

Abstract

Text documents are growing rapidly on the internet, digital libraries, and articles. The information on the web, especially news articles, is useful for users. A lot of news articles are available on the internet. So, it is hard for users to get the news articles that they want. Hence, news articles categorization based on information content is needed. So, it can be categorized based on the topic.

Documents/articles clustering is one of the ways to mine the information contents of documents/articles. Some related documents are located in the same cluster by clustering. A document can have high dimension and huge volume of data. For that reason, a technique is needed to handle it.

Generally, clustering algorithms focus on one-way clustering, for example, clustering based on words distribution or based on document distribution. Therefore, co-clustering, which makes clusters simultaneously on both of table dimensions, is developed to reduce the dimension effectively and efficiently.

Keywords : co-clustering, information theory, mutual information