

Analisis dan Implementasi Identifikasi Parafrase *Tweet* Menggunakan Algoritma *Smith Waterman Distance*

Rino Sumarno¹, Moch Arif Bijaksana²

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

Abstrak

Dalam penelitian ini dilakukan identifikasi parafrase pada dua buah kalimat (*tweet*). Metode yang digunakan untuk mengidentifikasi dua buah kalimat ini yaitu algoritma *Smith Waterman Distance*. Dalam mengidentifikasi kalimat parafrase dua buah kalimat *Smith Waterman Distance* akan mengidentifikasi kemiripan kedua buah kalimat tersebut dengan mencari nilai terbesar dari local similaritiesnya.

Confusion matrix merupakan salah satu metode untuk mengukur kinerja suatu metode klasifikasi, dan pada penelitian kali ini metode klasifikasi yang akan digunakan adalah algoritma *Smith Waterman Distance* yang dimana berguna untuk menentukan kesamaan atau pengukuran tingkat kemiripan sebuah kalimat dengan kalimat lain dengan cara melihat kesamaan isi (*local similarities*) dari kedua kalimat tersebut. Dari hasil penelitian ini, diperoleh akurasi sebesar 0.609; F1 sebesar 0.480; recall sebesar 0.405 dan precision sebesar 0.590.

Kata kunci : *Paraphrase, Smith Waterman Distance, Tweet, Confusion Matrix.*

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.*