

Towards Agile Knowledge Engineering: Methodology, Concepts and Applications

(University of Leipzig, Diss., 2006)

Abstract

The goal of this thesis is to help make the development and use of knowledge bases more efficient. For that purpose, a new, agile knowledge engineering methodology, called RapidOWL is proposed. This methodology is based on the idea of iterative refinement, annotation and structuring of a knowledge base. A central paradigm for the RapidOWL methodology is the concentration on smallest possible information chunks (i.e. RDF statements). The collaborative aspect comes into play, when those information chunks can be selectively added, removed, annotated with comments or ratings. Design rationales for the RapidOWL methodology are to be light-weight, easy-to-implement, and support of spatially distributed and highly collaborative scenarios.

The idea of applying agile paradigms to Knowledge Engineering with respect to the specific characteristics of Knowledge Representation is the major innovation of the presented approach. Such specific characteristics are for example: lack of a unique knowledge serialization, spatial separation of actors, and involvement of a large number of domain experts. Agile Software Development methodologies, on the contrary, assume a small team of programmers working closely (in particular spatially) with domain experts.

For efficient practical application of the methodology, strategies and techniques are presented supporting knowledge base versioning and evolution as well as rapid querying and view generation on the basis of the RDF statement paradigm. The approach to knowledge base evolution is efficiently implementable, since it can rely on mature relational database technologies. It is furthermore flexible enough to also capture information about changes of a knowledge base on higher conceptual levels than the one of statements. The application of the approach is not restricted to a specific knowledge representation language; the only requirement is a form of expression as RDF statements.

The caching and invalidation strategy presented in this document supports the RapidOWL methodology in that it enables the rapid querying, extraction and aggregation of knowledge bases. It allows a selective invalidation on knowledge base updates. Compared with caching strategies developed for database systems the presented approach allows both, an efficient straightforward implementation and selective invalidation of cached objects on updates.

The implementation of semantic-web applications is supported by Powl, a comprehensive, ready-to-use framework for semantic-web application development in PHP. OntoWiki is a semantic-web application based on Powl supporting Knowledge Engineering on the basis of the RapidOWL methodology. Finally evidence for the efficiency of RapidOWL is provided by a case study from the Virtual Organization domain.