Master's Thesis (Abstract)

Ontology-Based Knowledge Modelling

– usability as the semantic basis of the Semantic Web –

Michael Stollberg

This is an abstract of a master's thesis handed in at the Institute of Media and Communication Studies at the Free University of Berlin, Germany, on 29th October 2002. As the original thesis is written in German, the purpose of this paper is to give a brief summery of the major aspects and the conclusions of the thesis in English language.

Aim of the thesis

Ontologies are proposed to be a basic building block of the Semantic Web. They shall describe information structures on the World Wide Web by defining the meaning of terms and their relationships and thereby provide a machine-readable, uniform set of vocabularies that can be used as the semantic basis for information processing web-applications.

The aim of the thesis is to investigate on why ontologies are the favoured technique to create the semantic basis of the Semantic Web. Therefore the methods and techniques for creation, representation and management of ontologies are analysed with regard to their suitability to provide this functionality. The first three chapters each focus on one aspect of ontology-based knowledge modelling. The forth chapter deals with the use of ontologies in the Semantic Web.

1 Ontology-Based Knowledge Modelling

The first chapter works out the foundations of ontologies as a knowledge modelling technique in Artificial Intelligence. The most commonly accepted characterization is the definition of an ontology as a formal, explicit specification of a shared conceptualisation.

This means that an ontology is a conceptual model of a universe of discourse which represents knowledge structures of a domain. The people concerned with the domain should accept this model as a shared, common understanding of the domain. The intended meaning of terms should be clearly defined in an ontology so that misinterpretations can be avoided. By representing this model in an appropriate formalism an ontology can be processed by machines and thus be used in computer-based information systems.

The reason why ontologies are used in information systems is that they offer a formal specification of the structure of a domain including definitions of the meaning of the terms. This set of vocabulary can be used in information processing systems – either as an integrated component of a system or as a complementary tool in system development.

2 Ontology Development

The second chapter discusses methodological approaches for ontology development, which is the main focus of Ontology Engineering research. As no comprehensive methodology for ontology development has been created at this point of time, this chapter compiles existing methodological approaches for ontology development.

The process of ontology development consists of four major phases. At first the purpose and requirements for the ontology are defined. Then, the conceptualisation underlying the ontology is elaborated by describing the domain structure using concepts, properties and relations as ontological modelling primitives. As the conceptual model is meant to be an explanation of the domain, the knowledge structures must be represented correctly. Next, the conceptual model is conveyed into a formal representation to ensure that the ontology can be read by machines. Appropriate formalisms must be able to specify the knowledge structures exactly and explicitly in the way they are defined in the conceptual model. As the last step of the development process the ontology is evaluated.

To be usable as components in information systems ontologies have to meet high requirements on conceptual correctness and adaptability to the system. To ensure suitable quality of ontologies, design principles are proposed that should be followed throughout the development process. Further on, techniques for

post-evaluation of ontologies are announced – for example the ´OntoClean-Methodology´ for heuristic validation of taxonomic structures.

3 Ontologies in Information Systems

The third chapter deals with improvements achievable by the use of ontologies – especially for the basic information processing abilities of web-applications. Therefore different scenarios for using ontologies in information systems are distinguished.

The first scenario is to use an ontology as a set of pre-defined terms that can be employed as shared vocabulary by different applications. An example is using a common ontology to secure semantically correct communication between autonomous agents in multi-agent-systems. The second scenario is the usage of an ontology to solve semantic heterogeneities between information sources to enhance information exchange facilities. Therefore either a common conceptual model for all information sources is defined in a global ontology, or each source is described by a local ontology: Semantic interoperability between the sources is then permitted by defining mappings between the ontologies. In the last scenario an ontology can be used for indexing information and thereby improve information retrieval techniques. The major benefit of ontology-based information retrieval is to enable semantic matching: An ontology allows to determine the semantic context of a query so that the result set can be optimised to only contain the information relevant to the context intended by the query.

The use of ontologies in information systems demands techniques for handling ontologies. The main challenges are to make an ontology reusable for different applications and to ensure interoperability between heterogeneous ontologies. Although there is a lot of research activity in this field, existing techniques do only support these tasks with semi-automatic tools while the actual handling of ontologies relies on human investigation.

4 Ontologies in the Semantic Web

The last chapter of the thesis examines the use of ontologies in the Semantic Web. The Semantic Web as proposed by the World Wide Web Consortium envisions the future development of the Internet by enriching information processing capabilities of web-technologies.

As stated above, the proposed architecture for realising the Semantic Web claims ontologies as the technique for creating a uniform semantic basis of the information structures on the Internet. This basis is required to attain interoperability of web-applications and shall be provided by a ´web of ontologies´: Therein each information source on the web is described by an ontology and the interoperability of the sources shall be accomplished by defining mappings between their ontologies.

To achieve this functionality, languages for the formal specification of ontologies are required that are compatible with existing web-technologies. The W3C is working on the OWL Ontology Web Language as a standardized ontology-specification-language for the Semantic Web. As this language builds up on RDF – a basic technique for meta-description of web resources – and is founded on the DAML+OIL specification-language, it seems to be a suitable solution for this challenge.

The second challenge is to develop adequate technical solutions for the management of the desired web of ontologies. Apart from tools to facilitate the usability of ontologies, methods for managing mappings between ontologies are of major importance. As worked out before, techniques for integrating ontologies require humans to carry out the actual integration task. Maintaining a web of interoperating ontologies would necessitate an inappropriate amount of human investigation. Therefore a different approach for building the semantic basis is proposed: Separate information spaces shall be described by standardized ontologies so that the necessity of human-driven integration of ontologies can be reduced.

Conclusion

Ontologies seem to be a proper technical solution for creating a semantic basis for the Semantic Web. But if the standardisation of ontologies is technically required, the question arises if ontologies are really needed in the Semantic Web – because formally specified semantics only seem to be necessary if the interoperability between information spaces with heterogeneous conceptual models has to be guaranteed.

Berlin, Germany, November 2002

For remarks of for further information please contact me at: mail@michael-stollberg.de