Using Part of Speech Tags as Deep Syntax Indicators in Determining Short Text Semantic Similarity

V. Batanović, D. Bojić

References:

COMPUTER SCIENCE AND INFORMATION SYSTEMS - COMSIS, Vol. 12, No. 1, pp. 1-31, Jan, 2015

Abstract:

This paper presents POST STSS, a method of determining short-text semantic similarity in which part-of-speech tags are used as indicators of the deeper syntactic information usually extracted by more advanced tools like parsers and semantic role labelers. Our model employs a part-of-speech weighting scheme and is based on a statistical bag-of-words approach. It does not require either hand-crafted knowledge bases or advanced syntactic tools, which makes it easily applicable to languages with limited natural language processing resources. By using a paraphrase recognition test, we demonstrate that our system achieves a higher accuracy than all existing statistical similarity algorithms and solutions of a more structural kind.

Keywords:

statistical similarity, corpus-based measures, bag-of words model, short-text semantic similarity, natural language processing, POS weighting, syntactic information, part-of-speech tags