Semi-automatic Ontology Integration Framework

Lihua Zhao and Ryutaro Ichise

National Institute of Informatics, Tokyo, Japan,
{lihua, ichise}@nii.ac.jp

Abstract. A semi-automatic ontology integration framework is introduced in this paper which can discover related ontology schema by analyzing SameAs graphs and retrieve important and frequent properties from core classes using a machine learning method. The framework can construct a high-quality integrated ontology from the linked data sets.

Keywords: linked data, semantic web, ontology integration

1 Introduction

With the growth of the Linked Open Data (LOD) cloud, many Semantic Web applications have been developed by accessing the linked data sets [1]. However, in order to use the data sets, we have to understand the heterogeneous ontologies of the data sets in advance. The ontology heterogeneity problem has become a popular issue, which can be solved by constructing a global ontology that integrates ontologies from various data sets.

In this paper, we propose a framework that can semi-automatically integrate heterogeneous ontologies by merging the ontologies created from two frameworks. The first framework can retrieve related properties and classes among different ontologies by applying ontology matching methods on SameAs graph patterns. The second framework can find important and frequent properties from core classes of each data set. By combining these two framework, we can construct a concrete global ontology for various linked data sets.

2 Semi-automatic Ontology Integration Framework

Constructing a global ontology by integrating heterogeneous ontologies of linked data can effectively integrate various data resources. The ontology integration framework that can solve the ontology heterogeneity problem is shown in Figure 1, which consists of two frameworks that generate ontologies from liked data sets and an ontology merger.

The first framework semi-automatically finds related properties and classes by analyzing SameAs graph patterns in the linked data sets. It consists of the SameAs graph pattern extractor, <Predicate, Object> (PO) pairs collector, ontology matching system, and ontology aggregator. This framework can create a high-quality integrated ontology with minor manual revision [3].

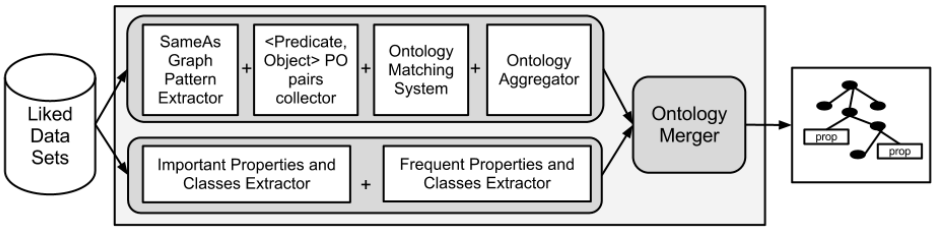

Fig. 1: Framework for Ontology Integration

However, the first framework only integrates related properties and classes between different ontologies. Therefore, we need another framework to find important properties and classes that frequently appear in the linked data sets. The second framework consists of two functions, where one is to find important properties from core classes, and the other is to find frequently used properties and classes. By applying machine learning methods, we can find important properties that are used to describe instances of a specific class. The weight of each property in an instance can be calculated as the product of property frequency (PF) and the inverse instance frequency (IIF) in a similar way as the TF-IDF [2]. The PF is the frequency of a property in an instance and the IIF is the logarithm of the ratio between the number of instances in a data set and the number of instances that contain the property. The frequent properties and classes can be found by analyzing the distribution of the usage in the instances.

With the ontology merger, we can merge two ontologies created with the above two frameworks, and construct a global integrated ontology that can help us easily access various data sets.

3 Conclusion

In this paper, we proposed a semi-automatic ontology integration framework that can integrate heterogeneous ontologies. The integrated ontology consists of important and frequent properties and classes that can help Semantic Web application developers easily find related instances and query on various data sets. With the ontology, we can also detect misuses of ontologies in the data sets and can recommend important properties for describing instances.

References

1. Bizer, C., Heath, T., Berners-Lee, T.: Linked data - the story so far. International Journal on Semantic Web and Information Systems 5(3), 1–22 (2009)
2. Manning, C.D., Raghavan, P., Schtze, H.: Introduction to Information Retrieval. Cambridge University Press (2008)
3. Zhao, L., Ichise, R.: Graph-based ontology analysis in the linked open data. In: Proceedings of the Eighth International Conference on Semantic Systems. pp. 56–63 (2012)