

A First Step towards a Tool for Extending Ontologies

Mina Abd Nikooie Pour¹, Huanyu Li^{1,3},
Rickard Armiento^{2,3}, and Patrick Lambrix^{1,3,4}

¹ Department of Computer and Information Science, Linköping University, Sweden

² Department of Physics, Chemistry and Biology, Linköping University, Sweden

³ The Swedish e-Science Research Centre, Linköping University, Sweden

⁴ Department of Building Engineering, Energy Systems and Sustainability Science,
University of Gävle, Sweden

`firstname.lastname@liu.se`

Abstract. Ontologies have been proposed as a means towards making data FAIR (Findable, Accessible, Interoperable, Reusable). This has attracted much interest in several communities and ontologies are being developed. However, to obtain good results when using ontologies in semantically-enabled applications, the ontologies need to be of high quality. One of the quality aspects is that the ontologies should be as complete as possible. In this paper we propose a first version of a tool that supports users in extending ontologies using a phrase-based approach. To demonstrate the usefulness of our proposed tool, we exemplify the use by extending the Materials Design Ontology.¹

1 Introduction

In many areas there is a recent interest in making data FAIR, i.e., Findable, Accessible, Interoperable, and Reusable [10]. Findable refers to the fact that data and metadata should be easy to find, accessible to the fact that it should be clear how to access the data, interoperable to the fact that the data needs to be integrated with other data and be usable by applications and workflows, and reusable to the fact that data and metadata are well described such that the data can be replicated or combined in different settings. Ontologies can alleviate some of the issues towards making data FAIR, such as in ontology-based access to multiple data sources for Interoperability and using ontologies to describe data source capabilities for Findability.

Ontologies need to be of high quality. One of the quality aspects is that the ontologies should be as complete as possible.² If an ontology is not complete,

¹ Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

² This relates to the requirement of domain coverage in [9]. In practice, knowing when an ontology is complete is difficult but it is possible to define an ‘is more complete than’ relation between ontologies which can be used for comparing completeness [5].

it may lead to missing inferences and thus incomplete results in semantically-enabled applications.

Many techniques exist for finding missing information in ontologies and extending them. The ontology extension problem that we tackle deals mainly with concept discovery and concept hierarchy derivation which are also two of the tasks in the problem of ontology learning [3]. Therefore, there are many techniques that can be used as shown by a recent survey [2] that discusses 140 research papers. These techniques are usually linguistics-, statistics-, or logic-based. One such technique is phrase-based extension where frequent phrases are collected from text and proposed to a domain expert as candidates for creating concepts with similar names. An example of such a method that has been applied to ontologies in the materials science field is found in [6,1]. When using the method the phrases were presented in an Excel file to the domain expert who then annotated the file with the concepts and axioms to add to the ontology. Then this information was transferred into OWL format manually or via an ontology editor. It was clear that user support for these tasks was needed. Therefore, in this paper we present a first version of a tool that takes as input a list of phrases (generated using the method in [6,1] or any other method that outputs phrases) and supports the user in turning phrases into concepts and in defining axioms related to the new concepts. We show the use of the tool in a use case where we extend the Materials Design Ontology.

The remainder of the paper is organized as follows. In the background section we describe the method for ontology extension on which our proposed tool is based (sub-section 2.1), as well as the ontology that we extend to exemplify the method (sub-section 2.2). In section 3 we briefly discuss the purpose, intended users and development of the tool. We describe our tool in section 4. In section 5 we show a use case in the materials science domain where we extend the Materials Design Ontology and briefly discuss a user interface evaluation. The paper concludes in section 6.

2 Background

2.1 Method for extending ontologies

Our tool is based on the approach for extending ontologies presented in [6] and extended in [1]. Figure 1 gives an overview.

