Abstract

Automated text summarization based on query is a process of taking the most important information of some text based on the query entered by user, aimed to create shorter version of the text using computer-based applications.

In this final project will be implemented a technique of Automatic Text Summarization Based on Query based on Graph approach for multi-document. This process will produces output in form of extractive summary sentences. The method for get summary is to use the calculation of similarity values and the centrality and making of a closure graph on a selected node. The centrality method that used in this final project is the Okapi Weighting method, which will considers the interests of a node based on queries entered by users. The similarity method used in this final project is IDF modified cosine similarity, which is inherited directly from LexRank. Enumeration methods which are applied, Top1 enumeration, Multiresults enumeration and Modification enumeration. The selected Closure Graph is the shortest path, and will be extracted as summary.

Evaluation of the summary will use ROUGE evaluation toolkit. The result from using enumeration methods shows that the method of Multiresults enumeration using Okapi Weighting and IDF modified cosine similarity has better accuracy than the Top1 enumeration. Modification Enumeration has the best result among the others method which is implemented in this final project.

Keywords: multi document, enumeration, text summarization, query, stopword, okapi weighting, idf modified cosine similarity