Ontology Learning for Facilitating Ontology Matching in Automotive

Ondrej Šebek[1], Václav Jirkovský[1], Nestor Rychtyckyj[2], Petr Kadera[1]

[1]Czech Institute of Robotics, Informatics, and Cybernetics
Czech Technical University in Prague,
Zikova 4, 166 36 Prague, Czech Republic
[2]Ford Motor Company
Dearborn, MI, USA

{ondrej.sebek,vaclav.jirkovsky,petr.kadera}@ciirc.cvut.cz
nrychtyc@ford.com

Abstract. All manufacturing companies need to monitor a large number of devices and from which critical data must be captured and analyzed. The increasing complexity of these ecosystems emphasizes the requirement for a flexible and versatile data model architecture. Ontologies may facilitate a proper understanding of the problem domain as well as the interoperability with surrounding systems using ontology matching approach. However, data models of surrounding systems are not always ontologies. Thus, concepts and relations among them have to be extracted from the models to enable their integration with the ontology. The definition of concepts, their hierarchy, relations between concepts, and properties from a general architecture is a complex task and has to be tailored to an application's needs. In this paper, we propose an involvement of the ontology learning approach to the process of ontology matching in the automotive.

Keywords: Ontology, Heterogeneity, Ontology Learning, Ontology Matching,

1 Introduction

All manufacturing companies need to monitor a large number of devices from which critical data must be captured and analyzed. The increasing complexity of these ecosystems emphasizes the requirement for a flexible and versatile data model architecture. Furthermore, common data models are becoming insufficient for conducting analytical tasks due to the systems complexity and corresponding exacting integration of new devices due to a complicated understanding of system data model.

Thus, the essential requirement is the proper understanding of given data models (sensors, machines, etc.) for ensuring a faultless processing of a huge amount of data. Moreover, the solution should also allow for easy maintainability as there will be frequent additions and modifications to the data model. The mentioned semantic integration problem [1], as well as the problem of easy maintainability, may be solved by

J. Steinberger, M. Zíma, D. Fiala, M. Dostal, M. Nykl (eds.)
Data a znalosti 2017, Plzeň, 5. - 6. října 2017, pp. 31-34.

the employment of Semantic Web technologies and model description in ontologies [2]. This approach also supports knowledge expressiveness and reasoning as well as the ability to keep track of the source of each data item.

The Semantic Web technologies were originally intended for the representation of various highly heterogeneous data models (e.g. web page metadata) within the Internet. Thus, these technologies were developed mainly for facilitating an integration of various data models. This research branch is named ontology matching and mapping [3] and many promising methods have already been developed. On the other hand, many of these methods were proposed as fully-automatic approaches – mainly for the integration of large ontologies. This approach is not suitable for all domains because it is required to achieve the highest possible precision in domains such as automotive, medicine, etc. In our previous work, MAPSOM [4] framework was developed for semi-automatic ontology matching based on Kohonen's self-organizing map[1] and active learning[2].

In this paper, we introduce a utilization of MAPSOM framework for an integration of heterogeneous data models in automotive. Within the framework of this work, we utilized our previous experiences with the development of manufacturing ontologies and will be building upon those ontologies in this work [5][6]. We provide a short overview of the problem of matching MS Excel sheet containing data about Ford spare parts to Ford Supply Chain ontology. The main obstacle in such matching may be found mainly in curtness and hard meaning understanding of data descriptions in catalogs.

If we would like to integrate data sources which are unstructured (e.g. text files, etc.), then we may exploit ontology learning [7] algorithms to derive new relations, concepts, and relations among entities. On the other hand, this approach has only limited applications in the case of structured data sources such as catalogs – there are no explicitly defined relations among entities, and there are typically no suitable additional text data for a widespread utilization of ontology learning algorithms. Nevertheless, we briefly describe conducted experiments which demonstrate that even limited utilization of ontology learning approaches may improve previously mentioned ontology matching task.

2 Ontology Learning for Facilitating Ontology Matching

The goal of ontology matching is to find correspondent entities expressed in different ontologies. The subsequent goal is an enrichment of captured knowledge in the first ontology by knowledge from the second one and vice versa. In this paper, we use a hybrid matching system prototype [8] which is responsible for matching elements from an MS Excel file (representing Ford spare parts) to the Ford Supply Chain ontology.

The Ford supply chain ontology captures the risk managements in the Ford global supply chain – every car model depends on many different suppliers, and important

[1] https://en.wikipedia.org/wiki/Self-organizing_map

[2] https://en.wikipedia.org/wiki/Active_learning_(machine_learning)

capability is to be able to determine which vehicles at which plant would be impacted by a potential shortage. The XLS file contains Ford spare part records and has about 62 various columns identifying particular parts. A predominant number of columns contain specific numerical codes or strings composed of abbreviated labels. A manual integration of such a data would be very exacting because of the big volume of records and their attributes.