Given a corpus of unstructured text, in a first step a phrase-based topic model is produced. The output of this step is a list of frequent phrases as well as a number of topics with representative phrases. In the method described in [6,1] this is done using the ToPMine system [4]. The approach in [6] uses a formal topical concept analysis-based approach to mine topics. The frequent phrases, the topics, and the result of the formal topical concept analysis are suggestions that a domain expert should validate or interpret and relate to concepts in the ontology.

The method in [6] introduced different categories for the phrases related to processing that could lead to the ontology extension based on the phrases. The

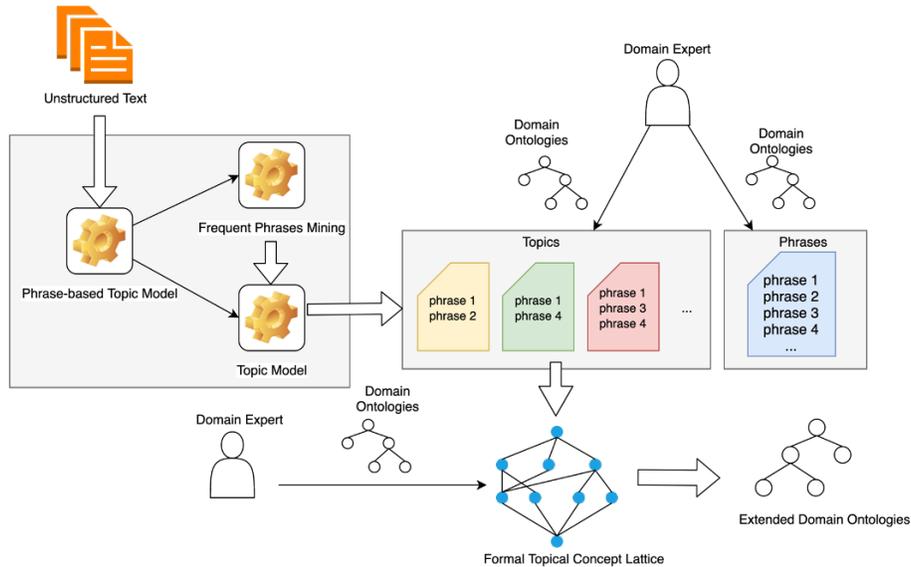


Fig. 1. Approach: The upper part of the figure shows the creation of a phrase-based topic model with unstructured text as input and phrases and topics as output. The lower part shows the formal topical concept analysis with as input topics and as output a topical concept lattice. In both parts a domain expert validates and interprets the results. [6].

ADD category represents the addition of concepts where the new concept has the same name as a phrase. For the ADD-m category a new concept is added with a name that is a modification of a phrase. The EXIST and EXIST-m categories do not represent a change, but indicate that there already is a concept in the ontology with the same name as the phrase or a modified version of the phrase, respectively. The No category represents that the phrase cannot be used to create a concept for the ontology, while the No-g category represents the fact that the phrase should not lead to a new concept in this ontology, but could be used in a more general ontology. We also introduce here NEW which represents the addition of a new concept that is not related to a phrase. We note that this categorization is interesting for experimenting and gathering data regarding the kinds of operations a user performs when extending an ontology, but not necessarily for an end user using the tool.

During the work on extending ontologies using this approach, it became clear that the domain and knowledge engineering experts needed tool support. In this paper we have implemented a tool for helping a domain expert validating phrases and creating concepts and axioms related to these phrases in the ontology. The validation of topics and axioms related to topics is left for future work. Large parts of the current functionality related to phrases will be reusable for topics, but some new visualizations will be needed.

