Chapter 14
Linguistics-Based Modeling Methods and Ontologies in Requirements Engineering

Florian Lautenbacher
University of Augsburg, Germany

Bernhard Bauer
University of Augsburg, Germany

Tanja Sieber
University of Miskolc, Hungary

Alejandro Cabral
Oracle Strategic Program, Argentina

ABSTRACT
Developing new software based on requirements specifications created by business analysts often leads to misunderstanding and lack of comprehension, because of the different backgrounds of the people involved. If requirements specifications instead have a clearly defined structure and comprehensive semantics, this obstacle can be resolved. Therefore, we propose to structure the requirements specifications using existing linguistics-based modeling methods and annotate the used terms with ontologies to enhance the understanding and reuse of these documents during the software engineering process.

1. MOTIVATION
Not only normal software development, but also the upcoming research area semantic-based software development (de Cesare, 2007) typically has an iterative software development process starting with the requirements engineering and requirements analysis phase. Before beginning with the development of software, the needs of the customer must be clarified and summarized into requirements specifications. These requirements contain all (or nearly most) of the details about the software product to be developed and are normally described in natural language. Some companies have therefore defined styleguides. However, most of the used terms are not defined in a concrete way which leads to misinterpretation
and incomprehension, i.e. the semantics are not defined clearly. Sometimes glossaries are used to describe the expressions, but even those can be interpreted differently by various readers/writers. Missing or not clearly defined requirements lead to change requests for the software product once it is tested or, in the worst case, when it is used by the customers. The customers might have thought of something different, but their requirement has not been described properly in the requirements specification. Therefore, it is critical to specify the requirements as precisely as possible in the first place to avoid unnecessary changes to the finished product afterwards and to build the product on time and in budget.

As stated in Rupp (2006), software (S) is a combination of documentation (D) and code (C), i.e. S=D+C. The documentation should not only cover the source code and its comments itself, but also the description of using the product afterwards (software documentation), any kind of technical specification and documentations, like functional and non-functional aspects, UML diagrams or database descriptions, etc. In document engineering, which is concerned with these issues, internal and external document engineering can be distinguished. The former refers to the documentation produced during the whole software development process, while the latter refers to the documentation produced for the system’s users after the product is released (Rueping, 2003). Requirements specifications can be seen as a typical example of internal documents, whereas user manuals are typical external document examples. There are some linguistics-based modeling methods that are widely used in external document engineering, which could also be used for internal document engineering, e.g. for gathering requirements. Using these modeling methods, the structure of documents and their underlying dependencies can already be reflected in the modeled segmentation of the documents, making it easier to be derived and annotated with semantic data afterwards. This semantic annotation is based on ontologies and can be used to describe the meaning of the constructs in a way that computers can not only read but also interpret.

We will therefore show how the semantics of requirements specifications can be gathered using linguistics-based modeling methods and that an annotation of these documents with ontologies can foster reuse and personalization.

This article is structured as follows: in the next section we describe the challenges of current documents and the difference of understanding some data between sender and recipient. Additionally, we describe our definition of data and how the communication between different persons takes place. Afterwards, we show how different linguistics-based modeling methods can be used to clarify the underlying meaning of terms. We evaluate several linguistics-based modeling methods and show a summary of our evaluation. We then use an example to clarify the usage of the modeling methods as well as introduce the process and benefits of semantic annotation through the usage of ontologies. Subsequently, we show some related work before we conclude describing the benefits of using linguistics-based modeling methods and ontologies.

2. CHALLENGES OF SEMANTIC REQUIREMENTS ENGINEERING

In this section we introduce the basics of linguistics such as the Speech act theory, before we introduce models for the description of data and the process of communication that are required for understanding the problems in Requirements Engineering and possible solutions.

Speech Act Theory

John Langshaw Austin developed his Speech Act theory in such a way that today, more than 45 years later, we find it useful to conduct our research on semantics in requirements engineer-