Exploring Faceted Ontologies for the Indexing of Materials Science Literature

Elijah Kellner, David Venator, Carl Shen, Harry Zhao, Xintong Zhao, Scott McClellan, Jane Greenberg
Metadata Research Center (MRC), College of Computing and Informatics, Drexel University, Philadelphia, PA

Motivation

- Materials scholarly big data has expanded greatly with increasing internet access.
- With millions of available documents there is now too much information to review manually.
- Tools such as ontologies are a potential way to quickly and automatically extract knowledge from articles.
- Extracted knowledge assists discovery efforts for novel materials through inferential reasoning based on relationships between concepts.

Ontology Development

- Searched literature for review and synthesis articles containing information about the processing, structure, properties, and performances of materials
- Created simple ontologies using WebProtégé
- Identified synonyms and preferred labels for extracted terms
- Connected terms using hierarchies and other lexical-semantic relationships

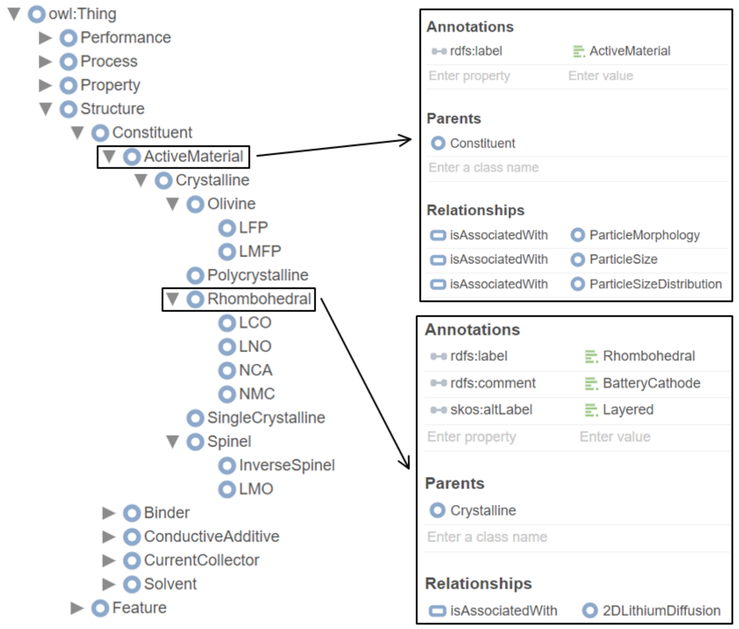

Figure 1: WebProtégé Ontology Structure and Relationship Examples

Goals, Methods, & Steps

Goals

- Develop two baseline faceted ontologies for lithium-ion battery cathode materials and aerogels.
- Input ontologies into **HIVE4MAT (Helping Interdisciplinary Vocabulary Engineering for Materials Discovery)** and demonstrate automatic indexing with the baseline ontologies.

Methods & Steps

- Explore literature and extract terms related to the processing, structure, properties, and performance of materials.
- Add terms to ontologies and identify synonyms, hierarchies, and other relationships between terms.
- Add ontology to HIVE4MAT and test indexing articles.

Ontological Relationships

- Terms in the ontologies that were related but nonhierarchical were connected using 5 relationships.
- These relationships were isAssociatedWith, isPrecededBy, isDependentOn, isDerivedFrom, and isSynthesizedBy.
- Use of these relationships allowed for connections across facets and gave greater structure to the ontologies.

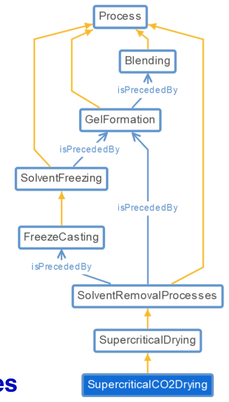

Figure 2: WebProtégé Ontological Entities Relationship Display

Indexing using the Ontologies

- Once completed the ontologies were exported from WebProtégé and incorporated into HIVE4MAT.
- Indexed multiple articles in HIVE4MAT through both the aerogel and lithium-ion battery cathode material ontologies.
- To index articles were first converted to .txt files and then uploaded to HIVE4MAT

HIVE Vocabulary Server
Helping Interdisciplinary Vocabulary Engineering for Materials Discovery (HIVE-4-MAT)

HIVE4MAT Demonstration

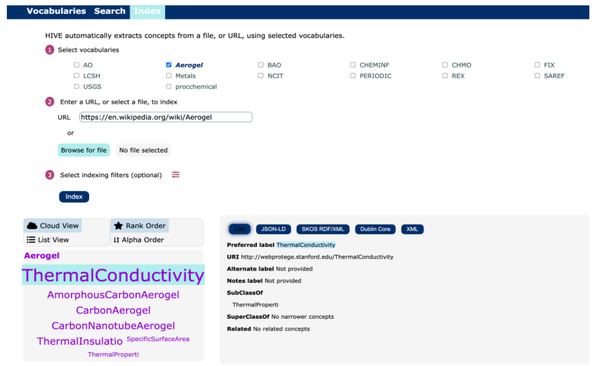

Figure 3: HIVE4MAT Automatic Indexing Output Example

- Demonstration of Aerogel ontology automatic indexing with standardized ontology vocabulary.
- HIVE4MAT expedites indexing and permits the use of multiple ontologies during a single indexing operation. Synonymous concepts can be missed if they were not included in the ontologies.

Summary/Future Work

Accomplishments

Successfully created two baseline faceted materials science ontologies focused on aerogels and lithium-ion battery cathode materials and demonstrated indexing articles in HIVE4MAT using ontologies.

Next steps

- Expand existing ontologies to include more terms and relationships and create more ontologies covering different materials and with focuses on specific facets.
- Pursue performance evaluation of HIVE4MAT for indexing scholarly big data.

Acknowledgments

We acknowledge support of NSF-HDR-OAC #2118201 Institute for Data Driven Dynamical Design. We also acknowledge REU infrastructure support via NSF-EEC-ENG #1949718 Smart Manufacturing Research Experiences for Undergraduates (SMREU).

HIVE4MAT: https://hive4mat.cci.drexel.edu/