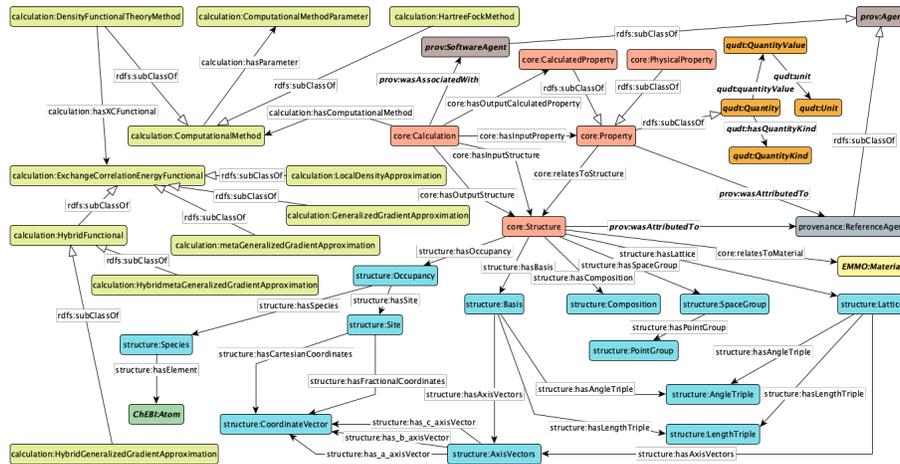


Fig. 2. The Materials Design Ontology [7]. The different modules *Core*, *Structure Calculation* and *Provenance* as well as other connected ontologies have different colors.

2.2 Materials Design Ontology

The ontology that we use to exemplify our approach is the Materials Design Ontology (MDO) presented in [7] and available at <https://w3id.org/mdo>. MDO is an ontology covering basic concepts for materials design (Figure 2). The development was guided by the schemas of the Open Databases Integration for Materials Design (OPTIMADE³) project. The OPTIMADE project aims at making materials databases interoperable by developing a common API and the consortium includes many of the important database providers in the field. Therefore, the OPTIMADE schemas are based on a consensus reached by several of the important stakeholders in the field. The aim is to use MDO for providing semantic search and integration of materials databases, e.g., via OPTIMADE.

3 Requirements and Development

Before we present our tool in section 4, we discuss the system requirements to clarify some of its aspects and use the dimensions of the design space for user studies in Semantic Web contexts in [8] as guidelines.

Purpose. The purpose of the tool is to support a user in extending an ontology using a phrase-based approach. The tool supports the user in creating concepts from phrases as well as in adding subsumption axioms related to the new concepts. The main purpose of the tool is *creating and managing* as defined in [8] as it is essentially an ontology authoring tool that supports the creation of new concepts and axioms for an ontology. However, there is also an aspect

³ <https://www.optimade.org/>

of *learning and understanding* as a user needs to understand the phrases, the concepts and their contexts to be able to make informed decisions about which phrases can lead to concepts and axioms that should be added to the ontology. Therefore, in addition to supporting tasks related to creating, adding and managing concepts and axioms based on phrases, also tasks related to exploration and searching for information in the frequent phrases and ontology need to be supported.

Users. The tool is supposed to be used by users or teams of users that have skills and experience in two fields. As it is an ontology authoring tool the user team needs to be well-versed in ontology development and understand such things as what an ontology can contain, how reasoning works and what the consequences are of adding an axiom to an ontology. Further, the user team also needs expertise in the domain of the ontology, e.g., in our use case exemplifying the use of the tool, the user team needs expertise in materials science.

Development. The tool was developed in an agile way, with meetings about once a week with ontology engineering and materials science experts. This allowed for fast feedback and the discussions led to changes in the design choices. Some examples of changes were the use of tabs instead of windows, the addition of the possibility to mark phrases as finalized, and the possibility to add additional concepts when defining a new concept. One item that we have currently kept in the tool is the ability to classify the type of action (using ADD, ADD-m, EXIST, EXIST-m, No, No-g, see explanation in section 2.1). In the final version of the tool for end users this is not needed and the parts of the user interface referring to this will be removed, but while experimenting with the tool this gives us valuable information.

