Ontology Design Patterns for Large-Scale Data Interchange and Discovery

Pascal Hitzler

Data Semantics (DaSe) Laboratory, Wright State University, USA
pascal.hitzler@wright.edu
http://www.pascal-hitzler.de

Abstract. Data and information integration remains a major challenge for our modern information-driven society whereby people and organizations often have to deal with large data volumes coming from semantically heterogeneous sources featuring significant variety between them. In this context, data integration aims to provide a unified view over data residing at different sources through a global schema, which can be formalized as an ontology. From the end-users perspective, the data integration problem can be seen as a data access problem whereby the emphasis is on how such a unified view should help the nontechnical end-users in accessing the data from such heterogeneous sources. Early efforts to solve these problems led to a number of relational database integration approaches which have been very useful in specific situations. Unfortunately, they still require very significant manual efforts in creating and maintaining the mappings between the global and local schema, as the resulting integrations are often rigid and not transferable to new application scenarios without investing even more human expert resources, and furthermore, the global schema expressivity is limited which makes it difficult for the end-users to pose ad-hoc queries for their information needs.

Ontology design patterns have been conceived as modular and reusable building blocks for ontology modeling. We argue that a principled use of ontology design patterns also improve large-scale data integration under heterogeneity, as compared to the use of a monolithic ontology as global schema. In particular, the adoption of ontology design patterns can simplify several key aspects of the ontology application life cycle, including knowledge acquisition from experts, collaborative modeling and updates, incorporation of different perspectives, data-model alignment, and social barriers to adoption.

We report on recent progress we have made with this approach as part of our work on improving data discovery in the Earth Sciences, and point out key challenges on the road ahead.

Acknowledgments. This work was supported by the National Science Foundation awards 1017225 III: Small: TROn – Tractable Reasoning with Ontologies, 1354778 EAGER: Collaborative Research: EarthCube Building Blocks, Leveraging Semantics and Linked Data for Geoscience Data Sharing and Discovery

(OceanLink), and 1440202 EarthCube Building Blocks: Collaborative Research: GeoLink – Leveraging Semantics and Linked Data for Data Sharing and Discovery in the Geosciences. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

1. Cheatham, M., Hitzler, P.: String similarity metrics for ontology alignment. In: Alani, H., Kagal, L., Fokoue, A., Groth, P., Biemann, C., Parreira, J.X., Aroyo, L., Noy, N., Welty, C., Janowicz, K. (eds.) ISWC 2013, Part II. LNCS, vol. 8219, pp. 294–309. Springer, Heidelberg (2013)
2. Cheatham, M., Hitzler, P.: The properties of property alignment. In: Proceedings OM-2014, The Ninth International Workshop on Ontology Matching, at the 13th International Semantic Web Conference, ISWC 2014, Riva del Garda, Trentino, Italy (to appear, October 2014)
3. Hitzler, P., Janowicz, K.: Linked Data, Big Data, and the 4th Paradigm. Semantic Web 4(3), 233–235 (2013)
4. Hitzler, P., Krötzsch, M., Rudolph, S.: Foundations of Semantic Web Technologies. Chapman & Hall/CRC (2010)
5. Jain, P., Hitzler, P., Sheth, A.P., Verma, K., Yeh, P.Z.: Ontology alignment for linked open data. In: Patel-Schneider, P.F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J.Z., Horrocks, I., Glimm, B. (eds.) ISWC 2010, Part I. LNCS, vol. 6496, pp. 402–417. Springer, Heidelberg (2010)
6. Jain, P., Hitzler, P., Yeh, P.Z., Verma, K., Sheth, A.P.: Linked Data is Merely More Data. In: Brickley, D., Chaudhri, V.K., Halpin, H., McGuinness, D. (eds.) Linked Data Meets Artificial Intelligence, pp. 82–86. AAAI Press, Menlo Park (2010)
7. Janowicz, K., van Harmelen, F., Hendler, J.A., Hitzler, P.: Why the data train needs semantic rails. AI Magazine (to appear, 2014)
8. Janowicz, K., Hitzler, P.: The Digital Earth as knowledge engine. Semantic Web 3(3), 213–221 (2012)
9. Krisnadhi, A., Arko, R., Carbotte, S., Chandler, C., Cheatham, M., Finin, T., Hitzler, P., Janowicz, K., Narock, T., Raymond, L., Shepherd, A., Wiebe, P.: An ontology pattern for oceanograhic cruises: Towards an oceanograhper's dream of integrated knowledge discovery. Tech. Rep. 2014.1, OceanLink Technical Report (2014), http://pascal-hitzler.de/