Applying Ontology Modularization for Corporate Ontology Engineering

Gökhan Coskun, Markus Luczak-Rösch, Ralf Heese, and Adrian Paschke
(Freie Universität Berlin, Germany
AG Corporate Semantic Web
coskun,luczak,heese,paschke@inf.fu-berlin.de)

Abstract: As an innovative approach for developing and maintaining continuously evolving ontologies in corporate context the Corporate Ontology Lifecycle Methodology aims at providing an intuitive understanding of rising costs per iteration. In this regard having a close look at ontology partitioning techniques as well as module extraction approaches is very promising, because modular reuse of ontologies as well as ontology composition of building blocks have a reducing impact on costs. In this paper we are discussing the modularization aspects of the Corporate Ontology Lifecycle Methodology and provide a first approach for applying ontology modularization techniques in corporate contexts.

Key Words: Corporate Semantic Web, Ontology Modularization
Category: H.2, H.5.4

1 Introduction

Fast creation of economic value out of knowledge is essential for modern companies. Therefore the ability to adopt new achievements into the business processes rapidly is very important. In this regard semantic technologies are very promising for the knowledge management within companies. Corporate Semantic Web\(^1\) (CSW) is working on the adoption of semantic technologies in corporate contexts [Coskun et al. 2009]. Originated in this effort the Corporate Ontology Lifecycle Methodology [Luczak-Rösch and Heese 08] (COLM) is an innovative lifecycle model for continuously evolving ontologies in corporate contexts. The main intention is to provide an intuitive understanding of rising costs per iteration.

Thus having a close look at ontology modularization is very promising, because modular reuse of ontologies as well as ontology composition by building blocks have a significantly reducing impact on the rising costs for engineering as well as maintenance. In this paper we are discussing for which phases of COLM ontology modularization is beneficial and which aspects of ontology modularization are necessary in each phase.

The paper is organized in the following way. In Section 2 we present our understanding of ontology modularization. Afterwards, in Section 3 we provide a brief introduction to COLM and identify phases which can benefit in which

\(^1\) http://www.corporate-semantic-web.de
way from ontology modularization. In Section 4 we discuss how existing modularization mechanisms can be used in COLM, before finally closing this paper with a conclusion and an outlook in section 5.

2 Ontology Modularization

Component-based development of large and complex software systems by small well defined building blocks simplifies the understandability as well as the management and leads to reusable software modules and a scalable overall system. Accordingly, designing ontologies in a modular way is intuitively promising in order to benefit from the same advantages. However, modeling knowledge represented by ontologies is fundamentally different than software modules, which are mainly describing processes. The overall goals of modularization are ontology reuse, scalability, complexity management, understandability, and personalization. Due to different understandings of the term modularization in literature there is a need to clarify how this concept is understood in this work.

From an engineering perspective it is a design principle [d’Aquin et al. 2007] or an approach on developing ontologies. Similar to component-based software engineering it proposes to design ontologies by breaking down the complexity in smaller pieces and create small self-contained ontologies, called modules, which are used in order to compose more complex ontologies. On the other hand modularization is understood as a process which can be executed on ontologies to create smaller ontologies (e.g. within the NeOn Project\(^2\)). This understanding is in turn divided into two categories. Partitioning ontologies is a process in which a bigger ontology is divided into smaller pieces like a puzzle. The second category, called module extraction, is more an application oriented activity which aims at extracting a particular part of an ontology mostly for a special use case. This means, that some content of the original ontology might be lost.

3 Modularization in COLM

In this chapter we will give a brief introduction to COLM. Afterwards we will discuss how modularization can be applied to COLM and which requirements modularization mechanisms have to fulfil.

3.1 COLM

COLM (Figure 1) is an innovative lifecycle model for continuously evolving ontologies in corporate contexts. The main intention is to provide an intuitive understanding of raising costs per iteration and in dependence on the duration and effort spent in each process phase.

\(^2\) www.neon-project.org, NeOn Glossary of Activities to references
Starting the process at selection/development/integration means to start the knowledge acquisition and conceptualization, to re-use or re-engineer existing ontologies, or to commission a contractor to develop an ontology. The result of this phase is an ontology, which is validated against the objectives. At the intersection point between the engineering and the usage cycles the ontology engineers and the domain experts decide whether the ontology suites the requirements or not. If this is approved the ontology is deployed and in use by applications. Then the ontology is populated, which means that a process for instance generation from structured, semi-structured and unstructured data runs up. Throughout the whole feedback tracking phase, formal statements about users feedback and behavior are recorded. A reporting of this feedback log is performed at the end of the usage cycle. That means that all feedback information, which were collected until a decisive point, are analyzed respecting internal inconsistencies and their effects to the currently used ontology version. The usage cycle is left and the knowledge engineers evaluate the weaknesses of the current ontology with respect to the feedback log. This point may also be reached, when the validation shows that the new ontology is inappropriate to the specification. The lifecycle starts again with the implementation of the results of the evaluation.

### 3.2 Modularization Aspects of COLM

Modularization of ontologies is in different ways integrable into the lifecycle. As illustrated and mentioned previously COLM facilitates cost estimation in the run-up of cost-intensive evolution steps. The modularization and integration dimensions of COLM helps to decrease the capital expenditure (CAPEX) as
well as the operational expenditure (OPEX) and supports to realize ontology adoption to the corporate environment by keeping the incentives of a company in view.

