Al Etaiwi, Awajan A., Suleiman D. 2019. Keywords Extraction from Arabic Documents Using Centrality Measures. Proceedings of Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). Granada, Spain. October 22-25, 2019. 10.1109/SNAMS.2019.8931808. Pages 237-241.

Abstract

In this paper, we propose a graph-based approach to extract keywords from Arabic documents. We represent the original text as a collocation graph, whose vertices represent original words and the edges represent the co-occurrence relation between all consecutive words. Then, a set of centrality measures are calculated and considered as features set. The vertices in the collocation graph are candidates to be keywords. We demonstrate the effectiveness of the proposed approach on news article documents. Four different machine learning algorithms are used in the experiments including Multi-layer Perceptron, Naïve Bayes, Random Forest, and OneR. The experimental results showed that the performance of the proposed approach achieved good results with comparison to other state-of-the-art results in terms of precision, recall, and F-score.