

## **Towards an Efficient Combination of Similarity Measures for Semantic Relation Extraction**

*Alexander Panchenko*

9:50-10:10 Semantics I DZ6

Semantic relations such as synonyms and hyponyms are useful for various NLP applications, such as word sense disambiguation (Patwardhan et al., 2003), query expansion (Voorhees, 1994), document categorization (Tikk et al., 2003), question answering (Sun et al., 2005), etc. Semantic relation extraction techniques aims to discover meaningful relations between a given set of words.

One approach for semantic relation extraction is based on the lexico-syntactic patterns which are constructed either manually (Hearst, 1992) or semi-automatically (Snow et al., 2004). We study the alternative approach, which relies on a similarity measure between lexical units (see Lin (1998) or Sahlgren (2006)). In spite of the significant improvements during the last years, the similarity-based relation extraction remains far from being perfect: Curran and Moens (2002) compared 9 measures and their variations and report Precision@1=76%, and Precision@5=52% for the best measure. Panchenko (2011) compared 21 measures and reports Fmeasure=78% for the best one.

Previous studies suggest that different measures provide complimentary types of semantic information. In our ongoing research we are trying to exploit the heterogeneity of existing similarity measures so as to improve relation extraction. We investigate how measures based on semantic networks, corpora, web, and dictionaries may be efficiently combined.

First, we present an evaluation protocol which is adopted to the similarity-based relation extraction. Second, we compare baseline corpus-, knowledge-, web-, and definition-based similarity measures with this protocol. Finally, we present our preliminary results on combination of different measures. We test three types of combination techniques – based on relation, similarity, and feature fusion.