Ontology templates: design pattern macros for large scale ontology development

Main Objective and Summary of the Project

The main objective of this project is to extend the state of the art in ontology engineering with advanced techniques for the definition, manipulation, maintenance and automatic extraction of ontology design patterns and ontology templates for the purpose of large scale ontology development. These techniques will play a key role in the creation of a ontologies in ongoing projects related to the SIRIUS\(^1\) centre (Norwegian Centre for Scalable Data Access) and the BIGMED project (the BIG data MEDical solution), to enable, among others, the implementation of scalable data access and big data analytics. The project will be executed in close collaboration with ontology experts and domain matter experts from industry and academia.

Project Background and Scientific Basis

A major obstacle in the industry’s digitalisation efforts of using data in smarter ways is the fact that the semantics of the data and the systems that process the data are not represented in formats that are accessible for automatic computation. Rather, the semantics are buried in systems code and usually informally represented in systems and standards documentation.

Ontologies represent domain models with clear semantics defined by formal logic. Ontologies and automatic reasoning methods enable the processing and access of the data on the semantical level, resulting in powerful techniques for using and validating the domain model, as well as for the integration of different, and partially overlapping, vocabularies and datasets.

However, developing and maintaining industrial scale ontologies is difficult and very costly. It requires the rare combination of deep knowledge in the relevant domain and in the theory that underpins ontologies and information modelling. Tools and methodologies that secure the quality of the ontology and allow ontology experts and domain experts to work together efficiently are necessary.

As a case in point, the engineering, procurement, and construction (EPC) company Aibel\(^2\) has developed a comprehensive ontology that describes various characteristics of the piping and structure domains [1]. This information is spread over several hundred of industry standards commonly published as PDF documents. Using its ontology, Aibel is now able to automatically consider and meet all the requirements of the standards when deciding on what bulk material to select across the complexity of major capital project developments. This is a task that previously could only be performed by engineers carefully comparing design specifications against standard requirements. To build the ontology, engineers at Aibel have used spreadsheets, formatted by ontology experts, to represent the information in standards. These spreadsheets are then fed into an elaborate system of transformations that outputs the ontology. Although the system succeeds in producing a high value ontology, there are severe drawbacks. The logic in the transformations is not available for consistency checks, and the task of manually creating and managing spreadsheet templates does not scale well, making it difficult to re-apply the system and methodology to other domains. Aibel sees that both further development and maintenance of its ontology would benefit greatly from more robust, semantically aware methods.

---

\(^1\) [http://sirius-labs.no/](http://sirius-labs.no/)

Uniform modelling is required to secure consistency of modelling across the engineering domains, and continuous maintenance is required to keep the ontology up to date with revised standards and changing industry requirements. Unfortunately, there are no tools or frameworks available today that take on the fundamental challenge of semantic integrity at the scale required.

Ontology design patterns\(^3\) (ODP) [2] are, like software design patterns, general reusable solutions to a commonly occurring ontology design problem. ODPs have become popular in ontology engineering as a means to construct ontologies using a set of structurally similar building blocks. Ontology templates (OT) [3] are parameterised ontologies that work just like macros (cf. [4]) in many computer languages. They define compact representations formats of ontological facts that may be expanded to a “full” ontological structure. Furthermore, OTs may be used to implement, use, and share ODPs, using existing semantic web standards. It is reasonable to believe that OTs could be used to both capture the data in Aibel’s spreadsheets and to fully represent the logic behind the transformations. This would allow the consistency of the transformations in Aibel’s system to be checked and the internal structure of the spreadsheets to be improved, e.g., identifying inconsistencies, overlaps and duplicates in the spreadsheet templates.

Although different implementations of template mechanisms exist, e.g., [5, 6, 7], these have not seen widespread uptake. The project aims at extending the state of the art with advanced techniques to structure and quality assess ontology templates and their use, allowing design patterns to be applied, defined and structured declaratively, using existing semantic web languages and infrastructure. The techniques will be directly applicable to ongoing ontology engineering projects fostered by the SIRIUS centre.

**Research Questions and Scientific Challenges**

There are many open challenges, and the project can be adjusted to fit relevant candidates:

- Define a suitable language and tools for representing and using templates and relations between templates
- Can templates be used as an ontology alignment mechanism (cf. [8]), mapping between different ontology languages on a ontology pattern level?
- Can templates be used to map the same domain knowledge into different styles of ontology modelling for different purposes?
- Can templates facilitate the use of logic-based techniques to check the inconsistencies?
- Create tools for discovering and extracting templates from ontologies
- Can we apply the designed techniques to other domains like biomedicine?
- Implement template mechanisms with semantic web technologies
- Implement tools that make is easy for domain experts to use templates
- Implement tools that make it easy for ontology experts to define and structure templates

**Ethics**

This project raises no ethical concerns.

**Project Timeline**

Tentative timeline according to above challenges:

---

\(^3\) [http://ontologydesignpatterns.org](http://ontologydesignpatterns.org)
Y1: Get a good overview of current research in Ontology Engineering and Ontology Design Patterns. First prototype to represent templates and their relationships.

Y2: Inclusion of an ontology alignment mechanism based on templates, and logic-based techniques to identify and correct inconsistencies in templates.

Y3: Implementation of the final toolkit for advanced ontology engineering based on ontology templates. Evaluation of the implemented techniques in a controlled scenario.

**Applicant Background and Competence**

The candidate must have competitive grades in subjects related to logic, artificial intelligence and databases. It is desirable, but not essential, with practical experience in applications of knowledge representation and semantic web languages like OWL, SPARQL and RDF. Good programming skills is also a plus.

**Supervision**

The PhD student will be affiliated with the Logic and Intelligent Data (LogID) group, based in the Department of Informatics (The Faculty of Mathematics and Natural Sciences). The research conducted in the group is based on well-established methods from logic, which we extend and enhance to tackle tomorrow’s challenges in fields like Semantic Web and Big Data.

The LogID group participates in SIRIUS, a Norwegian centre of excellence for research-based innovation, and the BIGMED project (BIG data MEDical solution), which consists of a multidisciplinary team of ICT researchers, biomedical researchers and industrial partners. SIRIUS is funded by the Norwegian Research Council, UiO, NTNU, University of Oxford and 11 partners from industry including IBM. SIRIUS aims at delivering world leading methods and technologies in data access through innovative research in, among other things, semantic and language technologies.

The thesis will be jointly supervised by Dr. Martin G. Skjæveland, Dr. Ernesto Jimenez-Ruiz, and Dr. Johan W. Klüwer.

**References**


