German Medical Data Sciences: Visions and Bridges
R. Röhrig et al. (Eds.)

doi:10.3233/978-1-61499-808-2-165

Expert2OWL: A Methodology for Pattern-Based Ontology Development

Kais TAHAR[a,1], Jie XU[a], Heinrich HERRE[a]
[a] *Institute of Medical Informatics, Statistics and Epidemiology, Medical Faculty, Leipzig University*

Abstract. The formalization of expert knowledge enables a broad spectrum of applications employing ontologies as underlying technology. These include eLearning, Semantic Web and expert systems. However, the manual construction of such ontologies is time-consuming and thus expensive. Moreover, experts are often unfamiliar with the syntax and semantics of formal ontology languages such as OWL and usually have no experience in developing formal ontologies. To overcome these barriers, we developed a new method and tool, called *Expert2OWL* that provides efficient features to support the construction of OWL ontologies using GFO (General Formal Ontology) as a top-level ontology. This method allows a close and effective collaboration between ontologists and domain experts. Essentially, this tool integrates Excel spreadsheets as part of a pattern-based ontology development and refinement process. *Expert2OWL* enables us to expedite the development process and modularize the resulting ontologies. We applied this method in the field of Chinese Herbal Medicine (CHM) and used *Expert2OWL* to automatically generate an accurate Chinese Herbology ontology (CHO). The expressivity of CHO was tested and evaluated using ontology query languages SPARQL and DL. CHO shows promising results and can generate answers to important scientific questions such as which Chinese herbal formulas contain which substances, which substances treat which diseases, and which ones are the most frequently used in CHM.

Keywords. Ontology development, Biomedical Ontologies, Knowledge Representation, Herbal Medicine, TCM, Expert2OWL

1. Introduction and related works

The formalization of expert knowledge about a domain requires a formal specification of the terms and relations the knowledge is built upon. In this paper, we apply the term ontology for knowledge systems represented in a formal language, such as OWL [1]. The formalization of expert knowledge enables a broad spectrum of applications, among them eLearning and expert systems [2]. In this work, we analyze various textual information about Traditional Chinese Medicine (TCM) from public resources such as TCM textbooks and online encyclopedias [3-4] as a source of domain knowledge as well as TCM expert experiences. Our aim is to propose a method for supporting the structural acquisition of knowledge in these resources and its formalization within OWL 2 DL [1]. The formalization of domain knowledge by OWL ontologies has the potential to offer to experts significant benefits in their domain. However, most domain

[1] Kais Tahar, IMISE, University of Leipzig, Härtelstr. 16-18, 04107 Leipzig, Germany;
E-mail: kais.tahar@imise.uni-leipzig.de

experts lack skills in using formal languages such as OWL or DL and usually have no experience in developing formal ontologies [2,5]. Domain experts therefore risk spending too much time in constructing ontologies that are in many cases impractical. To overcome these problems, experts need a domain core ontology (DCO) that defines and inter-relates the basic entities to formalize their domain knowledge in a useful manner. For that reason, domain experts should collaborate with ontologists to develop well-founded ontologies by using top-level ontologies such as GFO [6]. In addition, domain experts need easy access to authoring and refinement processes so that a close and effective collaboration with ontologists can be established.

We analyzed several existing tools mentioned in [2,5] such as *Mapping Master*, *Populous*, and *Excel2OWL* for their potential to build the CHO from our spreadsheets, but we concluded that none of these tools could be applied to our project. Due to the fact that these tools do not support pattern design with arbitrary restriction quantifiers and do not cover complex expressions such as intersection sets and equivalent classes axioms, we developed a new method and tool, called *Expert2OWL.*

Throughout this paper, we use standard semantic web terminology that refers to formal descriptions of ontological entities [1,7]. The rest of the paper is organized as follows: Section 2 explains *Expert2OWL* as an ontology engineering method, Section 3 presents the results obtained by applying this method in the field of CHM and finally, Section 4 concludes with a discussion and an outlook on future work.

