Abstract

The Quran is the primary source of law for Muslims. Quran has a total of 30 juzs, divided into 144 surahs and arranged of 6,236 verses and in the Quran discuss different topics and have many entities too, so someone sometimes has difficulty understanding the Quran. To make it easier to understand the Quran, we can get identification of essential entities in the Quran such as names of the people in the Quran. One way to do it is by extracting information on essential entities in Quran is with Named Entity Recognition (NER). NER automatically recognizes essential entities such as the people's names at the Quran. This paper builds a system for identifying the entity of the people in the Quran using CRF techniques with a multiple choice approach where the system will be introduced with a range of possibilities from the entity and adjusted to an input given to be able to detect entities. On system developing for identifying the entity of the people in Quran with Indonesian Quran translation dataset shows, this test produces an average performance of using the F1 generated at 0.77 for the use of training data as many as 36814 data from 954 verses in the Quran.

Our source code and data are available at: https://github.com/farhandzakyarvianto/peopleERinQuran

Keyword: *Quran*, *Named Entity Recognition*, *Conditional Random Field*