Abstract

Hadith is the second source of Islamic law after Al-Quran, making it important to study. However, there are some difficulties in learning hadith, such as to determine which hadith belongs to the topic of suggestions, prohibitions, and information. This certainly obstructs the hadith learning process, especially for Muslims. Therefore, it is necessary to classify hadiths into the topic of suggestions, prohibitions, information, and a combination of the three topics which also called as multi-label topic. The classification can be done with the K-Nearest Neighbor, it is one of the best methods in the Vector Space Model and is the simplest but quite effective method. However, the KNN has a lack in dealing with high vector dimension, resulting in the long time computing classification. For that reason, it is necessary to classify Sahih Bukhari's Hadiths into the topic of recommendations, prohibitions, and information using the Latent-Semantic Analysis - K-nearest Neighbor (LSA-KNN) method. Binary Relevance method is also employed in this research to process the multi-label data. This research shows that the performance of LSA-KNN is 90.28% with the computation time is 19 minutes 21 seconds and the performance of KNN is 90.23% with the computation time is 37 minutes 06 seconds, which means that the LSA-KNN method has a better performance than KNN.

Keywords: hadith, classification, latent semantic analysis, k-nearest neighbor, binary relevance