

ABSTRACT

Multi Document Summarization is a method to generate single summary from several documents which have similarities or interconnected. This method is helpful in solving the problem of overload information, because it generates a document summary of the documents to be read instead of having to read the many documents, without losing the main sentences from the documents.

At this Final Project, Multi Document Summarization is implemented by using Agglomerative Clustering and Sentence Specific Feature Extraction. In the implementation, the testing scenario is did by inputting 5 documents, 15 documents and 30 documents. It is figured that clustering using agglomerative clustering for 5 documents and 15 documents can generate a cluster which has best cluster variance in about 70% termination condition. Whereas clustering using agglomerative clustering for 30 documents can generate cluster of documents which has best cluster variance in about 40% termination condition. By looking through all testing, it is found that clustering for 5 documents and 15 documents generate a better cluster than clustering for 30 documents base on cluster variance value. Testing for the result of summarization is did with 10%, 20% and 30% compression rate. By using ROUGE Evaluation toolkit, the best summarize generated on 30% compression rate base on F-Measure value.

Key Words : clustering, document summarization, compression rate, sentence specific feature