

Abstract

Al-Hadith is a collection of words, deeds, provisions, and approvals of Rasulullah Shallallahu 'Alaihi wa Salam that becomes the second fundamental laws of Islam after Al-Qur'an. As a fundamental of Islam, Muslims must learn, memorize, and practice Al-Qur'an and Al-Hadith. One of venerable Imam which was also the narrator of Al-Hadith is Imam Bukhari. He spent over 16 years to compile about 2602 Hadith (without repetition) and over 7000 Hadith with repetition.

Automatic text categorization is a task of developing software tools that able to classify text of hypertext document under pre-defined categories or subject code[9]. The algorithm that would be used is Random Forest, which is a development from Decision Tree.

In this final project research, the author decided to make a system that able to categorize text document that contains Hadith that narrated by Imam Bukhari under several categories such as suggestion, prohibition, and information. As for the evaluation method, K-fold cross validation with F1-Score will be used and the result is 90%.

Keywords: *Automatic Document Categorization, Hadith Sahih Al-Bukhari, Random Forest, K-fold cross validation, micro average f1-score.*