Knowledge acquisition with geologists – a field report

S. Schade (1), P.Maué (1), J. Langlois (2) and E. Klien (1)

(1) Institute of Geoinformatics, Muenster, Germany
({schades,pajoma,klien}@uni-muenster.de)

(2) BRGM - Centre scientifique et technique, Orléans, France (J.Langlois@brgm.fr)

Introduction

One of the main tasks of the French geological survey (BRGM[1]) is to provide decision support documents. The traditional approach of the specialists, who usually come from the earth sciences domain, is the creation of reports by performing the following four steps:

1. retrieving information (data files) from the different providers,

2. collecting files on a local computer,

3. structuring, managing, and computing this information in a GIS, and

4. producing a report.

In the mineral resources domain, a typical discovery task (step 1) involves contacting about 20 different partners, identifying what information is available, defining a strategy for its exchange, and ensuring efficient local integration. Domain specialists deal with an ever increasing amount of relevant information, and have to face the increasingly complex challenge of finding and successfully integrating it. The work of BRGM represents just one case of many (Klien et al. 2006).

---

[1] More information available at http://www.brgm.fr

Driven by the European INSPIRE directive (INSPIRE 2003), a growing amount of geospatial information is made available online through Web Services (OGC 2005a) and searchable via Catalogue Services (OGC 2005b). Integration of such information has been simplified significantly through the standards proposed by the Open Geospatial Consortium (OGC). But efficient discovery of Geospatial Web Services is still an open challenge. New technologies based on the vision of semantically-enhanced discovery tools have been introduced to address this problem (Klien et al. 2006; Lutz and Klien 2006).

Ontologies have been identified as means to account for describing the semantics of information made available through Web Services (Stuckenschmidt 2003; Lutz and Klien 2006). An ontology is "an engineering artefact, constituted by a specific vocabulary used to describe a certain reality, plus a set of explicit assumptions regarding the intended meaning of the vocabulary words" (Guarino 1998). If an ontology can facilitate description of available resources as well as of queries, then reasoning mechanisms are able to automatically support the discovery and integration of (Geospatial) Web Services (Stuckenschmidt 2003; Lutz 2006).

Building software that allows for identifying appropriate information sources requires a systematic engineering process for creating the functionality (the program) and the models (the ontologies). Software engineering is an extensively studied area, and much has been learned during the early years of industrial software production from the software crisis (Summerville 2005). Building software based on ontologies is still in its infancy and ontology engineering nowadays has to face similar problems as software engineering in the past. Ontology engineers need to strongly collaborate with their clients, i.e. the domain experts, to assure acceptance of the product and to successfully complete projects. A primary issue for the development of stable and widely-accepted domain ontologies is the acquisition of knowledge from domain experts (Uschold and Grüninger 1996; Fernández-López et al. 2004).

In the remainder of this paper we describe a "best practice" in knowledge acquisition. The presented strategy has been developed and successfully deployed implemented in the scope of the EU-funded SWING project[2].

Background

Various procedures have been suggested to engineer ontologies. In the SWING project, the METHONTOLOGY (Fernández-López et al. 1997) has been identified as most promising. Here the development process is divided into five phases. Intended

use, reasoning requirements, and scope of ontologies in the application are identified in the *Specification Phase*. During *Conceptualization*, the knowledge of the domain of interest becomes structured. The transformation of the conceptual model into a formal or semi-formal model is subject to the *Formalization Phase*. It is followed by the *Implementation* of the ontology in a formal language. Once the ontology is released, updates and correction characterise the *Maintenance Phase*. Knowledge Acquisition is one of the accompanying activities to the ontology engineering process, and is continuously performed and validated throughout the engineering lifecycle (Klien et al. 2007).

Cooperation between engineers and domain experts is a key issue for a product that satisfies the requirements of end users. Again, ontology and software engineering are similar in this aspect. Building software requirements, use cases, user interfaces and underlying data models is performed in cooperation with the client. The actual implementation is in the responsibility of the specialists (with feedback from clients). The same is valid for ontology engineering, only the topic of collaboration and, the results are of a different type.

