On the Semiautomatic Generation of Romanian Noun Synsets

Claudia SIMA and Cristina VAȚA

Abstract

WordNet (WN) is a lexical knowledge base, first developed for English and then adopted for several European languages, which was created as a machine-readable dictionary based on psycholinguistic principles [Miller et. al., 90]. Its "building block" is a synonym set *(synset)* of all words belonging to the same part of speech that express a given concept. Various papers have discussed the *semiautomatic generation* of WNs for languages other than English, o topic of great interest since the existence of such WNs will create the appropriate infrastructure for advanced Information Technology Systems. The basic algorithmic approach to WN generation is introduced in [Nikolov and Petrova, 01), who rely entirely on a so-called "class method", namely one that uses as knowledge sources individual entries coming from bilingual dictionaries and WN synsets. This approach is extended in [Hristea, 02] where the need to combine such methods with structural ones is demonstrated. The present paper enriches the algorithmic approach of [Hristea, 02] with respect to noun synset generation, the focus being on the necessity to combine class methods with structural ones. The target language for performing tests will be Romanian. The proposed enriched method decrease failure in automatic noun synset generation with approximately 50%.