

ABSTRAK

Al-Qur'an adalah sumber dan petunjuk bagi umat Islam. Ini adalah dokumen yang berusia 15 abad yang lalu dan ditulis dengan baik dalam bahasa Arab. Banyak Muslim harus belajar bahasa Arab selain mengetahui artinya. Salah satu aspek terpenting dari bahasa Arab adalah morfologi dan mengidentifikasi deskripsi morfologi kata. Ini disebut dengan tugas analisis morfologis. Tugas ini penting karena dari aspek morfologi suatu kata dapat mengetahui perbedaan bentuk suatu kata, dan dari situ dapat diketahui maknanya. Makalah Gonzales telah berhasil membuat model untuk mengidentifikasi ciri morfologi (MSD) bahasa Arab. kata kerja saja. Penelitian ini akan fokus pada penambahan beberapa jenis kata Arab lainnya, yaitu kata benda. Mencoba menggunakan metode pendekatan state-of-the-art saat ini berbasis saraf dengan jaringan saraf berulang (RNN). RNN dapat menangkap lebih banyak informasi tentang urutan sub-kata seperti prefix, infix, root, dan suffix untuk membuat msdidentifier yang lebih baik. Inputnya adalah satu kata Arab Melalui ekstraksi pola, vektorisasi subkata, identifikasi bentuk kata kerja, identifikasi kata ganti dan jenis kata, dan akhirnya proses identifikasi MSD untuk melihat hasilnya. Studi ini membantu untuk memperbaiki kesenjangan terakhir (penganalisis morfologi pada kata benda POS) dan meningkatkan apa yang dapat diidentifikasi dari pendekatan sebelumnya untuk berkinerja lebih baik. Model ini berhasil mengidentifikasi MSD dengan akurasi 99% dan skor F1 97%. Metode ini juga telah dibandingkan dengan sistem Jabalin Gonzales menggunakan pengujian data sistem ini dan mendapatkan akurasi yang lebih baik dengan 99% dan Jabalin mendapatkan akurasi 39%.

Kata kunci: morphological analyzer, morphosyntactic description, recurrent neural network, Arabic word classification , sub-word vectorizing