

ABSTRACT

This research aims to build a website-based Automatic Text Summarization application by using web scraping techniques on news from news.detik.com. Extractive method with TF-IDF algorithm is used to summarize the text efficiently, focusing on sentence weighting and selection based on the highest value. The results are measured using the Recall-Oriented Understudy for gisting Evaluation (ROUGE) evaluation method with four variants: ROUGE-1, ROUGE-2, ROUGE-3, and ROUGE-L. In the research phase, text extraction through web scraping is followed by text preprocessing, text processing, and the arrangement of selected sentences based on their original order. The average ROUGE evaluation results on the given news case study and evaluated with ten human gold standard references show good performance on unigram information (ROUGE-1), with a recall value of 40.1%, precision of 50.6%, and F-Score of 42.9%. However, the performance decreases on bigram information (ROUGE-2) with a recall value of 21.8%, precision 32.1%, trigram (ROUGE-3) with a recall value of 16.6%, precision 27.5%, f-score 20.2%. ROUGE-L shows a balance between recall and precision, with a recall value of 30.8%, precision of 40.8%, and F-Score of 33.6%, indicating the system's ability to capture overall information. On the other hand. In conclusion, although the system is effective in capturing unigram information, its performance degrades on longer bigrams and n-grams.

Keywords: Text summarization, word weighting, news website, tf-idf