

ABSTRAK

Nama : Izki Zakiyah Al-Hamro

NIM : 1167010038

Judul Skripsi : Analisis *Part Of Speech* (POS) *Tagging* Menggunakan Metode HMM *Trigram* Pada Data Al-Qur'an

Part Of Speech (POS) *tagging* merupakan bagian dari *Natural Language Processing* untuk menentukan dengan benar label kata pada suatu kalimat dari input yang diberikan. Teknik POS *tagging* yang berbeda dalam beberapa literatur telah dikembangkan untuk teks bahasa Inggris, dan sedikit untuk teks bahasa Arab. Permasalahan ini menggunakan metode yang didasarkan pada model Markov tersembunyi kedua, yaitu mencari dua kata ke masa lalu atau lebih dikenal dengan sebutan metode HMM *Trigram*. Masalah utama POS *tagging* adalah *Out Of Vocabulary* (OOV) dan ambiguitas kata. Penelitian ini membahas POS *tagging* menggunakan metode HMM *Trigram* pada data teks Al-Qur'an. *Dataset* terbagi menjadi tiga kategori data yang berasal dari *quran corpus* terdiri dari 150 kalimat sempurna sederhana, 50 kalimat dengan S/P/O/K lebih dari satu dan 50 ayat Al-Qur'an pilihan. Percobaan data dilakukan dengan teknik validasi silang yaitu *k-fold cross validation*. Data diklasifikasi menjadi dua, yaitu data latih dan data uji. Data latih digunakan untuk mencari probabilitas emisi dan transisi, sedangkan pengujian data menggunakan algoritma Viterbi. Hasil percobaan mencapai akurasi rata-rata sebesar 86% untuk *dataset* sederhana, 60% untuk *dataset* sedang, dan 38% untuk *dataset* ayat lengkap.

Kata kunci: *Part Of Speech Tagging*, Metode HMM *Trigram*, Algoritma Viterbi

ABSTRACT

Name : Izki Zakiyah Al-Hamro

NIM : 1167010038

Title : *Analysis Part Of Speech (POS) Tagging Using HMM Trigram Method On Al-Qur'an Corpus*

Part Of Speech (POS) tagging is part of Natural Language Processing to determine the word label correctly in a sentence from the input given. The different POS tagging techniques have been developed in some literature for English text, and a little for Arabic text. This problem uses a method based on the second hidden Markov model, which is looking for two words from the past or as known as the HMM Trigram method. The main problems with POS tagging are Out Of Vocabulary (OOV) and word ambiguity. This study discusses POS tagging using the HMM Trigram method on Al-Qur'an text data. The dataset is divided into three categories of data derived from the quran corpus consisting of 150 simple perfect sentences, 50 sentences with more than one S/P/O/K and 50 selected verses of the Qur'an. The data experiment was carried out by using cross validation technique, namely k-fold cross validation. The data is classified into two, those are training data and test data. The training data is used to find the emission and transition probabilities, while the data testing uses the Viterbi algorithm. The experimental results achieved an average accuracy of 86% for the simple dataset, 60% for the medium dataset, and 38% for the complete verses dataset.

Keyword: *Part Of Speech Tagging, HMM Trigram Method, Viterbi Algorithm*