2. Methods

Expert2OWL provides efficient features to support the collaborative development of OWL ontologies using GFO. Essentially, this tool integrates the Excel spreadsheet as part of a pattern-driven ontology development and refinement process. *Expert2OWL* enables an automatic transformation of spreadsheet content into OWL axioms. Axioms for a domain are logical expressions, written in a formal language such as first-order logic or DL [1], which are valid in the considered domain. Relational propositions are expressions that specify certain relationships between entities of a given domain. Relational propositions can be either true or false, while axioms on the contrary are certainty propositions that have the highest logical value in a domain and therefore represent domain tautologies [8]. For this reason, we use domain core axioms (DCAs) in our method as a foundation for the formalization and analysis of domain knowledge. Essentially, *Expert2OWL* implements a method that separates the full axiomatization from the immediate compilation of graspable propositions that experts rated as significant and, hence, are assumed to be true in the considered domain (see Figure 1).

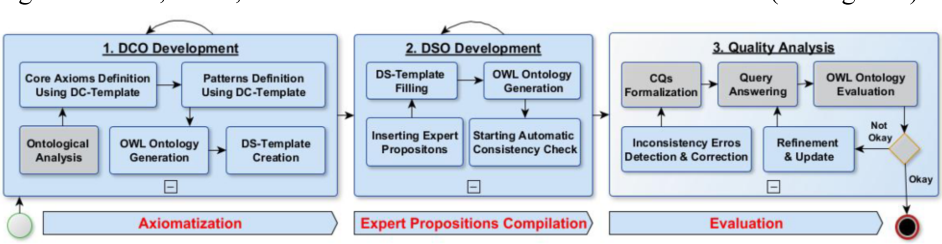

Figure 1. A pattern-based methodology for developing OWL ontologies using spreadsheets. The pale blue boxes show all tasks supported by the tool *Expert2OWL.*

These experts' propositions should be consistent with the defined DCAs. For example, in this work we collected propositions from TCM experts involved in the ontology development process. To automatically verify the consistency of these propositions, we implemented an inference engine that extends *Expert2OWL* with automated reasoning methods based on the previously declared DCAs. Our method is realized by a workflow that consists of three steps, called Axiomatization, Expert Propositions Compilation and Evaluation, which allow multiple iterations. This workflow is described using BPMN elements (see Figure 1). In the first step, ontologists develop a DCO based on interviews with domain experts by using a top-level ontology such as GFO. GFO provides a basic method for specifying the DCAs. The development of DCAs is the main and most creative task of the ontologist who collaborates with experts on this topic. We consulted TCM experts from the China Academy of Chinese Medical Sciences (CACMS) in Beijing, who provided answers to some important questions, such as the issues of compatibility of drugs and dose, the specification of meridians, and important information about clinical use cases including dosages of toxic herbal drugs. As a result, we developed an accurate and expressive DCO named GFO-CHO using *Expert2OWL*. GFO-CHO extends GFO with basic concepts, material objects, and symbolic structures illustrated in Table 1. The developed DCAs include axiom patterns, equivalent class axioms, disjoint class axioms, and embedding axioms. The Embedding axioms integrate the acquired knowledge in GFO (see Table 1).

Table 1. Embedding axioms for integrating the developed core ontology of Chinese Herbal Medicine in GFO.

GFO entities	GFO-CHO entities
gfo:concept	*gfo-cho:Chinese_herbal_fromula (CHF) , gfo-cho:TCM_therapeutic_use (TTU), gfo-cho:Health_problem (HP), gfo-cho:Disease, gfo-cho:Symptom, gfo cho:Drug_prescription (DP), gfo-cho:TCM_drug_prescription (TDP).*
gfo:Material_object	*gfo-cho:Budy_part (BP), gfo-cho:Meridian (M), gfo-cho:Plant (P), gfo-cho:Drug,, gfo-cho:Chinese_herbal_drug (CHD), gfo-cho:Substance (S).*
gfo: Symbol_structures	*gfo-cho:Code (C), gfo-cho:ICD10_code (IC).*

