
Ontology Integration for the Linked Open Data

Author:

Lihua ZHAO

Supervisor:

Ryutaro ICHISE

DOCTOR OF PHILOSOPHY

Department of Informatics

School of Multidisciplinary Sciences

The Graduate University for Advanced Studies

2013

A dissertation submitted to the Department of Informatics,
School of Multidisciplinary Sciences,
The Graduate University for Advanced Studies (SOKENDAI)
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Advisory Committee

Assoc. Prof.	Ryutaro Ichise	National Institute of Informatics, SOKENDAI
Prof.	Hideaki Takeda	National Institute of Informatics, SOKENDAI
Assoc. Prof.	Ikki Ohmukai	National Institute of Informatics, SOKENDAI
Prof.	Akiko Aizawa	National Institute of Informatics
Assoc. Prof.	Kouji Kozaki	Osaka University

Abstract

The Linked Open Data (LOD) cloud contains tremendous amounts of interlinked instances, from where we can retrieve abundant knowledge. In order to access to the linked data, we have to be familiar with the ontology of each data set. However, because of the heterogeneous and big ontologies, it is time consuming to learn all the ontologies manually and it is difficult to observe which properties are important for describing instances of a specific class. In order to construct an ontology that can help users easily access to various data sets, we propose the **F**ramework for **I**n**T**egrating **O**Ntologies (**FITON**) that can reduce the heterogeneity of the ontologies, retrieve core ontology schemas, and construct easily understandable integrated ontology.

FITON solves three main problems: ontology heterogeneity problem, difficulty in identifying core ontology schemas, and missing domain or range information problem. The three main components of FITON solve each problem, which are graph-based ontology integration, machine-learning-based approach, and integrated ontology constructor. The graph-based ontology integration approach solves the ontology heterogeneity problem by analyzing the graph patterns of the interlinked instances and integrates heterogeneous ontologies by retrieving related classes and properties that are critical to link the same instances in different data sets. The machine-learning-based approach retrieves core ontology schemas (top-level classes and frequent core properties) by applying Decision Table and Apriori, that can help Semantic Web application developers easily understand the ontology schemas of the data sets. Furthermore, the integrated ontology constructor automatically adds missing domain, range, and annotations that can provide us rich information about the ontology. The integrated ontology constructed by FITON can help us discover missing links, detect misused properties, recommend standard ontology schemas for the instances, and improve the information retrieval with simple SPARQL queries.

To my family and friends.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Ryutaro Ichise for the continuous support of my Ph.D study and research, for his patience, motivation, and immense knowledge. Without his guidance and persistent help this thesis would not have been possible. I appreciate all his contributions of time, ideas, and supervision to make my Ph.D. experience productive and stimulating.

I would like to thank my committee members, Prof. Hideaki Takeda, Prof. Akiko Aizawa, Prof. Ikki Ohmukai, and Prof. Kouji Kozaki, for their encouraging, insightful, and constructive comments. In addition, thank Prof. Keizo Oyama, Prof. Atsuhiko Takasu, Prof. Nigel Collier, and Prof. Ken Satoh, who provided valuable and thoughtful feedback.

I appreciate my lab members, Lankesh, Mizanur, Raul, and all the internship students, who supported me and helped me in my research. This thesis would not be possible without the support and encouragement of my friends: Osamnia, Raghvendra, Hamza, Mohamed, Yi-Ju, Dan, Simo, Xuping, Yunlong, Zhi, Yanlin, Yu, Lei, Ruijian, Tao, Jeffrey, Vanessa, Khanh, Hua, Lihong, Masakazu, and all my friends from NII and SOKENDAI.

Most importantly, none of this would have been possible without the love and patience of my family. I would like to express my heart-felt gratitude to my parents and sister, who have been a constant source of love, concern, and support all these years.

Lastly, I thank the educational and financial support from National Institute of Informatics and the Graduate University for Advanced Studies (SOKENDAI) during my doctoral studies. I appreciate all the incredible opportunities and memorable experience they offered during my Ph.D. study.