Abstract

Text summarization is one of text mining tasks. Text summarization is a computerized process of distilling the most important information of a source (or sources) for making a brief version of text (texts) to fulfill user's need or any tasks required.

Centroid-based summarization uses the centroid of a cluster to identify the salient sentences. A centroid is a set of words that are statically important to a cluster of documents. In this paper, the centroid is modified by using concepts as the centroid. A concept is a term that has a semantic role in the sentence. The concept is counted using statistical concept based mining model. As the input for this application are Indonesia and English documents. Centroid value, position value, title overlap value and redundancy penalty are the *feature*s extraction to determine the highest score among the sentences. Centroid value counts the centroid value of all words in the sentence. Similarity with tittle measured how similar the sentences with its document's tittle. The process produces the output in the form of extractive summary that consists of high ranked sentences.

Evaluation of the summaries uses ROUGE evaluation toolkit comparing the result from another summarizer, MEAD. The result of this experiment shows centroid based with concept as the centroid tends to have lower score than MEAD.

Key words: Text summarization, centroid-based, concept, *feature* extraction