

Abstract

The automatic text summarization is the process of taking the most important information from a text or some text to create a brief version of that text to fulfill user's need or any tasks required using a computer-based application. On this final project implemented an centroid based, stemming and stopword process. The centroid of cluster to identify the salient sentences, a centroid is a set of words that are statically important to a cluster of documents. Stemming and stopword process is also used to determine affect the accuracy of the summary. Centroid value, position value, title overlap value and redundancy penalty are the features extraction to determine the highest score among the sentences. Centroid value counts the centroid value of all words in the sentence. Similarity with title measured how similar the sentences with its document's title. 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. Evaluation results show that the method centroid based using preprocessing of Stemming and stopword process affect the accuracy of the summary generated.

Key words: Text summarization, stemming, stopword, features extraction, centroid based, concept, ROUGE evaluation toolkit.