

Malay documents clustering algorithm based on singular value decomposition.

ABSTRACT

Document categorization is a widely researched area of information retrieval. A research on Malay natural language processing has been done up to the level of retrieving documents but not to the extent of automatic semantic categorization. Thus, an approach for the clustering of Malay documents based on semantic relations between words is proposed in this paper. The method described in this paper uses Singular Value Decomposition (SVD) technique for the vector representation of each document where familiar clustering techniques can be applied in this space. The experimental results we obtained taking into account the semantics of the document that performed good document clustering by obtaining relevant subjects appearing in a cluster.

Keyword: Singular Value Decomposition (SVD); Latent Semantic Indexing (LSI); Document clustering; Malay natural language processing.