

Reusing Terminology for Requirements Specifications from WordNet

Katharina Wolter
HITeC e.V. c/o University of Hamburg
kwolter@informatik.uni-hamburg.de

Daniel Bildhauer
University of Koblenz-Landau
dbildh@uni-koblenz.de

Michał Śmiałek
Warsaw University of Technology
smialek@iem.pw.edu.pl

Hermann Kaindl
Vienna University of Technology
kaindl@ict.tuwien.ac.at

Abstract

In order to make requirements comprehensible to humans and as unambiguous as possible, a glossary and/or domain model is needed for defining the terminology used. Unless these are available from related projects, however, they are hard to create. Therefore, we propose to reuse terminology and its definition for requirements specifications from the semantic lexicon WordNet. For making this useful, however, we had to deal with the issue of disambiguation of the general terminology there for a given domain of a requirements specification.

1 Introduction and Background

Natural Language is still the most widely used requirements language, partly since specifications written in formal languages are not really comprehensible to most humans. Unfortunately, natural language is inherently ambiguous and does not facilitate automated manipulations well.

Therefore, we defined a new requirements specification language (RSL).¹ While it still allows the use of natural language, an important part of it provides constrained language in order to make the specification more precise.

Both for natural and constrained language, however, an important issue is the terminology used. In order to make requirements specifications both more precise and comprehensible to humans, usually domain-specific glossaries are provided. A more elaborate approach is to combine this with domain models [2]. It is hard to create such glossaries and/or domain models for new domains, however.

¹This research has been carried out in the ReDSeeDS project and is partially funded by the EU (contract number IST-33596 under the 6th framework programme), see <http://www.redseeds.eu>.

We propose to utilize terminology from *WordNet*² [1], since it offers some of the advantages of usual glossaries and of domain models i.e., definitions in natural language and relations between concepts, respectively. WordNet is based on the concept of synonym sets (called *synsets*) that group synonymic words of the English language. Amongst others, the following semantic relations connect synsets: *hypernyms / hyponyms* (is-a, is-a invers) and *holonym / meronym* (part-of, part-of invers). Words with different meanings participate in several synsets and a *frequency score* indicates which meanings are more common and which are used rather infrequently. Using WordNet, requirements descriptions refer to a newly defined representation of vocabulary, which is organized as a terminology representation that integrates a dictionary with a thesaurus.

2 Terminology Reused from WordNet

In RSL, the definitions of notions and words are kept separately from the requirements in a domain specification composed of Domain Elements (see Fig. 1), which can be Notions or Actors. Objects in constrained language sentences (e.g., “time schedule”) refer to Notions, while Subjects (e.g., “customer”) refer to Actors. All the Phrases containing the same main Noun are grouped within the same Notion.

Domain elements or phrases may contain words that have different meaning in different domains. Therefore, to make a requirements specification unambiguous and reusable in a global context, it has to have its own domain specification, which specifies the meaning of individual words unambiguously. RSL reduces the effort for the time consuming specification task by enabling reuse of definitions from former specifications and WordNet.

²WordNet is a semantic lexicon that was developed at the Cognitive Science Laboratory at Princeton University, see <http://wordnet.princeton.edu/>.

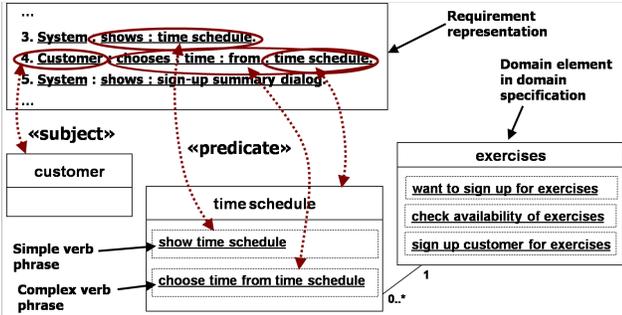


Figure 1: Hyperlinks contained in requirements representations referring to vocabulary elements

Our approach is based on a global terminology that subsumes all words of all domain specifications and combines the functionalities of a dictionary and a thesaurus. It contains words and their inflections as well as their definitions. The words are divided into content words and function words. *Function words* (Determiners, Prepositions, Conjunctions, Pronouns and ModalVerbs) have little semantic content on their own and mainly define grammatical relationships. In contrast, *content words* (Nouns, Verbs, Adjectives and Adverbs) have a clear lexical meaning [1]. We use WordNet in order to define the meaning of content words.

