Abstract

Stemming is a process to transform affixed words to be the root using specific rules. There are many algorithms that are used to perform stemming, especially Indonesian stemming consists of Nazief & Andriani algorithm, Yussof & Sembok algorithm, Idris & Mustofa algorithm, Vega algorithm, Arifin & Setiono algorithm, Indonesian Porter algorithm, Confix Stripping (CS) algorithm, and Enhanced Confix Stripping (ECS) algorithm. This research compares the performance of these Indonesian stemming algorithms to find the algorithm with the highest level accuracy. Evaluation of this research conduce the explicit and implicit evaluation. Explicit evaluation is done to measure and analyze directly the performance comparison from every algorithm. This evaluation uses accuracy parameter such as Word Conflation Class and Index Compression Factor. Then the implicit evaluation is done by implementing each of those stemming algorithms into the Indonesian text classification using Weka tools and Support Vector Machine (SVM) as the classifier. The result of that SVM is accuracy score and F1-Measure that becomes the performance comparison from each of the algorithms. The result of this research shows that Nazief & Andriani algorithm has the high score of explicit and implicit evaluation than the other Indonesian stemming algorithms.

Keyword : preprocessing, Indonesian stemming algorithms, text classification, explicit evaluation, implicit evaluation.