ABSTRACT

This thesis presents the results of the analysis and implementation of applications in building Part-of-Speech (POS) Tagging for Indonesian. In this implementation using POS Tagging Supervised Hidden Markov Model (HMM) where there are bernotasi training data so that the resulting optimal parameters. In building applications using POS Tagging Supervised HMM method, there are some problems, namely the handling of Out-of-Vocabulary (OOV) and transition probabilitas value equal to zero. The second problem is the focus of making this final task. To handle the transition probability value of zero is used Jelinek-Mercer Smoothing techniques (JMS) that will minimize or eliminate the value of zero on the transition probabilities. And to issues OOV Tree Affix the writer used to treat it, is associated with an increase in POS tagging accuracy parameters using the Supervised HMM. After doing research, the number does not affect the accuracy of OOV percentage generated by the system. Highly visible affect accuracy is JMS implementation, so that the accuracy produced by the system is higher than the results generated without implementing JMS. Beside that, the first N characters also affect accuracy generated by the system.

Keys : POS Tagging, JMS, Affix Tree, Supervised HMM, OOV