

Abstract

The text summarization is the process of taking the most important information from a text document or some text documents to create a brief version of that text using a computer-based application.

On this final assessment implemented multi document summarization based on clustering and nonnegative matrix factorization. The summarization produce extractive summaries consist of important sentences from the documents. NMF decomposes a sentence into combination of semantic feature. It can improve the quality of document summaries, because by using similarity between semantic feature and document topic, it can produce important sentences which have high relevance with the document topic. In the proposed method there is also clustering process to remove noise sentence and redundancy.

Evaluation of the summaries uses ROUGE evaluation toolkit. The test results show that summarization using NMF and clustering can produce summaries with higher precision score than summarization using NMF without clustering. NMF with clustering also produce summaries with fewer redundancy.

Keywords: text summarization, nonnegative matrix factorization, clustering.