Fig. 2 illustrates the relations between requirements specifications, domain specifications and the global terminology connected with WordNet. The figure also shows how ambiguity problems of natural language (e.g., caused by synonyms or homographs) are solved in RSL. One domain specification contains e.g., the word *Customer*, while the other two contain the word *Client*, but with different meanings. This homograph appears twice in the global terminology and is linked with different synsets in WordNet. Since the word *Customer* is used as a synonym for *Client* (someone who pays for goods or services), it is connected with the same synset.

The linking of words is done semi-automatically by our developed requirements management tool RSLEditor. It supports the requirements engineer to specify the meaning of new Domain Elements. For determining the lexical categories of words and for selecting the right definition, the tool allows browsing and reusing existing definitions in the global terminology and WordNet. Usage of definitions from mature semantic lexica like WordNet can lead to specifications of higher quality with reduced workload. Furthermore, the meaning of content words is specified in a way that is both comprehensible for humans and (partly) processable by automatic reasoning.

The latter is a major advantage regarding reuse of requirements specifications. Reasoning mechanisms can be used for finding requirements specifications with similar meaning and not only with similar words [3]. Semantic

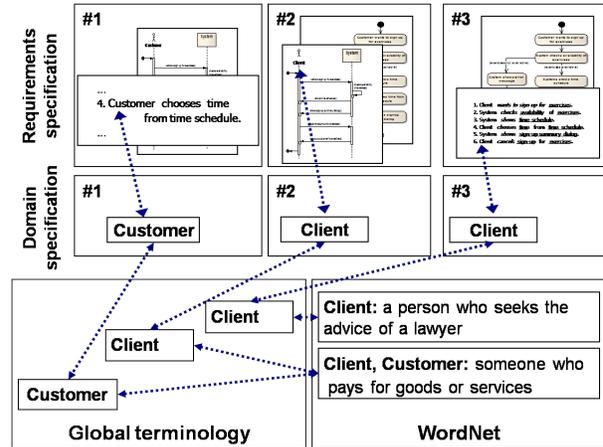


Figure 2: Relations between requirements representations, domain elements and the global terminology

relations can be evaluated in order to identify specifications that use not the same word but one that expresses the same or a similar meaning. Specifications that use the same word but with a different meaning (homographs) can be determined based on the relation between elements of the global terminology and synsets in WordNet. Thus, the approach enables semantic-based processing of requirements specifications without the need for sophisticated and completely formal knowledge modelling. A first implementation of these mechanisms is already available in the mentioned RSLEditor tool.

3 Conclusion

The terminology of requirements specifications using natural or constrained language is usually defined in glossaries or domain models. Instead of creating these from scratch for a new domain, we propose to utilize the general terminology from the semantic lexicon WordNet. It combines advantages from both glossaries and domain models. However, we had to deal with the issue of disambiguation of the general terminology given there. In our new language RSL, we provide for a terminology representation that links to the right entries of WordNet.

References

- [1] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [2] H. Kaindl. A practical approach to combining requirements definition and object-oriented analysis. *Annals of Software Engineering*, 3:319–343, 1997.
- [3] K. Wolter, T. Krebs, D. Bildhauer, M. Nick, and L. Hotz. Software case similarity measure. Project Deliverable D4.2, ReD-SeeDS Project, 2007. www.redseeds.eu.