ABSTRACT

Stemming is the process to find the base of a word. By removing all the good feed supplement consisting of the prefix (affixes), suffixes (suffixes) and confixes (a combination of prefix and suffix) on the word derivative. Stemming technique is different for each language. This is because the structure of words in each language has different rules of formation. Like stemming for English-language texts will be different from the Indonesian-language text.

From the implementation and testing results showed that the stemming algorithm Paice / Husk can be implemented in Indonesian language with accuracy and good power.

In analyzing the effect of stemming algorithms Paice / Husk in the process of text mining, categorization process is carried out using multinominal Naïve Bayes as a classifier, it was found that the process of stemming improve categorization process both in terms of accuracy, processing time, and F-measure when compared with the unstemmed documents. This is because the stemming process reduces the number of unique terms in the test document.

Key words : stemming, Text Mining Preprocessing, morphology of Indonesian Language, text categorization.