Initial development of an ontology for the semiconductor domain – SemicONTO

Huanyu Li[1,2,*], Chuanfei Wang[3,4,*] and Patrick Lambrix[1,2]

[1] *Department of Computer and Information Science, Linköping University, Linköping, Sweden*
[2] *Swedish e-Science Research Centre, Linköping, Sweden*
[3] *School of Materials Science and Engineering, Ocean University of China, Qingdao, China*
[4] *Department of Science and Technology, Linköping University, Norrköping, Sweden*

Abstract

Materials science domain is facing the fourth paradigm of science, i.e., data-driven science, which also encompasses the first three paradigms based on theory, experiment, and simulation. The semiconductor domain is one of many sub-domains of materials science, involving both mathematical models-based simulations and conventional experiments to study materials. A significant challenge in the semiconductor domain is the lack of interoperability between materials simulation data and experimental data. While there is existing work, such as the Materials Design Ontology, that enhances the interoperability of simulation data, there remains a need for representing experimental data with rich semantics. To improve the findability, accessibility, interoperability, and reusability of semiconductor experimental data, we present the initial steps in developing a semiconductor domain ontology, SemicONTO.

Keywords

Semiconductor, Ontology, Data Modeling,

1. Introduction

The materials science domain, aiming at better understanding and discovering materials characteristics and properties, has a typical workflow that relies on both experiments and simulations [1]. Taking the semiconductor field as a sub-field of materials science as an example, researchers may first use computational methods to model semiconductor materials, and then compare simulation results with experimental results, such as from spectroscopy experiments to learn materials' properties. Along such a workflow, massive materials data may be generated from computational simulations or experiments. Therefore, the materials science domain faces the big data challenges of volume, variety, velocity as well as variability [2, 3]. To organize materials data in a structured way and to share such data in a FAIR (Findable, Accessible, Interoperable and Reusable) [4] manner can not only help users better understand the data, but also enable the data to be used efficiently in different applications (e.g., machine learning-based materials design [5]). Several global efforts focus on dealing with data challenges in the domain,

SeMatS 2024: The 1st International Workshop on Semantic Materials Science co-located with the 20th International Conference on Semantic Systems (SEMANTiCS), September 17-19, Amsterdam, The Netherlands.

*Corresponding author.
huanyu.li@liu.se (H. Li); wangchuanfei@ouc.edu.cn (C. Wang); patrick.lambrix@liu.se (P. Lambrix)
http://huanyuli.se (H. Li); https://www.ida.liu.se/~patla00/ (P. Lambrix)
0000-0003-1881-3969 (H. Li); 0000-0002-6811-6689 (C. Wang); 0000-0002-9084-0470 (P. Lambrix)

e.g., the Materials Genome Initiative (MGI),[1] and the European Materials Informatics Network (EuMINe).[2] MGI aims at assembling and curating databases combining materials property data from both experiments and simulations. EuMINe targets at harmonizing resources and approaches within the domain of materials science.

To make materials data FAIR, ontologies has been realized as a way to represent semantics of materials data and thus alleviate the heterogeneity issue among different data sources [6, p. 22]. Essentially, ontologies contain: (1) concepts representing set of entities for a domain; (2) instances of the actual entities; (3) relationships and axioms representing facts which are always true in the topic area of ontologies. By developing ontologies, domain-based terms are possible to be organized at a conceptual and general level and to be connected to each other semantically. There exists many domain ontologies for the materials science field for different application purposes. Although efforts have been made to enable computational materials data FAIR, there is not much work that focuses on promoting experimental data encoded with semantics. Semiconductor domain is one of such domains that lack semantics-aware data management (e.g., based on ontologies).

In this **short paper**, we present initial work of developing the SemicONductor onTOlogy (SemicONTO) version 0.1. The remainder of the paper is as follows. We introduce the related work in Section 2. Then, the development and content of SemicONTO is presented in Section 3. We maintain SemicONTO in a public GitHub repository,[3] and publish the ontology with a permanent URI[4] through the w3id service. In Section 4, we present its initial usage. Finally, in Section 5, we present concluding remarks and future work.

