

Abstract

Part-of-Speech Tagging for Indonesian required a very large cost. It required a method in which the implementation does not need training data or text that has been tagged. Unsupervised Hidden Markov Models allow the system to perform text tagging without any training text. Unsupervised Hidden Markov Model is divided into two main process evaluation / training and decoding. The process of evaluation / training using Baum-Welch algorithm and forward algorithm to get the p probabilities, transition probabilities and emission probabilities that will be used to search for the word notation. Baum-Welch algorithm allows the system to estimate the probability of initial and forward algorithm is useful to obtain the most optimal probability. In the process of decoding, viterbi algorithm is used which to select the best of state sequence. Testing conducted to determine the effect of the number of tags and the number of words as to the accuracy of the system result. From the test results can be concluded that the method Unsupervised Hidden Markov Models can be used for the case of Part-of-Speech Tagging Indonesian but still produced poor accuracy and consistency. The number of tags that produces the best accuracy is 7 tag, and tags that generate the most consistent accuracy is 36 tags. The length of the sentence is not very influential, but has a tendency on the number of word variations. The best average accuracy are obtained only at 14.52%.

Keywords: *Part-of-Speech Tagging* for Indonesian, *Unsupervised Hidden Markov Model*, Baum-Welch, Viterbi