

An Approach to Support Collaborative Ontology Construction

Kais TAHAR^{a,1}, Michael SCHAAF^a, Franziska JAHN^a, Christian KÜCHERER^b,
Barbara PAECH^b, Heinrich HERRE^a and Alfred WINTER^a

^a*Institute of Medical Informatics, Statistics and Epidemiology, Medical Faculty,
University of Leipzig*

^b*Software Engineering Heidelberg, Institute for Computer Science, University of
Heidelberg*

Abstract. The increasing number of terms used in textbooks for information management (IM) in hospitals makes it difficult for medical informatics students to grasp IM concepts and their interrelations. Formal ontologies which comprehend and represent the essential content of textbooks can facilitate the learning process in IM education. The manual construction of such ontologies is time-consuming and thus very expensive [3]. Moreover, most domain experts lack skills in using a formal language like OWL[2] and usually have no experience with standard editing tools like Protégé² [4,5]. This paper presents an ontology modeling approach based on *Excel2OWL*, a self-developed tool which efficiently supports domain experts in collaboratively constructing ontologies from textbooks. This approach was applied to classic IM textbooks, resulting in an ontology called SNIK. Our method facilitates the collaboration between domain experts and ontologists in the development process. Furthermore, the proposed approach enables ontologists to detect modeling errors and also to evaluate and improve the quality of the resulting ontology rapidly. This approach allows us to visualize the modeled textbooks and to analyze their semantics automatically. Hence, it can be used for e-learning purposes, particularly in the field of IM in hospitals.

Keywords. Formal Ontology, Ontology Construction, OWL, Semantic Web, SPARQL, Ontology Evaluation, Information Management, e-learning, Education

1. Introduction and Related works

The increasing number of terms used in the Information Management (IM) literature, leads to challenges for medical informatics students. It is difficult for them to achieve an adequate understanding of the concepts denoted by the terms and to grasp the concepts' interrelations. We began developing a formal ontology named SNIK (Semantic Network of IM in Hospitals). This ontology, being a formal knowledge system, provides various benefits and advantages: it supports the learning process in the IM education, and it helps to aggregate and organize the IM entities in hospitals. Furthermore, a formal ontology is machine-readable, such that it can be used additionally for scientific analyses and queries, resulting in the explication of implicit knowledge by reasoning. Ontology construction from textual resources is an

¹ Corresponding Author: kais.tahar@imise.uni-leipzig.de

² <http://protege.stanford.edu>

elementary task for building formal ontologies related to a particular domain. Previous research in this area showed that formal ontologies can facilitate text understanding and also enable the automatic processing of its semantic meaning [1]. Choosing OWL [2] as a knowledge representation language allows us to visualize and analyze the semantics of the linked SNIK entities automatically. While the benefits of using OWL are convincing, the manual development of such an ontology is time-consuming and thus very expensive [3]. For this reason, a higher degree of automation of the process of ontology creation should be pursued. Furthermore, to have an authentic representation of IM in hospitals, experts from this domain must have easy access to the authoring process so that a close and effective collaboration with ontologists can be established. Unfortunately, there are barriers which make this task challenging. For example, the semantics of OWL and the intricacies of its development tools prevent domain experts from engaging in a collaborative ontology development [4, 5].

Related work in this area [5-8] focused on ontology population from spreadsheets using previously existing ontologies, but little attention has been paid to supporting the automatic generation of new OWL ontologies from spreadsheets. This particularly means that each relationship between classes may only be reused from other previously defined ontologies by using inline transformation rules for each relationship statement. These rules must be specified manually by domain experts during the mapping process. A further shortcoming of existing conversion tools [5-10] is that they are RDF-centric and are not designed to work directly with OWL because none of them covers all OWL constructors. For example, declarations of important constructors such as *owl:Axiom*, *owl:ObjectProperty*, *owl:DatatypeProperty*, *owl:Restriction* and *owl:annotatedTarget* are not supported. This is problematic because these constructors are necessary modeling primitives for building the SNIK ontology. To overcome these limitations, we developed a new tool, called *Excel2OWL* that enables an automatic transformation of spreadsheet content into OWL axioms without resorting to manually pre-configured transformation rules or pre-existing ontological entities encoded in OWL.

In this paper, we present a two-step ontology modeling approach to support domain experts and ontologists in collaboratively building ontologies from textbooks based on *Excel2OWL*. Furthermore, our approach provides a workflow model to support the construction, evaluation and use of SNIK ontology for e-learning purposes.

Throughout this paper, we use standard semantic web terminology that refers to formal descriptions of ontological entities [2]. The rest of the paper is organized as follows: the next section explains the ontology modeling approach. Section 3 presents the results which were obtained by applying this method to the content of two classic IM textbooks. Finally, section 4 concludes with a discussion and an outlook.