2. Related work

Our prior work presented in [3, 6, 7] has studied existing ontologies that are relevant to the materials science domain. For instance, the top-level ontology, Elementary Multiperspective Material Ontology (EMMO)[5] aims at developing a standard ontology framework according to knowledge of materials modeling and characterization. Given that materials science is a broad domain that has diverse sub-domains, existing work also focuses on knowledge representation for these sub-domains. For instance, MatOWL [8], NanoParticle Ontology [9], MMOY [10], and Dislocation Ontology [11] focus on representing materials. Furthermore, the study and research on materials involve two basic activities that are materials simulations and experiments. The former is to run computational models, while the latter is to conduct particular experiments on materials, for studying and investigating materials' structures, properties, etc. There has been some work focusing on these two directions. For instance, our prior work, Materials Design Ontology (MDO) [12, 7] is the first materials design ontology focusing on formally representing calculated materials data. The Platform MaterialDigital Ontology (PMD) [13] and its extended version (PMDco) [14] focus on describing materials science and engineering (MSE)

[1]https://www.mgi.gov/
[2]https://www.cost.eu/actions/CA22143/
[3]https://github.com/huanyu-li/SemicONTO
[4]http://w3id.org/SemicONTO
[5]https://github.com/emmo-repo/EMMO

processes. However, to our knowledge, there is no existing work on formally representing domain knowledge for the semiconductor field.

3. Ontology development

In this work, we choose the Linked Open Terms (LOT) [15] methodology for developing our SemicONductor onTOlogy (SemicONTO). LOT is a lightweight methodology aiming at aligning the ontology development with software development agile practices. There are also other ontology development methodologies such as NeOn [16], "Ontology Development 101" [17] and eXtreme Design (XD) [18]. Similar as LOT, they all include general ontology development steps such as requirements analysis, ontology implementation, publication and maintenance. All the above methodologies have been used in various applications. In addition, LOT is the first methodology that has a focus on publishing ontologies in accordance with the FAIR principles. Therefore we choose LOT for developing SemicONTO.

3.1. Requirements analysis

To develop SemicONTO, knowledge engineers and a domain expert (i.e., the second author) discussed domain interests regarding knowledge representation and data management for semiconductors. During the discussions, we identified some use cases and competency questions.

Use cases. SemicONTO aims to capture knowledge regarding: (1) composition of semiconductors in terms of basic chemical composition as well as structure (e.g., donors and acceptors for semiconductors); (2) organic semiconductor-based experiments.

Competency questions. After discussions between domain experts and knowledge engineers, we formulate four competency questions that the developed ontology should be capable to answer.

- **CQ1**: What are the different kinds of semiconductors?
- **CQ2**: What is the composition information of a semiconductor material?
- **CQ3**: Does a semiconductor have donors or acceptors? If it does, what are the donor and acceptor materials?
- **CQ4**: What are the different steps for a semiconductor experiment?

3.2. Development and implementation process

The ontology requirements specification process aims to clarify why the ontology is being developed, and identify requirements in the form of use cases and competency questions. For developing SemicONTO, we identified use cases and competency questions as introduced in Section 3.1. Then, the implementation process includes conceptualization based on the identified requirements, reusing existing ontologies and design patterns, encoding the conceptualization

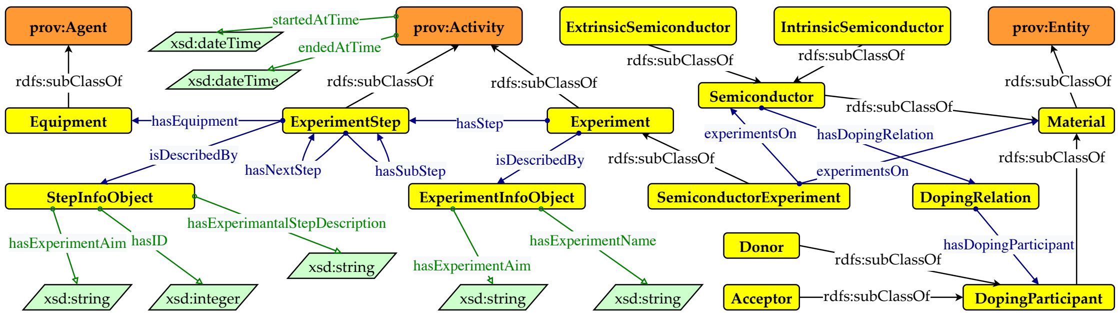

Figure 1: The core concepts and relationships of SemicONTO.

results, and evaluation. Tools including the OntOlogy Pitfall Scanner (OOPS!) [19] and OOPS! for FAIR (FOOPS!) [20], recommended by LOT, are used in evaluation.[6]

3.3. Conceptualization and formalization

The core concepts and relatiships of SemicONTO are shown in Figure 1 of which the formulation details are introduced as follows.

