

ABSTRAK

KLASTERISASI TERJEMAHAN AYAT-AYAT AL-QURAN

MENGGUNAKAN ALGORITMA K-MEANS++

ANISA KHAERANI - 1167050027

Jurusan Teknik Informatika

Clustering merupakan sebuah proses untuk melakukan pengelompokan data kedalam beberapa *cluster*, sehingga data didalam *cluster* tersebut memiliki karakteristik yang sama dan juga memiliki karakteristik yang berbeda pada antar *cluster*. Adapun algoritma untuk melakukan pengelompokan ayat-ayat al-quran yang mempunyai kemiripan tema yang sama kedalam beberapa *cluster* adalah menggunakan algoritma K-means++ *clustering*. Bedanya dengan algoritma sebelumnya atau k-means biasa adalah pada akurasinya, pada algoritma k-means penentuan awal clusternya dilakukan secara acak sehingga menghasilkan akurasi cluster yang kurang maksimal, sedangkan pada k-means++ penentuan awal clusternya menggunakan perhitungan probabilitas. Adapun hasil dari penelitian ini adalah bahwa k-means++ mampu mengelompokan terjemahan dari ayat-ayat Al-quran kedalam 7 kelompok cluster. Dengan cluster pertama berjumlah 600 ayat, cluster ke 2 715 ayat, cluster ke 3 997 ayat cluster ke 4 943 ayat, cluster ke 5 1441 ayat, cluster ke 6 693 ayat dan cluster ke 7 847 ayat. penentuan clusternya dihitung menggunakan Silhouette coefficient metode ini berfungsi untuk menguji kualitas dari cluster yang dihasilkan.

Kata kunci: *clustering*, *text mining*, *K-means*, *k-means++*,

ABSTRACT

Clustering is a process for grouping data into several clusters, so that the data within the cluster have the same characteristics and also have different characteristics between clusters. The algorithm for grouping Qur'anic verses that have the same theme into several clusters is using the K-means ++ clustering algorithm. The difference with the previous algorithm or ordinary k-means is in its accuracy, in the k-means algorithm, the initial determination of the clusters is done randomly, resulting in less than optimal cluster accuracy, while in k-means ++ the initial determination of the cluster uses probability calculations. The result of this research is that k-means ++ is able to classify the translation of the verses of the Koran into 7 cluster groups. With the first cluster totaling 600 paragraphs, the 2nd cluster 715 paragraphs, the 3rd cluster 997 paragraphs 4 943 paragraphs, the 5th 1441th clusters, the 6th 693th clusters and the 7th 847th clusters. The determination of the clusters is calculated using the Silhouette coefficient. This method is used to test the quality of the resulting clusters.

Keywords: clustering, text mining, k-means++, k-means