During the acquisition process, ontology engineers have to make sure that they structure the acquired domain knowledge in a way that enables its implementation in a formal representation language. However, in order to abstract from implementation details and complex logical formulas, an intermediate form of the representation is required for the collaboration (Novak and Gowin 1996). The details of formalisation are mostly irrelevant for domain experts and would produce unnecessary cognitive load. Nevertheless, the intermediate representation should support the (formal) structuring of natural language descriptions and the generation of logic statements in subsequent steps.

Ontology Engineering – Best Practise

A "best practise" of techniques for knowledge acquisition has been developed in the context of the SWING Project (Klien et al. 2007). It was tested in Knowledge Acquisition workshops, and the results have been used to develop and release[3] a set of ontologies for the domain of quarrying. In the following, we introduce the results in more detail.

Engineering in the Conceptualisation-Phase

The 20-question technique serves as introduction to knowledge acquisition. A domain expert and an ontology engineer perform an interview. The ontology engineer has a particular concept in mind, for example "Quarry", "Crushed Stone" or "Production

---

[3]The ontologies are available from BRGM's project page: http://swing.brgm.fr.

Rate". The domain expert has to guess the concept by asking up to 20 questions. The ontology engineer is only allowed to answer questions with "yes" or "no". The questions and answers are written down in protocols. This game-like approach is a simple, but also efficient solution to get a first impression on the taxonomic and other major relations used in the domain.

From Conceptualisation to Formalisation

As intermediate representation for the conceptualisation, we use concept maps (Novak and Gowin 1996). In a nutshell, concepts are visualised as labelled notes and relations between concepts are represented by labelled directed edges. The simple graphical representation provides the first step in structuring domain knowledge.

Based on the protocols from the first step, concepts and relations are identified and added to a preliminary concept map. It serves as material for direct evaluation of the transcribed knowledge and for a semi-structured interview in the same groups as before. The domain experts now become aware how we, the ontology engineers, intend to structure knowledge. The results of each group session are concept maps that capture the domain in the view of a single expert.

The maps from all groups provide input for a plenum in which a shared concept map is then generated from scratch. Central concepts and relations are identified and concepts and relations of subjective interest initiate group discussions.

In the next step, more complex information like natural language descriptions of concepts, and cardinalities and mathematical properties of relations are acquired. A matrix is used for this purpose; the elements building the concept map are directly filled into the more complex structure. Missing information is then added by domain experts. The matrix is the format that is used as basis for the creation of the ontologies, formalised, for example, in the Web Service Modelling Language (WSML) (de Bruijn 2006).

The first application of the techniques in the SWING workshop resulted in an ontology that was build out of 36 concepts and 39 relations. This ontology serves as the basis for describing the semantics of Geospatial Web Services in the context of the SWING project, which are all publicly available. Moreover, the basis of a domain vocabulary for the domain of mineral resources has been established and both, domain experts and ontology engineers got insides from a previously unknown area of knowledge.

Discussion and Conclusion

At the workshops, specialists have been introduced into the knowledge engineering process. Different utilized techniques and tools gave them an opportunity to re-

consider the experts' work procedures. In the specific domain of quarry management, it was a motivation for the experts to achieve a conceptual model, which allowed them for adjusting the concept maps and tables whenever new ideas were introduced. The rules for cooperative editing were well understood.

From an ontology engineer's perspective, the identification of requirements and risks were of central importance. Knowledge is an unlimited resource, defining a scope and limiting oneself in the number of concepts is crucial. Before starting with the conceptualisation, a team of domain experts having (diverse) knowledge on the setting has to be established. Identifying possible use cases is helpful to determine the required scope for the ontologies. During the process, engineers should regularly ask themselves if the current work contributes to solve the tasks defined in the use cases, or if it fills non-critical gaps, improves style, or is another occupation which just costs time and risks successful completion. Keeping focussed on the core concepts and allow only occasional side discussions is a necessary constraint. Regular reviewing of the conceptualised knowledge is necessary since knowledge is evolving over time. Our experience in knowledge modelling has shown that such changes are crucial for the acceptance of ontologies, but can also be reason for long discussions.

