

Abstract

This research identifies plagiarism between two documents (.txt). The method used in this research is Text Alignment and then processed with N-Gram characters and Winnowing Algorithm. This research processed two documents in the form of Source Document and Suspicious Document, then do the calculation using Confusion Matrix to get Accuracy, Recall, Precision, and F1-Score results. The Winnowing algorithm alone has advantages compared to other Fingerprint document algorithms such as the Manber algorithm and the Rabin-karp algorithm.

Plagiarism is one of the most common problems. For now there are still few who use technology to solve this problem. The others mostly use visual. The result of this research is confusion matrix, where some scenarios are done to show the performance of each scenario. From the research conducted, the results obtained research for scenario 1 F1-score of 0.4624 (46.24%) and scenario 2 F1-score of 0.4719 (47.19%).

Keywords: plagiarism, winnowing, text alignment, n-gram, confusion matrix