



Prague, August 3, 2010

The Board of Doctoral Study
Faculty of Mathematics and Physics

Report on the PhD Thesis "Methods for effective querying of RDF data"

by RNDr. Jiří Dokulil

This thesis summarizes works covering the author's study and research of so called Semantic Web, particularly Resource Description Framework (RDF).

In Chapter 1 the author specifies the topics of his research and the structure of the Thesis. Chapter 2 introduces shortly RDF basic notions, models and standards. The real work concerns three relatively stand-alone research areas, whose common topic is just RDF: RDF Visualization, a new query language over RDF data – TriQuery as well as a design and implementation of the Bobox parallelization framework for processing queries over RDF data. Chapters 3-5 can be conceived as the core of thesis. They cover the respective research areas mentioned above. Related works are included locally in these chapters.

Chapter 3 describes an original visualizer of RDF data developed by the author in cooperation with J. Katreniaková. The most valuable result presented here is a triangle layout algorithm used for incremental navigation in large RDF graphs and its theoretical analysis. A new algebra for RDF data is developed in Chapter 2. Technically, TriQuery query capabilities over RDF data are embedded into the XQuery language. A more software part of the thesis is represented by Chapter 5 describing the Bobox framework. Due to its low-level architecture, the framework is pretty general. It can support not only RDF-oriented query tools like SPARQL and SeRQL, but also, e.g., SQL language. Experiments done with SPARQL queries from SP³Bench benchmark confirm the power of the Bobox. It is well to emphasize that such big project like Bobox is a team work. Appendix 1 specifies the contribution of J. Dokulil in detail.

Appendix B provides an interesting approach to formalization of TriQuery language, i.e. a graph data model and particular algebraic operations are described there. My opinion is that this material could be organically included into Chapter 2.

The results presented in the thesis have been published in papers of representative international conferences, like e.g. SEMAPRO, Information Visualization, DEXA, SOFSEM, and ICSC as well as in the respected journal Information Sciences. The conference proceedings have been published by well-known Springer and IEEE publishers. Some papers of the author have been also cited internationally in a number of scientific works. The DBLP Bibliography Server records 11 items of which Dokulil is the author or co-author. Consequently, I recommend that the candidate be awarded the Doctor degree.

Prof. Jaroslav Pokorný
supervisor