

Suleiman D., Awajan A., Al-Madi N., 2017. Deep Learning Based Techniques for Plagiarism Detection in Arabic Texts. Proceedings of the International Conference on New Trends in Computing Sciences (ICTCS 2017), Amman, Jordan. 978-1-5386-0527-1/17 © 2017 IEEE, DOI 10.1109/ICTCS.2017.42. Pages 216- 222

Abstract

Plagiarism detection is very important especially for academician, researchers and students. Although, there are many plagiarism detection tools, it is still challenging task because of huge amount of online documents. In this research, we propose to use word2vec model to detect the semantic similarity between words in Arabic language which can help in detecting plagiarism. Word2vec is a deep learning technique that is used to represent words as features of vectors with high precision. The quality of vectors representation depends on the quality of corpus used in training phase. In this paper, we used OSAC corpus for training word2vec model. Moreover cosine similarity measure is used to compute the similarity between words' vectors. The similarity measures show how simple changes in text such as changing one word, or changing the position of verbs and nouns results with similarity value equal to 99% which provide the possibility to detect plagiarism even if the text is altered by replacing words by their synonyms or changing the words order.