4 Ontology Extension System

In this section we describe our tool with its functionality and current user interface. We use screenshots from extending MDO. We decided to model the interface in a way that reflects the workflow of the user. There are three main steps: *Set-up* where ontology and phrases are loaded, *From Phrase to Concept* where phrases are used to define concepts, and *From Concept to Axiom* where axioms related to the concepts are defined. Each of the steps in the workflow is represented by a tab in the tool. The workflow is iterative and the steps are repeated for different phrases and concepts.

4.1 Set-up

The first step in the workflow is loading the necessary files including the ontology that we want to extend and the frequent phrases. In Figure 3 MDO has been loaded from OWL files (as four different modules) and the frequent phrases have been loaded from a file that was generated earlier.

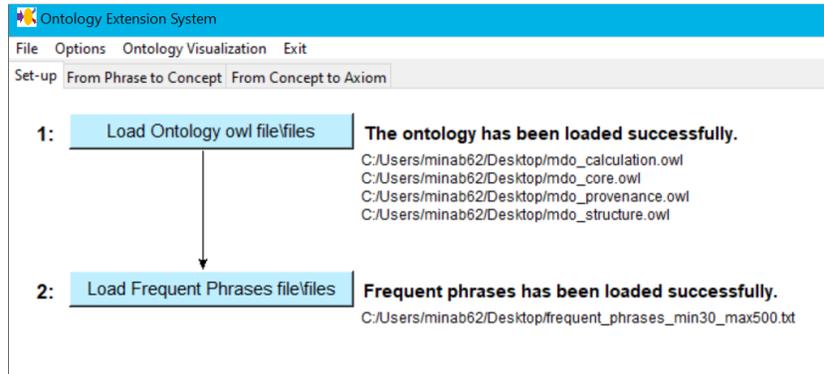


Fig. 3. The *Set-up* tab of the Ontology Extension System.

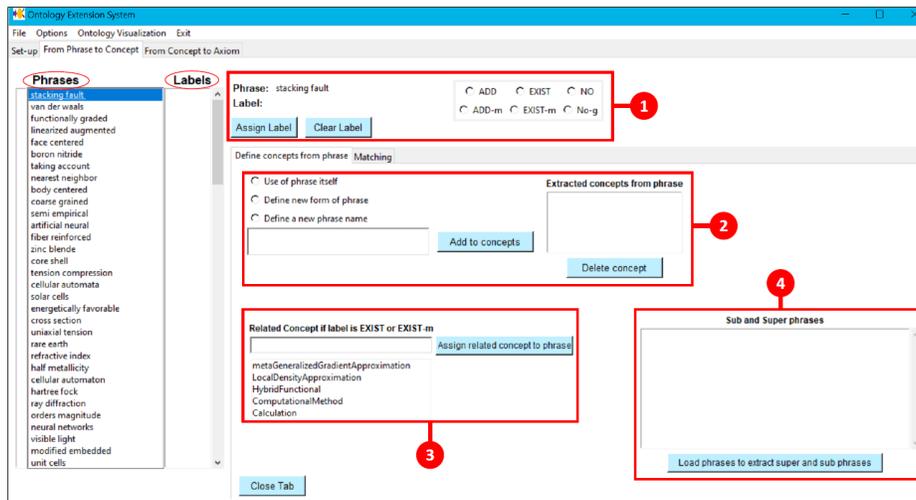


Fig. 4. The *From Phrase to Concept* tab of the Ontology Extension System.

4.2 From Phrase to Concept

In this step of the workflow the user processes the frequent phrases and decides for each phrase whether one or more concepts related to the phrase can be defined in the ontology. Figure 4 shows the tab for this phase.

At the left hand side of the *From Phrase to Concept* tab two main columns, *Phrases* and *Labels* are shown. The *Phrases* column contains the frequent phrases loaded or generated in the previous step. The *Labels* column will contain one or more of ADD, ADD-m, EXIST, EXIST-m, No, No-g in case the phrase has been (partly) processed through assigning label functionality.

By clicking on a phrase in the *Phrases* column, inner tabs are opened that support the processing of the phrase.

