The Next Generation of Ontology Development – Usage-based Ontology Engineering

Markus Luczak-Rösch (luczak@inf.fu-berlin.de), Freie Universität Berlin, Networked Information Systems Workgroup

Nowadays we look back on round about twenty years of research in the ontology engineering discipline. More than two decades ago ontologies were introduced to the ICT landscape. Since that date **three generations of ontology engineering** were passed through: **(1) Ontology engineering from scratch as the art of experts, (2) Ontology engineering inspired by software engineering, (3) Collaborative ontology engineering.** Ontologies can be used to enhance search, to add structured metadata to information, to infer knowledge, or to share a common understanding of a domain of discourse. Their development is well researched and standard ontology languages exist.

The Web is commonly understood as one global information space, characterized by its openness and driven by the easy principles invented in the 1980s: HTTP, URLs, HTML, hyperlinks. Due to that character the breakthrough of heavyweight semantic applications on the Web or of Web-scale reasoning efforts has not happened yet. The most successful application area of ontologies on the Web is the semantically rather lightweight approach of Linked Data. Relying on another simple set of principles (HTTP, URIs, the RDF standard, RDF links) this initiative gained momentum outside of the research community, at the latest when the public sector, online retailers and the media started to publish Linked Data. Reverse, this data is also at the beginning of being used. One can observe that the vocabularies underlying this data are rather small, simply structured and do not exploit the whole expressivity of ontology languages. To put this assumption on a basis we ran a survey in the end of 2010. It was intended to find out if people, that host and maintain datasets (taken from the Linked Data cloud diagram as of October 2010), developed own ontologies to describe their data and which ontology engineering efforts they spent. It was confirmed that ontologies which underlay Linked Data are semantically rather rudimentary, that people sparsely adapt to any engineering methodology for building them and that they see the need to evolve their ontologies in the future.

But, which is the factor that indicates ontology evolution for providers of semantic data on the Web if they do not follow any ontology engineering methodology that defines such a point? We claim that the current uptake of Linked Data usage opens a new direction in ontology engineering that helps to answer this question in a very agile way. In a similar fashion as website operators capture the requirements of website users to refine the website's structure, the analysis of usage data now can help to capture new knowledge or to detect weaknesses and bugs within shared data and the underlying ontologies. This brings together researchers from diverse fields such as the Semantic Web as well as Web usage mining and it **foments a new generation of ontology engineering – usage-based ontology engineering**. These visions are fundamentally inspired by our work in the field of collaborative ontology engineering[1] which, from our point of view, still left some issues with respect to the exhaustive technical insight people need when they follow one of the proposed methodologies in this area. In 2008 we proposed an ontology life cycle[2] that reflects the relevance of ontology usage to the engineering process with the goal to limit the necessity of engineering experience at least in some phases of the process. Since that date we also developed a method for log file preprocessing to facilitate an in-depth analysis of the usage of data and ontologies on the Web of Data with the purpose to recommend necessary ontology changes[3] and recently co-initiated with an interdisciplinary team an initial workshop[4] covering these emerging topics.

---

[1] C. Tempich, E. P. B. Simperl, M. Luczak, R. Studer, and H. S. Pinto, "Argumentation-Based Ontology Engineering.," IEEE Intelligent Systems, vol. 22, iss. 6, pp. 52-59, 2007.

[2] M. Luczak-Rösch and R. Heese, "Managing Ontology Lifecycles in Corporate Settings," in International Conference on Semantic Systems (I-SEMANTICS), 2008.

[3] M. Luczak-Rösch, "Towards Agile Ontology Maintenance.," in International Semantic Web Conference, 2009, pp. 965-972.

[4] 1st International Workshop on Usage Analysis and the Web of Data (USEWOD), co-located with the 20th International World Wide Web Conference (WWW2011), http://data.semanticweb.org/usewod/2011/