Extended Tagging in Requirements Engineering

Günther Fliedl, Christian Kop, Heinrich C. Mayr,
Jürgen Vöhringer, Georg Weber, Christian Winkler

University of Klagenfurt

Abstract. In this paper standard tagging mechanisms are discussed and partly criticized. We propose a semantically and morphosyntactically enriched mechanism in which many of the shortcomings of standard POS-taggers are eliminated. Therefore rule based disambiguation steps are presented. They include mainly a specification of contextually motivated verbal subcategories. We need this fine grained system for better preprocessing of requirements texts which have to be highly explicit and non ambiguous. This linguistic preprocessor can support an interpretation tool used to extract semantic concepts and relations for the requirements engineering step in software development.

1 Introduction

Classical tagging approaches use standardized (POS) tag sets. Such kind of standardized tagging (e.g., Brilltagger [1], TnT [2], Q-Tag [11] Treetagger [8], [10] etc.), however, show weakness in the following three aspects:

– Tags like 'VAINF' provide only basic categorial and morphological information;
– Ambiguity cannot be made explicit;
– Chunking and identification of multiple tokens is not possible.

To avoid such deficiencies, we developed a system called NIBA[1]-Tag which allows tagging of German with an extended tagset, and inheritance of morphosyntactic and morphosemantic features. Morphosemantic tagging in our sense is labeling words by morpho-syntactically relevant semantic classifiers (sem-tags like 'tvag2', 'eV', 'indef', 'poss', etc.; see Appendix 1 for a rough comparison with the Treebank [7] STTS [S*99]), It has proved to be an efficient method for extracting different types of linguistically motivated information coded in text. The XML-Tagger-output for the German PP (=P2) ***Bei Eintreffen des Auftrags*** in Table 1 shows how the tagging result is structured.

As can be seen in Table 1, our tagging system has the following linguistic competence:

---

[1] NIBA is a German acronym for Natural language based Requirements Engineering. The project NIBA is supported by the "Klaus Tschira Stiftung Heidelberg".