The interesting tasks such as data preprocessing, searching acronyms in external vocabularies, and a process of elements matching are not mentioned in this paper due to the scope limitations, but it is presented in [8].

The outcome of common matching process results in the situation when some of the spare part records are mapped to individuals of the concept “BPNO” – base part number object (e.g. BPNO6C358) or to the concept “Part”. Unfortunately, many correspondent individuals have no additional properties (i.e., data and object properties) in the case of “BPNO”. In this situation, we have only limited understanding of a meaning of records gained after matching. Similarly, records mapped to the concept “Part” are mapped only to this concept and any other specialized sub-concepts are missing for some records. For example, there are no specialized sub-concepts like “break” or “crankshaft”, etc.

Thus, we need to derive new specialized sub-concepts or more detailed relationships between concepts for example with the help of ontology learning methods. These methods are usually applied to, for example, text documents (manuals) containing required information. In this task, the question is what information should be used for ontology learning task – part number, part description, and what else?

As the first approach, the base part number categories may be used for deriving new concepts and their relations. In our experiments, we used WordNet dictionary where hyponyms/hypernyms and holonyms/meronyms information may be found.

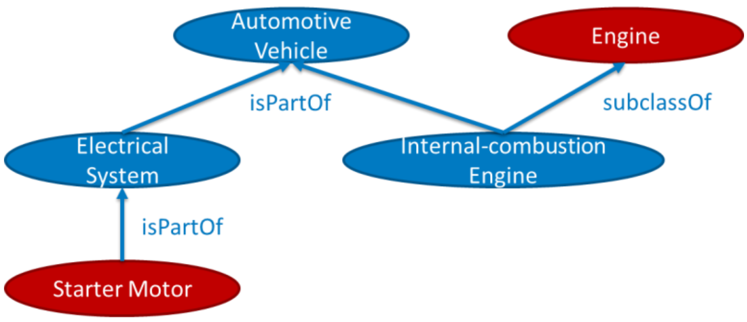

Figure 1. Deriving relationship between the concept "Starter Motor" and the concept "Engine"

Fig. 1 illustrates how the relationship may be found between “Starter Motor” and “Engine”. There is not only the new relationship but also new derived concepts such as “Electrical System”, “Automotive Vehicle”, and “Internal-combustion Engine”.

The main disadvantage is that there should be user verification (there are many various candidate relationships) as well as the search space is very extensive in some cases.

3 Conclusions

In this paper, we have shortly introduced how new concepts, as well as relationships between them, may be discovered even with very limited sources.

The future work will reside in deriving a more detailed/complex concept structure and relations with the help of external sources (text files – manuals, etc.) and part description in the form of abbreviated terms from the spare parts Excel sheet.

Acknowledgment: This work is supported through the Ford Motor Company University Research Proposal (URP) program and by institutional resources for research by the Czech Technical University in Prague, Czech Republic.

References

1. A. Doan and A. Y. Halevy, "Semantic integration research in the database community: A brief survey," *AI magazine*, vol. 26, no. 1, p. 83, 2005.
2. M. Obitko and V. Mařík, "Integrating transportation ontologies using semantic web languages," in *International Conference on Industrial Applications of Holonic and Multi-Agent Systems*, Springer Berlin Heidelberg, 2005.
3. J. Euzenat and P. Shvaiko, Ontology matching, Heidelberg: Springer, 2007.
4. V. Jirkovský and R. Ichise, "Mapsom: User involvement in ontology matching," in *Joint International Semantic Technology Conference*, Seoul, 2013.
5. D. Ostrowski, N. Rychtyckyj, P. MacNeille and M. Kim, "Integration of big data using semantic web technologies," in *IEEE Tenth International Conference on Semantic Computing (ICSC)*, 2016, 2016.
6. N. Rychtyckyj, V. Raman, B. Sankaranarayanan, P. S. Kumar and D. Khemani, "Ontology Reengineering: A Case Study from the Automotive Industry," *AI Magazine*, vol. 38, no. 1, 2017.
7. A. Maedche and S. Staab, "Ontology learning," in *Handbook on ontologies*, Springer Berlin Heidelberg, 2004, pp. 173-190.
8. V. Jirkovský, P. Kadera and N. Rychtyckyj, "Semi-Automatic Ontology Matching Approach," in *HoloMAS 2017*, Lyon, (forthcoming 2017).