We define axiom patterns as a specific type of DCA that is more complex and contain free variables. These complex patterns can be reused for assisting the implementation of experts' propositions such as semantic relationships between concepts or individuals (see Table 2). To support the development of these design patterns, *Expert2OWL* provides a domain core Template (DC-Template). This template offers a practical formalism for the specification of DCAs including pattern axioms with arbitrary restriction quantifiers. Using this template content, *Expert2OWL* generates an OWL ontology automatically. The resulting DCO contains OWL axiom patterns that restrict the basic types and relations within this domain. Based on these axiom patterns, *Expert2OWL* also creates a partly restricted domain specific template (DS-Template) automatically. This DS-Template is restricted with a drop down list (DDL) of binary relationships that point out the corresponding design patterns (see Table 2).

The second step – Expert Propositions Compilation – was also supported by *Expert2OWL*. During this step, experts compiled domain specific knowledge in the form of relational propositions that represent simple statements of the form SPO (subject predicate object). Hence, the structure of the DS-Template consists of three columns: one column for binary relations (predicate) and two columns for the relations' arguments. The domain expert selects the corresponding relationship from the DDL and adds a relational proposition regarding two domain entities (see Table 2). The variable Symbols *?X* and *?Y* are replaced with the first and second arguments of the domain expert's proposition. These arguments can be added arbitrarily because the consistency of each added propositions can be verified automatically using reasoning on axiom patterns of the developed DCO. The benefit of this modeling method is that it allows domain experts to describe the semantic network of CHM with relational propositions that are close to natural language sentences. This provides domain experts with an easy way to access the development process of formal ontologies without having to resort to learning a complex ontology language such as OWL. As a result, *Expert2OWL* generates a domain specific ontology (DSO) from the DS-Template.

Once the development process is finished, the next step is quality analysis, where the performance and quality of the developed ontology are evaluated and rapidly improved by using *Expert2OWL*. The earlier versions of CHO contained many inconsistency errors. *Expert2OWL* provides an inference engine that detects inconsistent expert's propositions and corrects them automatically. To evaluate the resulting ontology, we implemented several competency questions (CQs) [2] using SPARQL and DL[1].

Table 2. The first column shows exemplary relationships of the DDL. The second column illustrate the corresponding axiom patterns expressed in Manchester OWL Syntax with argument variables ?X and ?Y.

Relationships	OWL axiom patterns expressed in Manchester OWL Syntax
hasIngredient	*Class: ?X SubClassOf: CHF AND hasIngredient some ?Y*
	Class: ?Y SubClassOf: CHD
hasPlantSource	*Class: ?X SubClassOf: CHB AND hasIngredient only ?Y Class: ?Y SubClassOf: P*
hasMeridian	*Class: ?X SubClassOf: CHB AND hasMeridan only ?Y Class: ?Y SubClassOf: M*
relatedToCode	*Class: ?X SubClassOf: gfo:Entity AND relatedToCode some ?Y*
	Class:?Y AND SubClassOf: ID
instanceOf	*Individual: ?X Tpyes: ?Y Class: ?Y*
equivalentTo	*Class: ?X EquivalentTo: ?Y Class: ?Y*
hasProcedure	*Class: ?X SubClassOf: CHF AND Annotations:hasProcedure "?Y"*

3. Results

The proposed methodology was implemented as a tool named *Expert2OWL* and applied in the field of CHM. We consulted TCM experts and analyzed public textual resources [3-4] to discover important common basic entities and properties of Chinese herbal formulas and drugs such as prime ingredient, procedure, indication, dosage, disease, TCM therapeutic use, meridian, chemical compositions, and toxicity. As a result, *Expert2OWL* generates a formal CHO. It contains 2046 classes, 14 object properties, 1808 individuals and 16847 axioms that restrict CHO entities. A total of 4103 of these axioms were inferred automatically after inconsistency detection and correction with *Expert2OWL*. The resulting CHO is composed of two ontologies. The DCO contains 14 object properties and 131 DCAs, which are specified and implemented using *Expert2OWL* to integrate CHO entities in GFO. The DSO contains all essential individuals that instantiate 139 CHF, 193 CHD, 187 P, 576 S, 340 IC and 405 HP. All HPs including diseases and symptoms were coded in ICD-10 to promote the internationalization of CHM, and to harmonize it with Western medicine.

