Abstract

Automatic keywords extraction is very useful for text summarization, information retrieval and other natural language applications. This paper proposes a keyword extraction method that extracts keywords from Arabic documents based on the semantic similarity of words. The proposed method groups words into classes and put the words that are similar in the same class. Word2vec word embedding model is used to represent words using vectors. Accordingly, the semantic similarity between words can easily be computed using cosine similarity. Therefore, the words that have high context semantic similarity will be grouped in the same class. This method can also generate abstractive keywords that do not exist in original text using Word2vec. The experiments are conducted using three documents. Results showed that, the proposed model enhances the performance of keyword extraction with comparison to the previous models in terms of precision, recall and F-measure.