Abstract

Human entity extraction is one of the sub-tasks of Named Entity Recognition or NER which is specifically for identifying human entities or entities that are considered to have the same meaning as humans. NER has been widely used in fields such as business, news, search engines, or in the field of customer service. The Koran is the holy book for followers of Islam which consists of 114 letters. Not many studies have carried out human entity extraction in the Al-Quran, especially for the Indonesian translation of the Al-Quran. Therefore, we need a system that can carry out the task of extracting human entities for the Indonesian translation of the Al-Quran. The system will work by using input in the form of a translation of the Al-Quran in Indonesian and will produce output in the form of words that have been given an entity label. This research will conduct experiments using the Bi-LSTM model and then evaluate the value of the evaluation metrics. The experiment will also carry out a comparison between the Bi-LSTM model and CRF. The research results show a fairly high accuracy value, 0.93. The author's contribution is to create a human entity extraction system using the Bi-LSTM method and then analyze the results of the system that has been built.

Keywords: Al-Quran, human entity extraction, Bi-LSTM, CRF