

Abstrak

Perkembangan teknologi semakin canggih dengan adanya internet. Internet yang dapat dengan mudah diakses untuk mencari informasi dan dokumen dapat memicu adanya tindak plagiarisme. Setiap dokumen yang melakukan tindakan plagiarisme akan susah dikenali tanpa adanya sistem yang dapat mengenali kesamaan dokumen. Sistem yang dapat mendeteksi plagiarisme dengan mencari kemiripan pada dokumen dibutuhkan. Dalam penelitian ini digunakan algoritma *Smith-Waterman* untuk mengidentifikasi kesamaan yang paling signifikan (*local alignment*) dari dua buah rangkaian *sequence string* proses (rangkaian/susunan) *nucleotide* atau *protein sequences* sehingga kemiripan antara dua dokumen tersebut dapat terlihat. Algoritma *Nazief & Andriani* merupakan algoritma *stemming* pada *text-preprocessing* sebagai algoritma pendukung dalam proses penentuan nilai kemiripan dokumen teks. Hasil akhir yang didapatkan pada penelitian ini adalah perbandingan dua *sequence* dengan bantuan *preprocessing* memiliki tingkat perhitungan *similarity* yang lebih besar dalam mendeteksi kemiripan dokumen. Dari rata-rata pengujian pada dokumen asli dan dokumen uji menunjukkan tingkat kemiripan yang dihasilkan lebih dari 50%, yang berarti hasil tersebut dapat dinyatakan terjadi tindak plagiarisme.

Kata Kunci: Dokumen, Plagiarisme, Algoritma *Smith-Waterman*, Algoritma *Nazief & Andriani*