

Abstract

Evaluation of summaries is an important problem in development of automatic text summarization (ATS). Evaluation is a process that purpose to assess a automatic text summarization based on its output.

In this assignment implemented method variation ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) which is a method for evaluating automatic Text Summarization system (ATS), that is method of ROUGE-NW representing merger of method of ROUGE-1(N-GRAM 1) with method of ROUGE-W. ROUGE is an evaluation technique used to determine the quality of an automatic summary by comparing the summary produced by the ATS with another summary (ideal) created by human[5]. Moreover, as comparator applied also original method of ROUGE that is ROUGE-1, ROUGE-2, and ROUGE-W.

Evaluation is done by looking at the correlation or the association of score generated by each assessment method ROUGE with the produced by human (human judgment). The result of this experiment shows that the computation of ROUGE-2 and ROUGE-NW has more correlation with human judgment rather than ROUGE-1 and ROUGE-W.

Keywords : evaluation, ROUGE method, n-gram(s), correlation