Semiconductor and structure. In our current ontology, a semiconductor is distinguished as either of the type *ExtrinsicSemiconductor* or *IntrinsicSemiconductor* (Axioms 1 and 2). In terms of structural composition, an extrinsic semiconductor has composing particles called acceptors and donors. An acceptor is a material (e.g., an atom or a molecule) which can bind to or accept an electron, therefore is able to form a positive hole in an extrinsic semiconductor (Axiom 3). A donor is a material (e.g., an atom or a molecule) which can provide an electron, therefore contributes a conducting electron in an extrinsic semiconductor (Axiom 4). Moreover, we define a relationship, *hasStructure* to represent that a material can be associated with some structural information in terms of chemical compositions (Axioms 5 and 6), and reuse the *Structure* and *Composition* concepts from MDO.

$$ExtrinsicSemiconductor \sqcup IntrinsicSemiconductor \sqsubseteq Semiconductor \quad (1)$$

$$ExtrinsicSemiconductor \sqcap IntrinsicSemiconductor = \emptyset \quad (2)$$

$$ExtrinsicSemiconductor \sqsubseteq \exists hasAcceptor.Material \sqcap \forall hasAcceptor.Material \quad (3)$$

$$ExtrinsicSemiconductor \sqsubseteq \exists hasDonor.Material \sqcap \forall hasDonor.Material \quad (4)$$

$$\top \sqsubseteq \forall hasStructure.Material \quad (5)$$

$$\exists hasStructure.\top \sqsubseteq mdo{:}Structure \quad (6)$$

Experiments and experimental steps. In our ontology, we represent detailed operations in an experiment. An *Experiment* can have a number of *Experimental Steps* (Axiom 7). Furthermore, an experimental step can have sub-steps and uses corresponding *Equipments*. The *hasSubStep*

[6]Pitfall report from OOPS! and FOOPS!: https://github.com/huanyu-li/SemicONTO/issues/3

relationship is transitive with *Experimental Step* as both the domain and range (Axioms 8, 9 and 10). In addition, we define *Experiment* and *Experimental Step* as sub-concepts of *Activity* from the Provenance Ontology (PROV-O).[7]

$$Experiment \sqsubseteq \exists\, hasExperimentalStep.ExperimentalStep \sqcap \forall\, hasExperimentalStep.ExperimentalStep \quad (7)$$

$$hasSubStep \circ hasSubStep \sqsubseteq hasSubStep \quad (8)$$

$$\top \sqsubseteq \forall hasSubStep.ExperimentalStep \quad (9)$$

$$\exists\, hasSubStep.\top \sqsubseteq ExperimentalStep \quad (10)$$

Contextual information of experiments and experimental steps. To represent contextual information of experiments and experimental steps, we use the **InformationObject** based on DOLCE [21]. Each experiment or experimental step is described by a specialized information object in which such information object captures information to describe an experiment or experimental step.

$$Experiment \sqsubseteq \exists\, isDescribedBy.ExperimentInfoObj \sqcap \forall\, isDescribedBy.ExperimentInfoObj \quad (11)$$

$$ExperimentalStep \sqsubseteq \exists\, isDescribedBy.StepInfoObj \sqcap \forall\, isDescribedBy.StepInfoObj \quad (12)$$

4. Usage of SemicONTO

We identified four relevant experiments for organic semiconductor materials as uses cases to be represented by the developed ontology. These four experiments are (1): *Battery Device Preparation and Parameter Characterization Experiment*; (2): *External Quantum Efficiency Testing Experiment*; (3): *Single Electron Device Fabrication and Charge Mobility Testing Experiment*; and (4): *Photoelectron Spectroscopy Testing of Film Properties and Interface Properties Experiment.* We populate the developed ontology with instances to represent these four kinds of experiments, and then publish a SPARQL server.[8] Figure 2 showcases part of an instantiation for the first kind of experiment.

In accordance with competency questions outlined in Section 3, we formulate four SPARQL queries. In Listing 1, we provide an example SPARQL query corresponding to CQ4. For each competency question (CQ1-CQ4), we formulated one or more SPARQL queries.

5. Concluding remarks and future work

This paper introduces our initial work on SemicONTO, an ontology for the semiconductor domain. The ontology captures basic semantics that can represent experiments for semiconductors. We showcase usage scenarios in which SemicONTO is utilized: **(1)** to annotate four different kinds of experiments on semiconductor materials, and **(2)** for asking queries over such

[7] https://www.w3.org/TR/prov-o/
[8] SPARQL server showcase: https://huanyu-li.github.io/SemicONTO/demo/

