

Abstraksi

Dalam pemakaian *Part-of-Speech Tagging* bahasa Indonesia diperlukan biaya yang sangat besar. Untuk itu diperlukan sebuah metode dimana dalam pengimplementasiannya tidak diperlukannya data training atau teks yang sudah ternotasi. *Unsupervised Hidden Markov Model* memungkinkan sistem untuk melakukan penotasian teks tanpa adanya teks training. *Unsupervised Hidden Markov Model* dibagi menjadi 2 proses utama yaitu evaluation/training dan decoding. Proses evaluation/training menggunakan algoritma Baum-Welch dan forward untuk mencari probabilitas π , probabilitas transisi dan probabilitas emisi yang akan digunakan untuk pencarian notasi kata. Algoritma Baum-Welch memungkinkan sistem untuk mengestimasi probabilitas awal dan algoritma forward berguna untuk mendapatkan probabilitas yang paling optimal. Dalam proses decoding, digunakan algoritma viterbi yang berfungsi untuk memilih jalur state yang paling baik. Pengujian dilakukan untuk mengetahui pengaruh jumlah tag dan jumlah kata terhadap akurasi yang dihasilkan sistem. Dari hasil pengujian dapat diambil kesimpulan bahwa metode *Unsupervised Hidden Markov Model* dapat digunakan untuk kasus *Part-of-Speech Tagging* bahasa Indonesia namun masih menghasilkan akurasi dan konsistensi yang buruk. Jumlah tag yang menghasilkan akurasi terbaik adalah 7 tag, dan tag yang menghasilkan akurasi yang paling konsisten adalah 36 tag. Panjang kalimat tidak terlalu berpengaruh namun memiliki kecenderungan pada jumlah variasi katanya. Rata-rata akurasi terbaik yang didapat hanya sebesar 14,52 %.

Kata kunci : *Part-of-Speech Tagging* bahas Indonesia, *Unsupervised Hidden Markov Model*, Baum-Welch, Viterbi