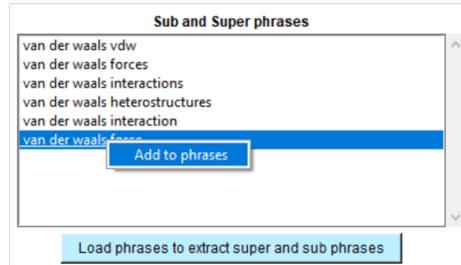


Fig. 5. Extracting sub- and super-phrases from another frequent phrases list.

In the part marked by red frame 2 in Figure 4 the user can define a new concept for the ontology. There are three options represented by the three radio buttons. For the name of the new concept the phrase can be used as is (ADD, first button), or a modified version of the phrase can be used, e.g., removing plural ‘s’ (ADD-m, second button). The third button allows to define a new unrelated concept which was deemed necessary by the domain experts (NEW). Further, all concepts created in relation to the phrase are shown in the text box on the right side of the frame.

The part marked by red frame 1 in Figure 4 is mainly interesting for our experiments and allows a user to directly assign or remove categorization labels.

In the part marked by red frame 3 in Figure 4 the user can relate an existing concept to the phrase and implicitly assign an EXIST or EXIST-m label.

The part related to red frame 4 in Figure 4 allows for searching sub- and super-phrases in other lists. This is useful, for instance, when different frequent phrase lists are generated by a system with different parameter settings. This functionality allows to also use parts of the other lists by adding these phrases as well. An example is shown in Figure 5.

In the *Phrases* list, the phrases can be marked or unmarked as processed by right-clicking on the desired phrase. This allows the user to mark progress. An additional help, using the *Matching* functionality, is to search for phrases or concepts with a name similar to a phrase. The algorithm uses fuzzy matching on the phrases and concepts and a string matcher on their stemmed version. This information could be used, for instance, to relate concepts and phrases or as extra information to the decision process regarding adding new concepts based on phrases.

4.3 From Concept to Axiom

Figure 7 shows the tab related to the next step in the workflow for extending ontologies using the phrase-based approach. In this step axioms related to the new concepts are added to the ontology.

At the left hand side of the tab there is a column of concepts which consists of concept names from the ontology loaded during the *Set-up* phase together with the ones that have been added by the user in the *From Phrase to Concept*

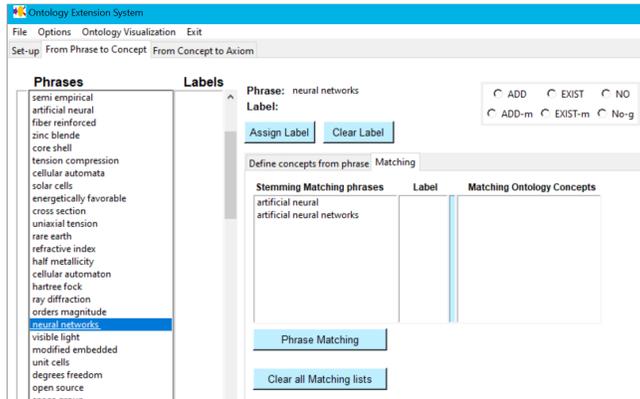


Fig. 6. Matching a phrase with other phrases and concept names.

phase. The origin relates a new concept to the phrase that was used for defining the concept.

In the upper part of the tab, axioms can be defined. By clicking on a concept in the concept list, the concept name appears at two places with radio buttons to help the user define sub- and super-concepts of the concept. These can be selected from the list shown in the text box shown on the right side.

The existing axioms as well as the axioms defined by the user are displayed in the bottom part of the tab in the list labeled as *Axioms defined for the concept*. To provide more information regarding the context of an axiom, clicking on an axiom shows the axioms related sub-concepts of the concept in the left-hand side of the axiom and super-concepts of the concept in the right-hand side of the axiom. An example is given in Figure 8.

