

Abstract

The narrator's name in the Hadith is one of the most important components in determining the validity of a hadith, but with the large number of Hadiths that exist, causing the process of determining the validity of a Hadith manually becomes difficult, especially in the Indonesian Hadith translation. Named Entity Recognition (NER) is a method that aims to find entities in a text document, in this case the entity includes the name of the person, location, organization, etc. This study will discuss the implementation of the Named Entity Recognition to the Indonesian translation of the Hadith collection to find the names of narrators from each Hadith. In this study 200 Hadiths from 9 different books consisting of 31010 tokens and 2241 narrator name entities will be used as datasets. Because of the variety of entity forms and the amount of data used, this study will use a supervised-learning approach, and to maximize performance from the NER system, Support Vector Machine (SVM) is chosen as a classifier model that is known to have good generalization capabilities in classifying data and ability to deal with high-dimensional data. Some combinations of test scenarios on the NER model showed the highest F-1 results of 0.9 with training data totaling 140 Hadiths consisting of 1564 entities and testing 60 Hadiths consisting of 677 entities. The narrator name produced by the NER system will then be used as an index of the Hadiths that have been narrated by the narrator using the Inverted Index method.