

Abstract

Clustering is the grouping of data without the process of label into groups - groups or clusters, that have a quality parameter values formed clusters (cluster purity), and before doing to create clusters, in the clustering have been doing the data management process into a collection of text is called text mining . In text mining there are two types of words : word groups that have affix and group of words that have been lost the affix. To perform basic word processing, required a process called stemming. Affix removal stemming algorithm is one stemming algorithm that can be applied in the Indonesian language text data, as text data can stemming with values reaching 90% accuracy. In the case study translated text clustering verses the Qur'an, proved that the effect of stemming on the text mining application, able to reduce the amount of text data will be processed on clustering and cluster purity capable of delivering value better than the clustering does not apply stemming.

Keywords: *Clustering, Cluster, Text Mining, Affix Removal Stemming, Purity Cluster, Al-Qur'an.*