Jurnal Tugas Akhir Fakultas Informatika

Abstract

In Natural Language Processing (NLP), Named Entity Recognition (NER) technology is one part of the NLP method and is widely used such as information extraction, information retrieval, machine translation and question answering systems and others, so this research focuses on information extraction. Named Entity Recognition (NER) has the main purpose of identifying entity names with special meanings in the text, especially personal names, locations, organizations, times and other entities. The source of the data used is the Indonesian language news text which is reproduced manually using several tags, namely personal name, location, organization and time. Therefore, this study uses the Bidirectional LSTM-CRF method. Bidirectional LSTM utilizes pre-context (previous context) and post-context (next context) by processing data from two directions which are then classified using CRF. In this study, there are several processes carried out, namely preprocessing (case folding, filtering, tokenization), labeling, word2vec, training, testing and the final evaluation process. The results of this study indicate that the Bidirectional LSTM-CRF method for the NER system in Indonesian texts obtains f1-score results for organizational entities of 86%. These results are based on three test scenarios, namely setting the word embedding dimension, units and batch size.

Keywords: Named Entity Recognition, Natural Language Processing, Bidirectional LSTM-CRF

