

Clustering topic groups of documents using k-means algorithm: Australian embassy Jakarta media releases

Wishnu Hardi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920549784&lokasi=lokal>

Abstrak

The Australian Embassy in Jakarta stores a wide array of media release document. Analyzing particular and vital patterns of the documents collection is imperative as it may result new insights and knowledge of significant topic groups of the documents. K-Means algorithm was used as a non-hierarchical clustering method which partitioning data objects into clusters. The method works through minimizing data variation within clusters and maximizing data variation between clusters. Of the documents issued between 2006 and 2016, 839 documents were examined in order to determine term frequencies and generate clusters. Evaluation was conducted by nominating an expert to validate the cluster result. The result showed that there were 57 meaningful terms grouped into 3 clusters. “People to people links”, “economic cooperation”, and “human development” were chosen to represent topics of the Australian Embassy Jakarta media releases from 2006 to 2016. Text mining can be used to cluster topic groups of documents. It provides a more systematic clustering process as the text analysis is conducted through a number of stages with specifically set parameters.