

Abstract

The developing technology in cyber world has made the number of article as a part of information increased. Therefore, a method of articles is needed to ease the reader in seeking information by applying a functionality of data mining, which is categorization. To obtain a good categorization result, a good data preprocessing stage is also needed. The generally used data preprocessing stage is stemming.

Stemming is a process to obtain root word by separating all affixes that are attached on that word. This stemming process will be able to reduce dimension of the data in categorization process so that it can be included in one stem. There are several methods of stemming according to the process, one of them is affix removal. This final assignment will mention two methods of affix removal, which are Porter Stemmer and Krovetz Stemmer, as well as their effect on categorization process. Stemmer performance is calculated by accuracy and ICF (Index Compression Factor).

Based on testing result, Improved Porter Stemmer has better accuracy and ICF score than Porter Stemmer dan Krovetz Stemmer. However, from the best stemmer performance acquired by Improved Porter Stemmer does not absolutely increase the score of precision and recall in categorization.

Keywords: *stemming, affix removal, Porter Stemmer, Krovetz Stemmer, Index Compression Factor, accuracy.*