4.4 Additional functionality

In addition to the tabs described earlier other functionalities have been implemented as shown in the menu bar. The **File** menu contains functionalities related to creating, opening and saving files. For instance, the lists of phrases and concepts can be saved as Excel files including added parts such as defined axioms, which phrase was the origin for a concept, assigned labels and concepts created for a phrase. The extended ontology can be saved as an OWL file. We note that partial results can be saved and loaded again at a later time. The **Options** menu contains functionality to search phrases, concepts and axioms. Additionally, all the axioms (Figure 9) and equivalent concepts can be displayed as well. **Ontology Visualization** is used to visualize the ontology as well as the changes made. The new concepts and axioms are displayed in a different color from the concepts and axioms that were in the original ontology (Figure 10).

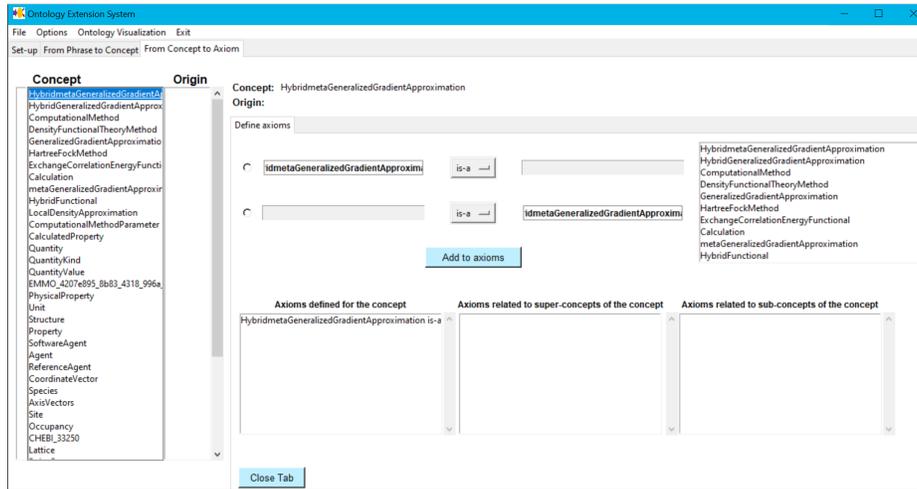


Fig. 7. The *From Concept to Axiom* tab of the Ontology Extension .

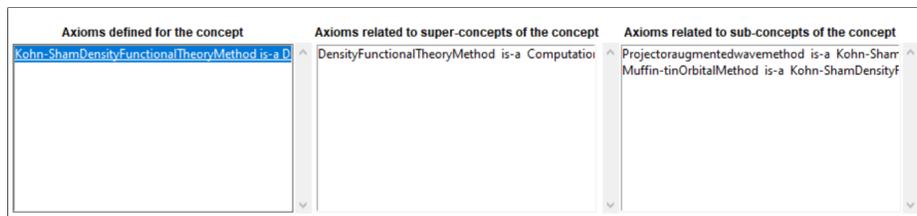


Fig. 8. Axioms related to sub- and super-concepts of concepts in an axiom.

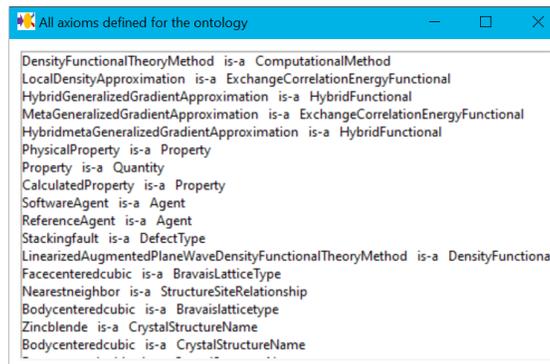


Fig. 9. Axioms of the ontology.

5 Experiments

We performed a feasibility study to evaluate the functionality and usefulness of our tool by extending MDO. Furthermore, we performed a small evaluation of the user interface.

