

A Metric for Word Similarity in WordNet

T. Hong-Minh¹ and D. Smith²

Abstract: Understanding concepts expressed in natural language is a challenge in Natural Language Processing and Information Retrieval. It is often decomposed into comparing semantic relations between concepts, which can be done by using Hidden Markov model and Bayesian Network for part of speech tagging. Alternatively, the knowledge-based approach can also be applied but it was not well explored due to the lack of machine readable dictionaries (such as lexicons, thesauri and taxonomies). However, more dictionaries have been developed so far. Following this approach, we present a measure for semantic similarity between concepts. By exploiting advantages of distance (edge-base) approach for taxonomic tree-like concepts, we enhance the strength of information theoretic (node-based) approach. Our measure therefore gives a complete view on word similarity, which can not be achieved by solely applying node-based approach.

^{1,2} School of Computing Sciences
University of East Anglia
Norwich, Norfolk, UK, NR4 7TJ
mtht@cmp.uea.ac.uk, djs@cmp.uea.ac.uk