We tried to give an account of our experience of past knowledge engineering projects, performed in cooperation with experts coming from different disciplines in the geospatial domain. In essence, we could identify the following four statements, which will be considered in future projects:

1. Collaboration between domain experts and ontology engineers is necessary throughout the engineering lifecycle.

2. Successful collaboration with non-experts requires an intermediate view, which needs to be separated from the final formalisations of the ontologies.

3. Knowledge engineering is an iterative process: knowledge is changing over time and this change needs to be reflected in the underlying model.

4. Game-like character of a knowledge acquisition meeting resolves sceptical and too respectful views on the notion of ontologies.

We stressed the similarities between knowledge engineering and software engineering several times. Model-driven engineering (MDE) makes use of models as engineering artefacts during the complete engineering lifecycle (Frankel and Parodi 2004). A well-known example for this is software engineering with the help of the Unified Modelling Language (UML). Developing ontologies with the help of tools which conform to the

MDA (Model-driven architecture) is a promising approach and will be further investigated. Having MDA-tools which are able to visualize the intermediate representation, but can also output ontologies encoded in formal languages, would support the ontology engineering process. MDA can also enable round-trip engineering, making it possible to easily visualize already existing ontologies, modify the concept maps together with domain experts, and re-export them as ontologies.

The specialists' objective is now to employ the ontologies in their application to improve the way in which they manage domain-specific knowledge. Particular attention for measuring the success of our ontology engineering activities in the SWING project will be paid to the trade-off between the benefits from using the ontologies in the application versus the efforts for ontology maintenance.

Acknowledgements

References

.De Bruijn, J. (2006). The Web Service Modelling Language WSML: An Overview. Springer.

Fernández-López, M., A. Gómez-Pérez, et al. (2004). Ontological Engineering with examples from areas of Knowledge Management, e-Commerce and the Semantic Web. Springer.

Fernández, M., A. Gómez-Pérez, et al. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Workshop on Ontological Engineering.

Frankel, D. S. and J. Parodi (2004). The MDA Journal: Model Driven Architecture Straight From The Masters. Meghan Kiffer.

Guarino, N. (1998). Formal Ontology and Information Systems. Formal Ontology in Information Systems. IOS Press.

INSPIRE (2003). Consultation Paper on a forthcoming EU Legal Initiative on Spatial Information for Community Policy-making and Implementation.

Klien, E., M. Lutz, et al. (2006). Ontology-Based Discovery of Geographic Information Services - An Application in Disaster Management. Computers, Environment and Urban Systems 30(1).

Klien, E., Schade, S. and Hoffmann, J. (2007). D3.1 Ontologies in the SWING Appli-

cation - Requirement Specification. Deliverable of the SWING Project, Muenster.

Lutz, M. (2006). Ontology-Based Descriptions for Semantic Discovery and Composition of Geoprocessing Services, GeoInformatica.

Lutz, M. and E. Klien (2006). Ontology-Based Retrieval of Geographic Information. IJGIS 20(2).

Novak, J. D. and D. B. Gowin (1996). Learning How To Learn. Cambridge University Press.

OGC (2005a). OpenGIS® Web Service Common Implementation Specification, Version 1.0.

OGC (2005b). OpenGIS® Catalogue Service Implementation Specification, Version 2.0.1.

Stuckenschmidt, H. (2003). Ontology-Based Information Sharing in Weakly Structured Environments. Vrije Universiteit Amsterdam. PhD.

Summerville, I. (2005). Software Engineering. Addison Wesley.

Uschold, M. and M. Grüninger (1996). Ontologies: Principles, Methods and Applications. Knowledge Engineering Review 11(2).