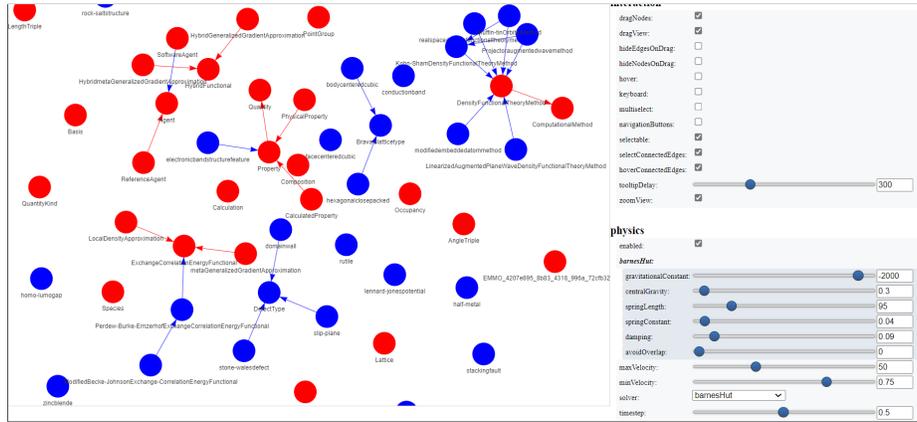


Fig. 10. Ontology visualization.

5.1 Use case - extending MDO

As a use case we decided to have a domain expert extend MDO using our tool. We used the list of 131 frequent phrases that was used in an earlier study described in [1] as well as the same domain expert.

The final result contains 29 new concepts (see Table 1) and 27 new axioms to be added to MDO.

The domain expert found that the tool gave support to the process of extending the ontology and was a large improvement over the Excel files that saved time and made the process easier. In particular, the fact that the tool guided the user through the different phases of the workflow was helpful. Further, the tool guides the user in different tasks and suggests possibilities. Also being able to search in the different lists and marking progress, as well as providing more information regarding the concepts and phrases, was very helpful. The labeling of the different kinds of changes was not always that easy and in the future we will remove this part for the end user. When we want to gather data about this, we will investigate in ways to do the labeling automatically.

When comparing the result with the result of [1], we noted some differences in the list of added concepts. The main reason for these differences was that the domain expert had a better understanding of the use of MDO and therefore sometimes made more restrictive decisions on what concepts should be part of MDO. Deciding on axioms is a task that the domain expert finds hard and it is clear that the help of an ontology engineering expert is highly recommended. Therefore, the current results regarding axioms should be considered as a first draft that should be discussed further.

5.2 User Study

As part of the design and implementation phase of our tool, we performed a small user study to receive feedback on the user interface and which resulted in some

Table 1. New concepts proposed to be added to MDO.

Stacking Fault	Linearized Augmented Plane Wave Density Functional Theory Method
Face Centered Cubic	Nearest Neighbor
Body Centered Cubic	Electronic Band Structure Feature
Rare Earth	Projector Augmented Wave Method
Brillouin Zone	Lennard-Jones Potential
Stone-Wales Defect	Rock-Salt Structure
Half-Metal	Realspace Density Functional Theory Method
Homo-Lumo Gap	Modified Becke-Johnson Exchange-Correlation Energy Functional
Valence Band	Perdew-Burke-Ernzerhof Exchange-Correlation Energy Functional
Slip-Plane	Kohn-Sham Density Functional Theory Method
Rutile	Hexagonal Close Packed
Conduction Band	Muffin-tin Orbital Method
Defect Type	Bravais Lattice Type
Zinc Blende	Modified Embedded Atom Method
Domain Wall	

changes to the user interface. The subjects were 4 PhD students in computer science at our department which did not have a background in ontologies nor in materials science. A small introduction including the purpose, the project and some basic ontology terminology was handed out beforehand. During a session with a subject via Zoom, the subject performed a number of typical tasks and was asked to think loud. The subjects did not think it was that easy to use the tool and we gathered a number of issues that led to changes that were implemented in the current version of the interface to improve the functionality of the tool. Some issues were left for a future version of the user interface. For instance, one issue that we will work on in the future is to make the interface more clean and not combine the functionality for an end user and a researcher gathering data about the use of the tool in one interface. Further, we will look into making the search functionality more easy to use and make other tasks more uniform.

