Abstract

Stemming is the process of forming the base of the words that have been getting modifications in its use. The use of words contained in the sentence structure of which have got Affixes include prefixes, suffixes, or infixes. Stemming is part od the Preprocessing, which is the last phase after Tokenization and Stoplist Removal. Stemming process is different in each language because the word has a different formation in each language. In Indonesia, there are several algorithms that can be used in the process of stemming, such like Arifin-Setiono Algorithm, Nazief-Adriani Algorithm and Enhanced Confix Stripping Stemmer (ECS) Stemmer. ECS algorithm is a refinement of the Confix Stripping Stemmer algorithm.

This final project will be presented and the implementations of ECS Stemmer algorithm and its modifications to the Indonesian language text. Enhanced Confix Stripping Stemmer Algorithm have drawbacks and limitations in dealing with infixes and few letter in th end part of word that such as suffix. Modifications of Enhanced Confix Stripping Stemmeris designed to repair the weakness. From the test result will be seen the difference between the accuracy of the Enhanced Confix Stripping Stemmer algorithm and its modification, it will be proven that the scheme has been modified to produce a higher accuracy value.

Keyword : *Stemming, Enhanced Confix Stripping Stemmer, Affixes, Preprocessing.*