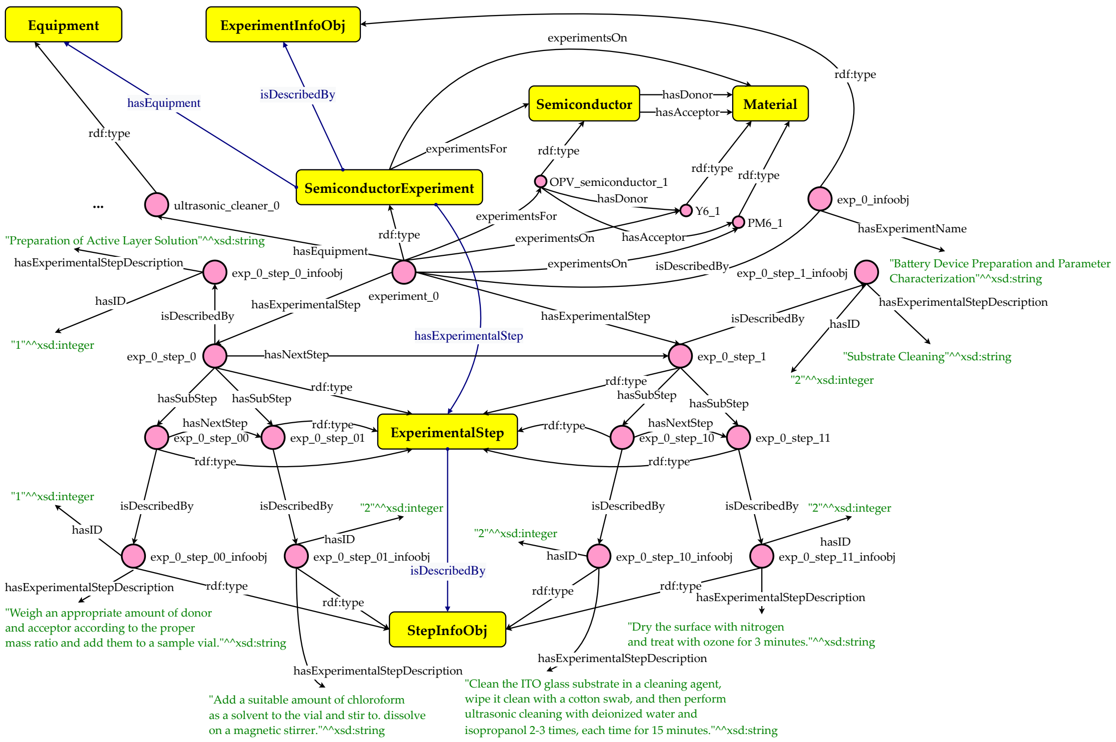

Figure 2: Part of an Instantiation for a Battery Device Preparation and Parameter Characterization Experiment.

Listing 1: Example SPARQL query (CQ4, What are the different steps for a semiconductor experiment?).

```
PREFIX semiconto: <http://w3id.org/SemicONTO/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?experiment ?sid ?ssid ?desc WHERE {
    ?experiment   rdf:type   semiconto:SemiconductorExperiment .
    ?experiment   semiconto:hasExperimentalStep   ?step .
    ?step   semiconto:hasSubStep   ?substep .
    ?step   semiconto:isDescribedBy   ?sinfo .
    ?substep   semiconto:isDescribedBy   ?ssinfo .
    ?sinfo   semiconto:hasID   ?sid .
    ?ssinfo   semiconto:hasID   ?ssid .
    ?ssinfo   semiconto:hasExperimentalStepDescription   ?desc .
}
ORDER BY ?sid ?ssid
```

annotated data. Since this paper presents the initial work of SemicONTO (version 0.1), we will continue working on SemicONTO towards introducing new concepts and relationships.

For instance, representing properties and quantities, annotating more material experiments, and aligning SemicONTO with general ontologies such as EMMO. Moreover, we will discuss with domain experts in terms of describing experimental steps in a more structured way. For instance, to represent inputs, outputs and condition details of an experimental step.

Acknowledgments

This work has been financially supported by the Swedish National Graduate School in Computer Science (CUGS) and the EU Horizon project Onto-DESIDE (Grant Agreement 101058682).

References

[1] A. Agrawal, A. Choudhary, Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science, APL Materials 4 (2016) 053208:1–10. doi:10.1063/1.4946894.

[2] P. Lambrix, R. Armiento, A. Delin, H. Li, Big Semantic Data Processing in the Materials Design Domain, in: Encyclopedia of Big Data Technologies, Springer, Cham, 2019. doi:10.1007/978-3-319-63962-8_293-1.

[3] P. Lambrix, R. Armiento, A. Delin, H. Li, FAIR Big Data in the Materials Design Domain, in: Encyclopedia of Big Data Technologies, Springer, Cham, 2022. doi:10.1007/978-3-319-63962-8_293-2.