2. Methods

2.1. Ontology construction using *Excel2OWL*

Let us be given a text *Txt*. The axiomatic structure of *Txt*, consisting of the relevant concepts and their interrelations, exhibits the skeleton of the knowledge, which is represented in *Txt*. Ontology construction, then, includes the extraction of the text's axiomatic structure and its transformation into a formal ontology, specified by a set of expressions of a language such as OWL. An abstract workflow model has been applied to support the collaborative process of developing an ontology. This workflow model is

described using BPMN elements shown in figure 1. The workflow is divided into two main steps: knowledge extraction and knowledge analysis. We adopted multiple iterations of these basic steps that allowed us a continuous evaluation and refinement of SNIK ontology. Initially, domain experts manually extracted the relevant SNIK entities from classic IM textbooks [11-12] and saved the captured semantic network (SN) in the form of Excel spreadsheets. The stored spreadsheets contained 15 predefined columns and multiple user-defined SNIK entities in each row. The filled rows were consolidated by peer reviews at the end of each cycle. Excel provides powerful sort and filter features to easily review entire spreadsheet rows and to compare them against each other. This offers domain experts and ontologists an easy way of consolidating the captured SN. Nevertheless, the consolidated data still contained much redundant information. This was due to the large volume of data captured. As a consequence, the semantics of the spreadsheet content rapidly became unclear. To facilitate the visualization and comprehension of the captured SN, the spreadsheet content needed to be encoded in OWL. Contrary to the first step, which was done manually, the second step – knowledge analysis - was supported by *Excel2OWL*. This tool provides an algorithm for the automated parsing and mapping of data, captured by spreadsheets, in a non-redundant OWL ontology. In this process, SNIK concepts, their relationships, properties, annotations, instances, axioms and restrictions were extracted automatically from the spreadsheet and transformed into OWL using mapping rules.

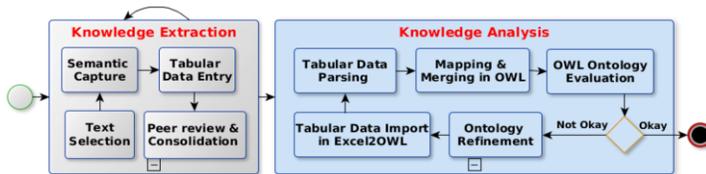


Figure 1. Basic scheme for transforming a text into an OWL ontology. The pale blue boxes show all tasks supported by *Excel2OWL*.

2.2. Error detection using semantic inconsistency inference

Once the ontology development process is finished, the next task is ontology evaluation. The earlier versions of SNIK ontology contained many inconsistency errors. The detection of all these errors is a difficult task for each ontologist. This motivated us to extend *Excel2OWL* by implementing an error detection component. With this extension, several domain axioms were verified automatically using SPARQL [13] queries and ontological reasoning. Here are some examples of domain axioms defined in first order logic that semantically constrain the arguments of usage, update and responsibility relationships to the following concept types: *Function*, *EntityType* and *Role*:

$$\forall xy (\text{uses}(x,y) \rightarrow \text{Function}(x) \wedge \text{EntityType}(y)) \quad (1)$$

$$\forall xy (\text{updates}(x,y) \rightarrow \text{Function}(x) \wedge \text{EntityType}(y)) \quad (2)$$

$$\forall xy (\text{isResponsibleFor}(x,y) \rightarrow \text{Role}(x) \wedge \text{Function}(y)) \quad (3)$$

To support the evaluation process, *Excel2OWL* detects the SNIK relationship statements which are inconsistent with these axioms; these false statements are saved as a log file.

2.3. Ontology evaluation using multiple query answering

A competency question (CQ) [14] is a generalized question that expresses a pattern of natural language questions. For example, which roles are responsible for which tasks? Or which data (data = *EntityType*) are needed to accomplish which task? In this case, roles or tasks (task = *Function*) can be substituted with relevant SNIK entities like CIO as a role and strategic alignment as a task. The use of CQs in ontology development processes is not a new idea [14] but the formalization of these CQs in SPARQL is very useful for the quality assurance. For this reason, *Excel2OWL* was extended by a multiple query answering method to test and also to support the evaluation and refinement of SNIK ontology. In this process, a list of several CQs were executed against SNIK ontology. The results were then exported as a report file (see figure 2a).

3. Results

The proposed basic scheme (see figure 1) was applied to the content of two classic IM textbooks for the management of health information systems (HIS) [11-12]. As a result, an accurate and highly expressive OWL ontology was generated automatically from the SNIK spreadsheets. It contained 1898 classes, 17 object properties, 219 data properties, 52 individuals, 12267 annotation axioms, 3993 class axioms and 3993 logical axioms that restrict the defined SNIK entities. The annotation axioms describe the concept definitions, pages, synonyms, ID, type and domain. The proposed basic scheme offers domain experts an easy and efficient way to build ontologies from textbooks without resorting to complex ontology editing tools or even OWL. Our approach thus facilitates the domain experts' involvement in a collaborative ontology development process. Furthermore, it enables ontologists to fix modeling errors and improve the quality of the developed SNIK ontology rapidly. Using *Excel2OWL*, the required development time and costs are considerably reduced. Beyond this, *Excel2OWL* facilitates the use of OWL-based software to visualize the SNIK ontology. For instance, using the *OntoGraf* plugin for *Protégé 5.0*, SNIK entities are visualized as a semantic graph (see figure 2b).

