

Knowledge Graph Extension for Word Sense Annotation

Kiril Simov and Petya Osenova

Institute of Information and Communication Technology, BAS,
Akad. G. Bonchev. 25A, 1113 Sofia, Bulgaria
{kivs,petya}@bultreebank.org

1 Abstract

The recent success of knowledge-based Word Sense Disambiguation approaches depends on the quality of the knowledge graph – whether the knowledge represented in terms of nodes and relations (arcs) between them is sufficient for the algorithm to pick the correct senses of ambiguous words. Several extensions of the knowledge graph constructed on the basis of WordNet have been proposed and already implemented.

In this paper we present an approach towards the enrichment of Word Sense Disambiguation knowledge bases with data-driven relations from gold standard corpora. We focus on Bulgarian (BulTreeBank) and English (SemCor) as use cases, but our approach is scalable to other languages as well. Such an approach is justified by the fact that the lexical databases are sparse with respect to the available knowledge, its density and appropriateness. Also, the predominance of paradigmatic knowledge (synonymy, hypernymy, etc.) is balanced by the addition of syntagmatic relations (valency). For the purpose of exploring such methods, the Personalized Page Rank algorithm is used. Analysis is provided on the impact of the various semantic relations, which lead to different variants of the knowledge graph.

In our current work we perform an analysis of the various semantic relations in WordNet and Extended WordNet knowledge graphs. The investigation is performed via experiments with different subgraphs that include only part of the semantic relations in these resources. Some of the relation types allow for inference to be applied over them. Thus, inferred semantic relations have been included in some of knowledge graphs as well.

In our experiments we separated the relations in WN into 16 sets of relations corresponding to the relations in PWN3.0. Here we present the combination of synsets in each relation as parts-of-speech. The parts-of-speech are: A — adjective, N — noun, R — adverb, and V — verb. Also we present the number of links for the relation in WordNet.

The inclusion of inference has only a modest impact on accuracy, while the addition of syntactic relations produces stable improvement over the baselines.