

ABSTRACT

Text classification is a part of problems in the classification field. Nowadays, Islamic content widely used in text classification research, including Hadith and the Alquran. Since both are become main references for muslim. The aim of this research is to help muslim for learning Islamic law based on the Alquran, since the muslim learn the Alquran for their guideline of life in the world. In addition, non-Muslim also learn the Alquran to expand their insight about Islam. We proposed a word embedding feature based on Tensor Space Model, which used to improve the accuracy of the system. Based on the experiment results and analysis, classification with the average of features of Lesk concatenate with word embedding vector result in a better performance. Furthermore, the best Hamming loss produced in this study was 0.10317 for the prediction of data test.

Keywords: Multilabel Classification; The Alquran; Tensor Space Model; Lesk Algorithm; Word Embedding.