

ABSTRAK

Bahasa Arab adalah bahasa dengan morfologi yang sangat kaya. Hal tersebut memiliki masalah dimana suatu item lexical dapat muncul dengan variasi bentuk yang sangat tinggi dalam korpus. Variasi yang sangat tinggi tersebut dapat mengurangi kemungkinan ditemukannya bentuk kata tunggal dan dapat mengurangi efektifitas dari pekerjaan lain dalam NLP (Natural Language Processing). Oleh karena itu, penelitian ini bertujuan untuk membangun model yang dapat melakukan pembentukan kata secara akurat. Model dibentuk menggunakan teknik *morphological reinflection* dengan menitik beratkan pada elemen-elemen penting pada pembentukan kata dalam Bahasa Arab. Elemen-elemen tersebut adalah kelas kata, *wazan* (verb-form), dan *dhamir* (kata ganti). Mereka direpresentasikan oleh fitur morfologi yaitu *morphosyntactic description* (MSD). Penelitian sebelumnya telah berhasil membangun model untuk *reinflection process* tanpa *wazan*. Sehingga, pada penelitian ini *wazan* merupakan fitur yang ditambahkan dan merupakan bagian penting dalam meningkatkan akurasi. Model dibangun menggunakan RNN (Recurrent Neural Network) seq2seq berbasis karakter. Model ini berhasil memetakan kata dengan tepat sebanyak 92.87% untuk task dengan input MSD-source dan 90.71% untuk task tanpa MSD-source. Akurasi-akurasi ini 1.78% dan 7.91% lebih tinggi dibandingkan dengan penelitian sebelumnya. Artinya penelitian ini menghasilkan prediksi yg lebih tepat.

Keywords: Bahasa Arab, bentuk kata tunggal, morphological reinflection, morphosyntactic description, RNN seq2seq.