In addition, the expressivity of CHO was tested using multiple CQs implemented in SPARQL and DL. CHO yielded promising results and generated answers to important scientific questions. One example showed that Liquoric Root (甘草, Gan Cao) is the most frequently used herbal drug. It contains the substances glycyrrhetinic acid, glycyrrhizin, licochalcone, licorice extract and liquiritin. This drug was used in 65 CHFs and treats several health problems such as palpilations, dyspnoea and cough.

4. Discussion and outlook

A major benefit of our methodology compared with existing ones mentioned in [2,5] is that it provides features that support the collaborative development of new design patterns and enables the integration of the acquired knowledge in GFO. Furthermore, with *Expert2OWL* we can also separate the resulting DCO from the DSO. This enables us to expedite the development process of OWL ontologies and modularize the resulting ontologies. Another main advantage of our method is that it provides a general common formalism that is largely independent of the application case since the architecture of *Expert2OWL* stipulates the separation of the full axiomatization from the immediate compilation of domain specific knowledge in the form of relational propositions (see figure 1). Consequently, this method enables experts to subsequently change the DCAs including the restriction quantifiers consistently and easily in the DC-Template without redefining the Java source code. Furthermore, it provides automated detection and correction of inconsistency errors. Our approach thus offers to experts an easy and efficient way to build domain ontologies without resorting to complex ontology editing tools or even OWL syntax. Hence, *Expert2OWL* facilitates the domain experts' involvement in a collaborative ontology development process and reduces the required development time and costs. *Expert2OWL* can also be transformed into a web-based tool for real time collaborative ontology development so that multiple domain experts could share their filled templates and work simultaneously on the same content but from different PCs to generate consolidated domain ontologies. The resulting ontologies can be integrated into a knowledge base for analyzing Big Data using GFO. The benefit of using this top level ontology is the semantic interoperability that allows harmonizing related domain knowledge such as chemistry, pharmacy, medicine and biology. The embedded reasoner can infer implicit knowledge and detect inconsistency errors automatically.

In addition, the query answering method can be extended to build an ontology-based software that supports drug discovery from herbal medicine. The application of such a tool could effectively support and expedite drug research.

Acknowledgement.

We thank Dr. Yan Zhu from CACMS in Nanxiaojie 16, Dongzhimennei, Beijing for his support in clarifying various important notions in TCM.

References

[1] Hitzler P., Krötzsch M., Rudolph S., Sure Y., Semantic Web. Grundlagen. Springer, Berlin, 2008.
[2] Tahar K, Schaaf M, Jahn F, Kücherer C, Paech B, Herre H, Winter A. An approach to support collaborative ontology construction. *Studies in health technology and informatics* **228** (2016), 369-373.
[3] Zhu Zhenheng, Ming Chenchong, *Danxi's Mastery of Medicine*, China Bookstore press, Beijing, 1986.
[4] http://www.wiki8.com (accessed Mars, 3 2017)
[5] Jupp S et al., Populous: a tool for building OWL ontologies from templates, *BMC Bioinformatics* ***13*** (2012), S5
[6] Herre H: General Formal Ontology (GFO): A Foundational Ontology for Conceptual Modelling. In *Theory and Applications of Ontology: Computer Applications*, Springer, Netherlands; 2010, 297–345.
[7] Geroimenko V., *Dictionary of XML Technologies and the Semantic Web*, Springer, London, 2004
[8] Mendelson, E. *Introduction to Mathematical logic*, Van Nostrand, Princeton, 1964.