At the very beginning of COLM during the selection and integration phases it is reasonable to look for existing ontologies for the sake of reusability, because developing ontologies from scratch is a very cumbersome and time-consuming task. Expecting candidate ontologies which perfectly fit into the targeted system is unrealistic, some customization will be necessary in order to adapt candidate ontologies and make them useful. At this point modularization is an important mechanisms to allow reusability even if the candidate ontologies are not usable in their original form. The possibility of extracting only relevant parts of existing ontologies and integrate them in order to achieve a useful ontology decreases CAPEX drastically and makes ontology application realistic and really attractive for companies. In this regard, some kind of semi-automatic module extraction mechanism which is driven by ontology engineers together with domain experts is needed. By defining relevant parts of candidate ontologies, the mechanism is responsible for extracting a useful module, which can be customized in further steps.

Modularization during the lifetime of ontologies is also possible. This can be done based upon diverse aspects. In the feedback tracking and reporting phases, the closed and controlled corporate environment allows obtaining information as relevance of concepts and relationships regarding departments and application. It also enables to observe the evolution of the ontology and allows to identify vague parts which change very often and well-established parts which change less frequently. During the development phase it is possible to create modules based upon the mentioned observations. The possibility to identify modules according to the probability to change increases reliability, while the efficiency is improved by relevance based modularization. For that reason, a partitioning mechanism is necessary, that is automatically able to split the ontology into different building blocks, based upon the mentioned aspects.

Finally, in case of context-sensitive and ontology-based applications, which we expect to arise in future, an additional aspect of ontology modularization is possible. Those applications should be able to define their ontology requirements. During the deployment phase, while application need to load ontologies, they can feed fully automatic working module extraction mechanisms with user-specific and application requirements. Based upon these requirements, optimized modules regarding the application context, can be identified and extracted in real-time. This would lead to personalization and increased efficiency of the ontology usage, because loading needless parts and wasting storage as well as computing power can be avoided.
4 Applying Modularization to COLM

In this chapter we provide a brief discussion how three selected mechanisms meet the three modularization aspects of COLM.

As mentioned before the possibility to reuse existing ontologies is expected to have a reducing impact on the capital expenditure. Using the extraction for reuse mechanism presented in [Doran et al. 2007] is very promising for the selection and integration phase of COLM. The mechanism is designed following three module requirements: self-containment, concept centered, and consistency. Given a start concept it produces self-contained and consistent modules which can be used as a starting point and can be customized in further steps. Ontology engineers can extend it with new concepts and properties or integrate it with other modules extracted from different candidate ontologies.

Structure-based partitioning [Stuckenschmidt and Klein 2004] and its extensions in [Schlicht and Stuckenschmidt 2006, Schlicht and Stuckenschmidt 2007, Schlicht and Stuckenschmidt 2008] is a candidate for the second modularization aspect of COLM, that is for creating modules based on the observations during the usage phases. In this regard the second step of the mechanism needs a modification, because the original mechanism uses only social network analysis methods in order to determine the strength of the dependency. Obtained information about the relevance of concepts for certain users or departments as well as information about the probability to change has to be taken into consideration. The determined strength of dependency has to reflect these observations, so that personalized modules for departments and users as well as evolution based modules can be created.

Module extraction for knowledge selection [d’Aquin et al. 2006] is a method which is very promising for the third modularization aspect of COLM. It is strongly tight to a particular application and fulfills requirements as minimizing the user interaction, creating the smallest part of the original ontology still covering needed terms while implicit information should be contained in the extracted module. In this regard, modules are created which are very efficient and application-specific. This is done without the need for ontology engineers.

5 Conclusion and Outlook

COLM provides a cost sensitive methodology for adopting ontologies by keeping the corporate setting in view. In this regard, we state that modularization is very promising in order to reduce CAPEX through modular reuse and OPEX through simplified complexity management. Accordingly, we have discusses in this paper how three existing mechanisms can be applied to COLM. However, they might need some modifications in order to take the corporate setting and its conditions into account.
Through this paper it comes clear that further work in this direction is promising and necessary. Especially the chance of reducing investment costs through reusing existing ontologies in a modular way needs further investigation, because it would lessen the barrier for companies to adopt ontologies and Semantic Web technologies in general. Furthermore, well-sized modules which evolve independently during the usage would lead to increased quality and improved user experience which is very important for long-lasting application of Semantic Web technologies in companies.

Acknowledgements

This work has been partially supported by the “InnoProfile - Corporate Semantic Web” project funded by the German Federal Ministry of Education and Research (BMBF) and the BMBF Innovation Initiative for the New German Länder - Entrepreneurial Regions.

References


[d’Aquin et al. 2006] d’Aquin M., Sabou M., Motta E.: Modularization: a key for the Dynamic Selection of Relevant Knowledge Components. In First International Workshop on Modular Ontologies, WoMO-2006, Athens, Georgia, USA


[Schlicht and Stuckenschmidt 2006] Schlicht A., Stuckenschmidt H.: Towards structural criteria for ontology modularization. In First International Workshop on Modular Ontologies, WoMO-2006, Athens, Georgia, USA


[Schlicht and Stuckenschmidt 2008] Schlicht A., Stuckenschmidt H.: A flexible partitioning tool for large ontologies. In International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT20’08, Sydney, Australia