

ABSTRACT

Arabic is a language with rich morphology and has problem where a lexical item can appear as a form with highly inflected forms in the corpus. This large variation can reduce the possibility of finding a single word form and reduce the effectiveness of other tasks in NLP (Natural Language Processing). Therefore, to handling the problem of finding single word form, this study aims to propose a model that can accurately perform word formation in Arabic. The model build using morphological reinflection techniques with an emphasis on the important elements of word formation in Arabic. These elements are type of words, *wazan* (verb-form), and *dhamir* (pronouns). The elements represented by morphological features namely MSD (Morphosyntactic Description). Previous research has success build a model for the reinflection process without *wazan*. In this study *wazan* is an additional feature and an important part in increasing accuracy. The model was built using character-based RNN (Recurrent Neural Network) seq2seq. This model successfully to map words correctly with 92.87% accuracy for task with MSD-source and 90.71% accuracy for task without the MSD-source. These accuracies are 1.78% and 7.91% higher than previous research. It means that this study produces more precise predictions.

Keywords: Arabic, single word form, morphological reinflection, morphosyntactic description, RNN seq2seq.