Abstract

Language is a means of communication between people to understand each other. The diversity of language causes less structured pattern language. The structure of the language can be recognized from the classes of each words. The induction process of word class is known as POS tagging. Grouping the words in induction process of word class based on proximity between the words. SVD method and X-means *cluster*ing can be used to group words.

SVD method is used to model the proximity problem between the word into a matrix of descriptors. Matrix descriptor contains the value of the left context and right context of each word in the corpus. Matrix descriptors used for *cluster*ing process with X-means *cluster*ing method. X-means *cluster*ing group the words according to the *cluster* that will continue to form until upper bound that has been set. Quality of *cluster*s formed are tested using Silhouette Coefficient.

Combination of SVD and X-means *cluster*ing produces better quality *cluster*s to the structure of the medium using fixed *centroid* initialization. Parameter of Top N1, RR1, Top N2, and RR2 in the method affects the obtained of SC. The higher the difference between Top N and reduced rank make the SC is better and the quality of the *cluster*s that formed are better too.

Keywords: *Part of speech tagging*, SVD, *X-means clustering*, *Silhouette Coefficient*.