

ANALYSIS AND IMPLEMENTATION OF PARAPHRASE IDENTIFICATION ON TWEET USING SMITH WATERMAN DISTANCE ALGORITHM

Rino Sumarno¹, Moch Arif Bijaksana²

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung
¹rhino.soemarno@gmail.com, ²arifbijaksana@gmail.com

Abstract

In this study, paraphrase identification is done on two sentences (tweet). The method used to identify these two sentences is Smith Waterman Distance algorithm. In identifying the sentence paraphrase two sentences Smith Waterman Distance will identify the similarity of the two sentences by looking for the greatest value of local similarities.

Confusion matrix is one method to measure the performance of a classification method, and in this research the method of classification that will be used is Smith Waterman Distance algorithm which is useful to determine the similarity or measurement of similarity level of a sentence with another sentence by looking at the similarity of content (local similarities) of both sentences. From the results of this study, obtained an accuracy of 0.609; F1 of 0.480; recall of 0.405 and precision of 0.590.

Keyword : *Paraphrase, Smith Waterman Distance, Tweet, Confusion Matrix.*