6 Conclusion

In this paper we have presented the first version of a tool that supports users with phrase-based ontology extension by representing the user’s workflow and guiding the user during each step. As future work, we will implement a version of the tool that is focused on end-users. We will also take the comments gathered from the use case and the user study into account. Additionally, it may be interesting to consider other kinds of relations besides *is-a* relation, as well as the fact that sometimes additions to the ontology may lead to deletions in the ontology as well.

Acknowledgements. This work has been financially supported by the Swedish e-Science Research Centre (SeRC), the Swedish National Graduate School in

Computer Science (CUGS), the Swedish Research Council (Vetenskapsrådet, dnr 2018-04147), and the Swedish Agency for Economic and Regional Growth (Tillväxtverket, 20201438).

References

1. Abd Nikooie Pour, M., Li, H., Armiento, R., Lambrix, P.: A first step towards extending the materials design ontology. In: ESWC Workshop on Domain Ontologies for Research Data Management in Industry Commons of Materials and Manufacturing (2021)
2. Asim, M.N., Wasim, M., Khan, M.U.G., Mahmood, W., Abbasi, H.M.: A survey of ontology learning techniques and applications. *Database* **2018**, bay101:1–24 (2018), doi: 10.1093/database/bay101
3. Buitelaar, P., Cimiano, P., Magnini, B.: *Ontology Learning from Text: Methods, Evaluation and Applications*. IOS Press (2005)
4. El-Kishky, A., Song, Y., Wang, C., Voss, C.R., Han, J.: Scalable topical phrase mining from text corpora. *Proceedings of the VLDB Endowment* **8**(3), 305–316 (2014), doi: 10.14778/2735508.2735519
5. Lambrix, P.: *Completing and debugging ontologies: state of the art and challenges* (2020), arXiv:1908.03171
6. Li, H., Armiento, R., Lambrix, P.: A method for extending ontologies with application to the materials science domain. *Data Science Journal* **18**(1) (2019), doi: 10.5334/dsj-2019-050
7. Li, H., Armiento, R., Lambrix, P.: An ontology for the materials design domain. In: Pan, J.Z., Tamma, V.A.M., d’Amato, C., Janowicz, K., Fu, B., Polleres, A., Seneviratne, O., Kagal, L. (eds.) *The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference, Athens, Greece, November 2-6, 2020, Proceedings, Part II*. *Lecture Notes in Computer Science*, vol. 12507, pp. 212–227. Springer (2020), doi: 10.1007/978-3-030-62466-8_14
8. Pesquita, C., Ivanova, V., Lohmann, S., Lambrix, P.: A framework to conduct and report on empirical user studies in semantic web contexts. In: Faron-Zucker, C., Ghidini, C., Napoli, A., Toussaint, Y. (eds.) *Knowledge Engineering and Knowledge Management - 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings*. *Lecture Notes in Computer Science*, vol. 11313, pp. 567–583. Springer (2018), doi: 10.1007/978-3-030-03667-6_36
9. Sabou, M., Fernandez, M.: Ontology (network) evaluation. In: Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E., Gangemi, A. (eds.) *Ontology engineering in a networked world*, pp. 193–212. Springer (2012), doi: 10.1007/978-3-642-24794-1_9
10. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J., Groth, P., Goble, C., Grethe, J.S., Heringa, J., ’t Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B.: The FAIR guiding principles for scientific data management and stewardship. *Scientific data* **3**, 160018:1–9 (2016), doi: 10.1038/sdata.2016.18