

Abstract

The development of large numbers of information like news articles are available on the internet. Hence text clustering is needed by applying clusterisation as one of *data mining* task. Nowadays, the method of text clustering still do not really address the special problem of text clustering such as the high dimensionality of the data and very large size of the database, therefore understandability of the cluster description still difficult to understand.

This application method can reduce the high dimensionality of the data and very large size of the database. There is some methods that can be used based on frequent term-based text clustering, such as hierarchical frequent term-based clustering (HFTC) and frequent term-based clustering (FTC).

The clusterisation's output that based on frequent term-based text clustering has the understandability of the cluster description. Based on experimental evaluation, it can be concluded on HFTC, f-measure value increasing while minsup decreasing thus the quality of cluster is better, on each minimum support, FTC has variation Entropy value, and the time to make cluster is decreasing while minimum support is increasing.

Keywords: *clusterisation, frequent term-based text clustering, HFTC, FTC, f-measure, Entropy*