

Using Priming in the Web-Based Probabilistic Measures of Semantic Similarity

Khaled Elghamry

The semantic similarity of linguistic units (words, phrases, sentences, and full documents) plays an important role in both Natural Language Processing (NLP) and psycholinguistics. In NLP, semantic similarity is central for tasks such as clustering, classification, word sense disambiguation, information retrieval/extraction, paraphrasing, construction of semantic lexicons, and summarization. In psycholinguistics, semantic similarity effects are powerful and well documented in word recognition (associative and semantic priming) and word production literature. In this talk, I will show how the priming-related findings in psycholinguistics can be used to improve the accuracy of the Web-based probabilistic measures of semantic similarity in the domain of computational linguistics. This talk is divided into three main parts. The first part is concerned with the different aspects of semantic similarity in NLP and reviews previous methods for measuring this similarity with a special focus on Web-based methods. The second part reviews priming-related research and findings in psycholinguistics. The final part presents methods for and experiments on incorporating the priming effect in the Web-based measures of semantic similarity. The initial results of these experiments show that priming-sensitive measures are more accurate than priming-insensitive measures in establishing semantic similarity.