ABSTRACT

FAKHIR MAHATHIR. Plagiarism Detection System on Indonesian Language Text Documents Using ROUGE-N, ROUGE-L, and ROUGE-W. Under direction of AHMAD RIDHA.

Plagiarism is a serious problem in education. This research uses ROUGE-N (N = 3 or trigrams), ROUGE-L, and ROUGE-W (with weighted function \( f(x^2) \)) at the sentence level to detect plagiarism on Indonesian language text documents. This research aims to obtain a suitable preprocessing for each method of assessment. The preprocessing includes stopword removal and stemming. This research uses clipping based on recall, precision, and f-measure. Analysis is restricted to preprocessing and calculation method used in each assessment method. Stemming improves ROUGE-N, while stopword removal negatively effects ROUGE-N. ROUGE-L and ROUGE-W perform well with f-measure clipping. ROUGE-W is better without stemming.

Keyword: Recall-Oriented Understudy for Gisting Evaluation (ROUGE), plagiarism detection, document similarity, information retrieval, Naïve Bayes Classifier.