

Abstract

Nowadays, Electronic news are most popular and interactive information media. Because of that, it grows fast. It is proved by the increasing of company sites and personal sites which is mean that more information and data. This fast increasing is paced by the increasing in using internet recently than before. As the effect, the volume of informations getting bigger and increase exponentially.

The number of data should be serve many advantages. Clustering is one of method to group documents by founding the document relations. Nowadays, the clustering methods mostly rely on term based similarity but ignore another aspect , such as Vector Space Model. On this final project, documents are grouped by *Document Index Graph* which combine term based similarity and phrase based similarity.

This method is tested by using Indonesian news article from world wide web. An exact selected similarity blend factor and similarity threshold will improve cluster's quality. The result of clustering will evaluate according precision and recall value.

Keywords: *clustering, Document, Index , Graph, similarity.*