Abstract

Hadith is believed to be the main source of Islam after Qur'an. The simplicity of obtaining hadith information is currently supported by global access using the internet. The abundance of hadith literature sometimes finds difficulties to obtain the information that needed. Therefore, information extraction is required to facilitate the searching of information in hadith. In this study, the name indexing in Indonesian translation of hadith from nine narrators was built. The model was built using Named Entity Recognition with Naïve Bayes classifier. The features used in this study are title case, POS tag and unigram. This study experimented with individual features and features that were combined. Precision, recall, and F1-Score are employed as evaluation metrics. F1-Score is used in this study to measure the performance of named entity and features. The results of experiments extracted 258 people's names from 13870 token data from 100 Indonesian hadith texts and show that implementing the combination of all features can achieve 82.63% of F1-Score.