[4] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 160018:1–9. doi:10.1038/sdata.2016.18.

[5] R. Armiento, Database-driven high-throughput calculations and machine learning models for materials design, in: Machine Learning Meets Quantum Physics, volume 968 of *Lecture Notes in Physics*, Springer, Cham, 2020. doi:10.1007/978-3-030-40245-7_17.

[6] H. Li, Ontology-Driven Data Access and Data Integration with an Application in the Materials Design Domain, Ph.D. thesis, Linköping University, Sweden, 2022. doi:10.3384/9789179292683.

[7] P. Lambrix, R. Armiento, H. Li, O. Hartig, M. Abd Nikooie Pour, Y. Li, The materials design ontology, Semantic Web 15 (2024) 481–515. doi:10.3233/SW-233340.

[8] X. Zhang, C. Hu, H. Li, Semantic query on materials data based on mapping matml to an owl ontology, Data Science Journal 8 (2009) 1–17. doi:10.2481/dsj.8.1.

[9] D. G. Thomas, R. V. Pappu, N. A. Baker, Nanoparticle ontology for cancer nanotechnology

research, Journal of Biomedical Informatics 44 (2011) 59–74. doi:10.1016/j.jbi.2010.03.001.

[10] X. Zhang, D. Pan, C. Zhao, K. Li, Mmoy: Towards deriving a metallic materials ontology from yago, Advanced Engineering Informatics 30 (2016) 687–702. doi:10.1016/j.aei.2016.09.002.

[11] A. Zainul Ihsan, D. Dessì, M. Alam, H. Sack, S. Sandfeld, Steps towards a dislocation ontology for crystalline materials, in: Proceedings of the Second International Workshop on Semantic Digital Twins co-located with the 18th Extended Semantic Web Conference (ESWC 2021), volume 2887 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2887/paper4.pdf.

[12] H. Li, R. Armiento, P. Lambrix, An Ontology for the Materials Design Domain, in: The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, volume 12507 of *Lecture Notes in Computer Science*, Springer, Cham, 2020, pp. 212–227. doi:10.1007/978-3-030-62466-8_14.

[13] M. Alam, H. Birkholz, D. Dessì, C. Eberl, H. Fliegl, P. Gumbsch, P. von Hartrott, L. Mädler, M. Niebel, H. Sack, A. Thomas, Ontology modelling for materials science experiments, in: Proceedings of the Poster & Demo track co-located with he 17th International Conference on Semantic Systems (SEMANTiCS 2021), volume 2941 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2941/paper11.pdf.

[14] B. Bayerlein, M. Schilling, H. Birkholz, M. Jung, J. Waitelonis, L. Mädler, H. Sack, Pmd core ontology: Achieving semantic interoperability in materials science, Materials & Design 237 (2024) 112603. doi:10.1016/j.matdes.2023.112603.

[15] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, R. García-Castro, Lot: An industrial oriented ontology engineering framework, Engineering Applications of Artificial Intelligence 111 (2022) 104755. doi:10.1016/j.engappai.2022.104755.

[16] M. C. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, The neon methodology for ontology engineering, in: Ontology Engineering in a Networked World, Springer, Berlin, Heidelberg, 2012, pp. 9–34. doi:10.1007/978-3-642-24794-1_2.

[17] N. Noy, Ontology development 101: A guide to creating your first ontology, 2001.

[18] V. Presutti, E. Blomqvist, E. Daga, A. Gangemi, Pattern-Based Ontology Design, Springer, Berlin, Heidelberg, 2012, pp. 35–64. doi:10.1007/978-3-642-24794-1_3.

[19] M. Poveda-Villalón, A. Gómez-Pérez, M. C. Suárez-Figueroa, OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation, International Journal on Semantic Web and Information Systems (IJSWIS) 10 (2014) 7–34. doi:10.4018/ijswis.2014040102.

[20] D. Garijo, O. Corcho, M. Poveda-Villalón, FOOPS!: An Ontology Pitfall Scanner for the FAIR Principles, in: International Semantic Web Conference (ISWC) 2021: Posters, Demos, and Industry Tracks, volume 2980 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2980/paper321.pdf.

[21] S. Borgo, R. Ferrario, A. Gangemi, N. Guarino, C. Masolo, D. Porello, E. M. Sanfilippo, L. Vieu, DOLCE: A descriptive ontology for linguistic and cognitive engineering, Applied Ontology 17 (2022) 45–69. doi:10.3233/AO-210259.