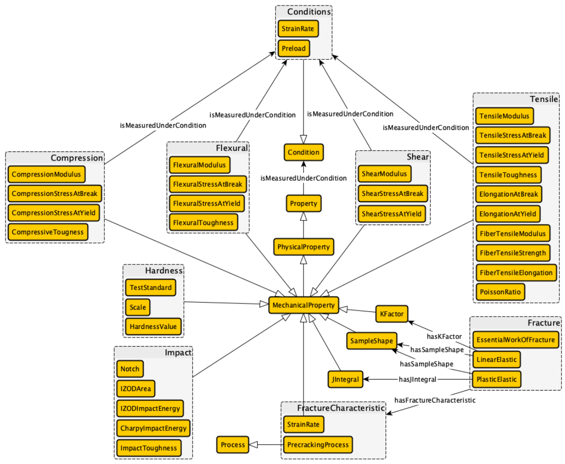

Figure 1: Schema Module for Mechanical Property

Discussion

In this brief work, we have detailed our schema that captures polymer concepts, with the goal of acting as a curated store of domain data and enabling an AI agent to assist in experimental design and execution. Having been constructed modularly, it will be adaptable and extensible as the domain evolves. Some considerations of the ontology's utilisation and improvement follow.

Since most polymer science data remains siloed, it is prudent to explore methods of obtaining and combining the data contained within them. In our case, this entails automating population of a KG using the constructed domain ontology with the objective of establishing a consolidated infrastructure, as manual approaches are time-consuming and unscalable. In related work [9], we have constructed an AI agent pipeline, the purpose of which is to populate the KG given a textual representation of the schema and a publication. While much work remains to perfect the pipeline, we have evaluated various parameter configurations for comparatively improved KG population. A potential avenue to improve results even further involves refining the schema to more closely resemble the mapping of data to ontology concepts, as this may reduce the liberty taken by an LLM and the ensuing hallucinations.

Furthermore, formalisation of the ontology in Web Ontology Language remains to be completed. By defining axioms for each of the modules, we aim to paint a more semantically clear picture of the ontology and limit ambiguity. This would prove beneficial not only in ensuring the accurate modelling of data but also in allowing for formal reasoning over the ontology [10]. It has been shown that in some cases, an axiomatised representation of an ontology yields better results for inference of AI agents due to heightened coverage [11]. Thus, we must determine if including axioms in the representation of a schema, either separately or in addition to textual triples, would improve LLM inference in the case of Saini et al. [9] and similar studies.

In the broader construction of KGWRAPS, this ontology will allow the research assistant to draw key insights necessary to predict experimental results. Specifically, by embedding the KG developed from it in a vector space, we will be able to make correlations among aspects like polymer properties and structures, along with experimental conditions, via link prediction and clustering tasks. These correlations will be crucial both in the reproducibility of existing and discovery of new polymers, as particular known characteristics are used to yield desired qualities in polymers. Thus, KGWRAPS will also be able to identify knowledge gaps and use the underlying repository of knowledge contained within the KG to fill them.

Acknowledgment. The authors acknowledge support from DAGSI RX24-22.

[3] `https://github.com/kastle-lab/KGWRAPS/tree/master/schema/diagrams/domain_schemas`

References

[1] T. Schuett, P. Endres, T. Standau, S. Zechel, R. Q. Albuquerque, C. Brütting, H. Ruckdäschel, and U. S. Schubert, "Application of Digital Methods in Polymer Science and Engineering," *Advanced Functional Materials*, vol. 34, no. 8, p. 2309844, 2024. doi: 10.1002/adfm.202309844 .

[2] A. Riedel, J. Gerlach, M. Dietsch, S. Herbst, F. Engelmann, N. Brehm, and T. Pfeifroth, "A deep learning-based worker assistance system for error prevention: Case study in a real-world manual assembly," *Advances in Production Engineering & Management*, vol. 16, no. 4, pp. 393–404, Dec. 2021. doi: 10.14743/apem2021.4.408 .

[3] D. Garisto, "Claimed superconductor LK-99 is an online sensation — but replication efforts fall short," *Nature*, vol. 620, no. 7973, pp. 253–253, Aug. 2023. doi: 10.1038/d41586-023-02481-0 .

[4] W. Wetsig, "Open-source software enables scientists to expedite research," https://www.afrl.af.mil/News/Article-Display/Article/2775183/open-source-software-enables-scientists-to-expedite-research/, 2021.

[5] M. Abolhasani and E. Kumacheva, "The rise of self-driving labs in chemical and materials sciences," *Nature Synthesis*, vol. 2, no. 6, pp. 483–492, Jun. 2023. doi: 10.1038/s44160-022-00231-0 .

[6] D. Walsh, W. Zou, L. Schneider, R. Mello, M. Deagen, J. Mysona, T.-S. Lin, J. De Pablo, K. Jensen, D. Audus, and B. Olsen, "CRIPT: A Scalable Polymer Material Data Structure," Sep. 2022. doi: 10.26434/chemrxiv-2022-xpz37 .

[7] C. Shimizu, K. Hammar, and P. Hitzler, "Modular ontology modeling," *Semantic Web*, vol. 14, no. 3, pp. 459–489, Apr. 2023. doi: 10.3233/SW-222886 .

[8] "NanoMine: an Online Platform of Materials Genome Prediction for Polymer Nanocomposites." [Online]. Available: https://brinsonlab.pratt.duke.edu/research/mgp/Nanomine

[9] A. Saini, C. D. Jaldi, J. Ethier, and C. Shimizu, "Research directions for ontology-guided domain-specific knowledge graph population using llms," in *Proceedings of the Seventh International Knowledge Graph and Semantic Web Conference*, 2025, to Appear.

[10] C. Shimizu, P. Hitzler, Q. Hirt, D. Rehberger, S. G. Estrecha, C. Foley, A. M. Sheill, W. Hawthorne, J. Mixter, E. Watrall, R. Carty, and D. Tarr, "The enslaved ontology: Peoples of the historic slave trade," *Journal of Web Semantics*, vol. 63, p. 100567, Aug. 2020. doi: 10.1016/j.websem.2020.100567 .

[11] C. D. Jaldi, A. Saini, E. Ghiasi, O. D. Eziolise, and C. Shimizu, "Knowledge conceptualization impacts rag efficacy," in *Proceedings of the Seventh International Knowledge Graph and Semantic Web Conference*, 2025, to Appear.