Towards building Wordnet for the tatar language: A semantic model of the verb system

Galieva A., Nevzorova O., Gatiatullin A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© Springer International Publishing Switzerland 2014. Wordnet is a lexical database where nouns, verbs, adjectives, and adverbs are organized in a conceptual hierarchy linking semantically and lexically related concepts to each other. This paper reports on the prototype of the Tatar Wordnet which currently contains about 5,500 Tatar verbs. Within our project we are creating a model of the semantic system of Tatar verbs as a hierarchical structure considering specifics of the Tatar language. For this purpose we use the entries of available Tatar dictionaries (explanatory dictionaries and those of synonyms). As the first step the extraction of available verbal synonyms from the dictionary of synonyms of the Tatar language was carried out. Then the most frequent 5156 Tatar verbs were selected and classified into several groups (synsets) according to their dominant semantic components with the purpose of adding new synsets and enriching those already existing (currently about 1,500 core synsets were distinguished). Then semantic relations between synsets were mapped (the verbs were linked according to their troponymy, entailment, and causality). The paper presents the results obtained, and discusses some problems encountered along the way.

Keywords

Synset, Tatar language, Verb, Wordnet