Intelligent formal analysis of heterogeneous data for semantic web

Mariano Julián Ferreirone[1,2,3], Mario Lezoche[1,2], Diego Torres[3,4], and Hervé Panetto[1,2]

[1] Univesité de Lorraine, Nancy, France

[2] CRAN, Nancy, France

[3] LIFIA, CICPBA-Facultad de Informática, UNLP

[4] Depto. CyT, UNQ, Bernal, Argentina

Abstract. The semantic web is having a main role in the development of many sectors of the industry, and is becoming necessary for being competitive in the market. The tourism in France is not an exception, and as part of their reorganization, the knowledge formalization modelling is required. This work is expected to improve the interoperability between the information represented on different knowledge formalization schema levels.

Keywords. Knowledge formalization, Semantic Interoperability, etourism.

1 Introduction

In the ”Grand Est” region, a new regional tourism policy has emerged with the elaboration of a Regional Tourism Development Plan (RTDP). The project aims to develop a digital tool to pave the way for a global, intelligent and connected system allowing the collection and processing of data in real time, in order to extract knowledge for tourism purposes. This data and information will then be aggregated and transformed by automatic inference systems to formalize the existing implicit knowledge. This formalization work will enable the use of data and information to provide more accurate and real-time services to all socio-economic actors. For the organization, a systems engineering approach [7] which consists in relying on different types and levels of abstraction or models will be used.

These models should express and formalize not only the ”structural” aspect of the system components, but also their behaviour [5], which may be limited by the specific requirements of the system domain (business rules) and the interoperability protocol(s), which may impose strict rules to provide inter-operable systems with properties such as autonomy, confidentiality and transparency[6].

The objective of this project is to model data from heterogeneous sources and to study the problems posed by model-driven engineering cooperative systems.

More specifically, the effort is going to be focused on the lack of formalization in the inter-operation of systems by minimizing semantic losses. Two popular formalization methods are going to be studied from the point of view of their interoperability: ontology and knowledge graphs. Since they can be found with different levels of schema, for example RDF (Resource Description Framework) or OWL (Web Ontology Language), it's interesting for the community to develop a way of connecting them minimizing the loss of semantics.

2 State of the art

Some interesting works related to combining knowledge formalizations with different schema levels or to increase the schema level on existent representations are as follows. In the tool named CoGui [3], the transformation is done from RDF to conceptual graphs, and the resulting OWL file is exported to different languages. When referring to RDF, currently, OWL rules, constraints and type disjunctions are ignored. Another interesting concept is the ontology alignment or ontology matching, which is based on generating a set of correspondences between concepts, properties or instances of different structured KGs, with the objective of unifying them into a new one [2]. Furthermore, Shapes Constraint Language [1] is a language for validating RDF graphs against a set of conditions, and adding formalization to RDF knowledge graphs. The characteristics of the property graphs [8] represent also an interesting model to take into account since they add some descriptive properties to the graph.

3 Problem Statement and Contributions

The contribution of this PhD project should include: a state of art research based on a systematic literature review methodology, the development of the semantic interoperability method, between heterogeneous data sources, which could be represented as a program or a set of rules that will be able to combine existent knowledge formalization models of different schema level.

4 Research Methodology and Approach

The research methodology is a process composed of the following steps: 1) Survey based on the state of art related to the combination between Knowledge formalization models of different schema level, guided by a defined literature review methodology named Scoping Literature Review (SLR) [4]. 2) Development of an approach to combine existent knowledge formalization models of different schema level. 3) Testing of the new approach using real data from an example of the corresponding knowledge representation. 4) Evaluation of the results, analyzing the equivalence of the semantics. 5) The conclusion of the advancements done in the last iteration, look for opportunities to define following challenge, write a scientific article

if corresponds. 6) Generalization of the created novel approach. 7) Develop prototype of the program. The process will be iterative from the step 2 to the step 5.

5 Evaluation Plan

The evaluation plan consists in doing several tests with different knowledge graphs (or ontologies) as inputs, measuring the semantic similarity between the input and output knowledge formalization model examples that are going to be part of the process. This will be done in order to be aware if there's a loss of semantics in the process. Furthermore, the time of execution is going to be noted and analyzed as well.

6 Conclusions and Learned Lessons

For the moment, the learned lessons are based on understanding how to find and read scientific articles, how to structure a review with a systematic methodology and concepts related to the semantic web and knowledge representation.

7 Ph.D. Stage

Early.

References

1. Shapes constraint language (SHACL). Tech. rep., W3C (Jul 2017), https://www.w3.org/TR/shacl/
2. Ardjani, F., Bouchiha, D., Malki, M.: Ontology-alignment techniques: Survey and analysis. International Journal of Modern Education and Computer Science 7, 67–78 (11 2015). https://doi.org/10.5815/ijmecs.2015.11.08
3. Baget, J.F., Chein, M., Croitoru, M., Fortin, J., Genest, D., Gutierrez, A., Leclére, M., Mugnier, M.L., Salvat, E.: RDF to Conceptual Graphs Translations. In: CS-TIW: Conceptual Structures Tool Interoperability Workshop. vol. LNAI, p. 17. Springer, Moscow, Russia (Jul 2009), https://hal-lirmm.ccsd.cnrs.fr/lirmm-00410621
4. Franciosi, C., Voisin, A., Miranda, S., Riemma, S., Iung, B.: Measuring maintenance impacts on sustainability of manufacturing industries: from a systematic literature review to a framework proposal. Journal of Cleaner Production 260, 121065 (2020). https://doi.org/10.1016/j.jclepro.2020.121065
5. Maier, M.W.: Architecting principles for systems-of-systems. Systems Engineering 1, 267–284 (1998).
6. Panetto, H., Zdravkovic, M., Jardim-Goncalves, R., Romero, D., Cecil, J., Mezgár, I.: New perspectives for the future interoperable enterprise systems. Computers in Industry 79, 47–63 (2016)
7. Ramos, A.L., Ferreira, J.V., Barcel´o, J.: Model-based systems engineering: An emerging approach for modern systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42(1), 101–111 (2012). https://doi.org/10.1109/TSMCC.2011.2106495

8. Tomaszuk, D., Angles, R., Thakkar, H.: Pgo: Describing property graphs in rdf. IEEE Access 8, 118355–118369 (2020).