

Abstract

This final project explains about analysis of *n-gram* method to recognize text based tribe languages from Indonesia. To see the accurateness from *n-gram* method, we can test with using different length of *n* from *n-gram* and different length of rank from *n-gram* language model to recognize tribe languages. The test also has been done to see how many words in document which we want to recognize that influence the accuracy of recognition. The system was built with *n-gram* method for modeling the tribe languages, and rank-order-statistic for classification. From the testing was had done, for accurateness tribe languages recognition we can use the minimum length of *rank=100* and the length *n* from *n-gram* which can be *n=3*, *n=4*, *n=5* with using *rank=100* the accuration is about 100% for *n=3*, 98,75% for *n=4*, 97,50% for *n=5*, and for ratio between the length of rank from *n-gram* language model and how many words in document which we want to recognize is *length of rank : number of words = 100 : 40*, with using minimum *length of rank=100* and minimum *number of words=40*.

Keywords: n-gram, performance, accuracy, rank-order-statistic.