Abstract

Al-Quran is the holy book of Muslims and consists of letters and verses that are many and have various meanings. With the content of various meanings, it has many entity structures, making it difficult to perform entity recognition manually because it takes a lot of time and effort. Therefore creating a human entity extraction system that can run automatically so as to simplify and speed up the extraction of human entities. Extraction of the human entity itself is one of the tasks of Named Entity Recognition (NER) which focuses on identifying human entities such as names or those that have the same meaning as humans. An example of input and one of named human entity is in Al-Quran Surah Al-Baqarah verse 8 in Arabic which has been transliterated in buckwalter, namely Wa, mina, {l, nˇaAsi, man, yaquwlu, 'aAma, nˇaA, bi, {llˇahi, wa, bi, {lo, yawomi, {lo, 'aAxiri, wa, maA, hum, bi, mu&ominiyna. In this verse these are words of named human entity : man, yaquwlu, 'aAma, nˇaA, bi, {llˇahi, wa, bi, {lo, yawomi, {lo, 'aAxiri. At this time there are still not many who make human entity extraction systems in the Al-Quran. Therefore, the author in this thesis will apply a human entity extraction system with input in the form of Arabic Al-Quran which has been transliterated in buckwalter, as well as an additional experiment that is comparing the performance of two human entity extraction algorithms on a supervised learning basis, namely CRF and BERT. Experiments were carried out, namely comparing the F1-Score values of two human entity extraction models, namely the CRF and BERT models and trying to modify the parameters contained in the BERT model. The result of the experiment is that the CRF model produces an F1-Score value of 0.72 and a BERT of 0.50. After trying to modify the parameters on BERT, it turned out that it was able to increase the F1-Score value. The author's main contribution is to create a system for extracting human entities from the Al-Quran in Arabic which has been transliterated in Buckwalter and analyzing the performance of the algorithm to be used.

Keywords: Al-Quran, Extracting Human Entities, Supervised Learning, Comparative Study