Abstract

Sentence similarity is considered the basis of many natural language tasks such as information retrieval, question answering and text summarization. The semantic meaning between compared text fragments is based on the words' semantic features and their relationships. This article reviews a set of word and sentence similarity measures and compares them on benchmark datasets. On the studied datasets, results showed that hybrid semantic measures perform better than both knowledge and corpus based measures.

References


Index Terms

Computer Science      Artificial Intelligence

Keywords

Word Similarity, Sentence Similarity, Corpus Measures, Knowledge Measures, Hybrid Measures, Text Similarity