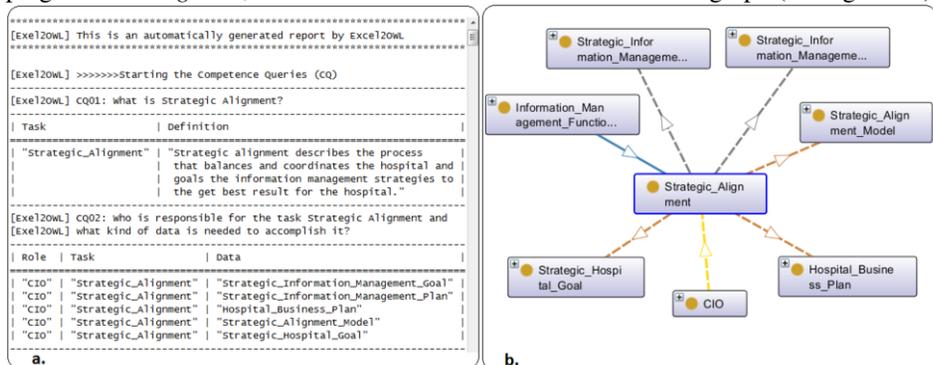


Figure 2. Result of the queries CQ01 and CQ02 (a) and its corresponding visualization using OntoGraf (b). The yellow arrow represents the responsibility relationship (see axiom 3) and the three brown dashed lines show the update relationships (see axiom 1). The usage relationships (see axiom 2) are represented by the grey dashed lines. The inheritance is visualized by a blue line

Typically, in textbooks, it takes many pages to describe what strategic alignment is and what kind of data is needed to accomplish it. A student could legitimately

encounter a question about this on an exam. The *OntoGraf* visualization and the multiple query answering results offer students an easy way of determining the correct response without spending time on searching through several abstract and complex chapters of IM books. Hence, our approach enables the use of the SNIK ontology for e-learning purposes, particularly in the master module “management of HIS”. In addition, essential CQs and exam questions about this module are answered automatically using the multiple query answering method. Subsequently, a report file about the query results is generated (see figure 2a). This document is well-structured and can be used for ontology evaluation and also for exam preparation.

4. Discussion and Outlook

Using *Excel2OWL*, we automated the construction of OWL ontologies from spreadsheet content. The other step – knowledge extraction – was done manually and was thus time-consuming and expensive. This situation motivates us to automate some parts of this process by using NLP techniques in order to reduce the development cost. In addition, the multiple query answering method showed promising results which could be extended to build an ontology-based e-learning module. In such a module, students' knowledge about SNIK could be tested automatically using multiple choice questions or annotation tasks. Their answers could be evaluated using logical reasoning.

Acknowledgement

This work is supported by the DFG (German Research Foundation) under the Project SNIK, Grant no. 1605/7-1 and 1387/8-1.

References

- [1] Biemann C., *Ontology Learning from Text: A Survey of Methods*, LDV-Forum 20(2) (2005), 75–93.
- [2] Geroimenko V., *Dictionary of XML Technologies and the Semantic Web*. Springer, London 2004.
- [3] Drumond, L., and Girardi, R. *A Survey of Ontology Learning Procedures*, *3rd Workshop on Ontologies and Their Applications* (2008), 1-12.
- [4] Vania Dimitrova et al., *Involving Domain Experts in Authoring OWL Ontologies*. 7th International Semantic Web Conference, *Lecture Notes in Computer Science* **5318** Springer (2008), 1-16.
- [5] Jupp S et al., *Populous: a tool for building OWL ontologies from templates*, *BMC Bioinformatics* **13** (2012), S5.
- [6] RDF Refine: a Google Refine extension for exporting RDF [<http://refine.deri.ie/>].
- [7] O'Connor MJ, Halaschek-Wiener C., Musen MA., *Mapping Master: A Flexible Approach for Mapping Spreadsheets to OWL*, Springer Volume **6497** (2010), 194–208.
- [8] Excel Import tool [http://protege.wiki.stanford.edu/wiki/Excel_Import].
- [9] Excel2RDF tool [<http://www.mindswap.org/~rreck/excel2rdf.shtml>].
- [10] Convert2RDF tool [<http://www.mindswap.org/~mhgrove/ConvertToRDF>].
- [11] Winter A. et al., *Health Information Systems: Architectures and Strategies*, Springer, Heidelberg 2011.
- [12] Ammenwerth et al., *IT-Projektmanagement in Krankenhaus und Gesundheitswesen*, 1. Auflage, Stuttgart: Schautauer GmbH 2005.
- [13] DuCharme B., *Learning SPARQL. Querying and Updating with SPARQL 1.1*, O'Reilly Media (2013).
- [14] Uschold M., Gruninger M. et al., *Ontologies: Principles, Methods and Applications*, *Knowledge engineering review* **11**(2) (1996), 93–136.