

Abstrak

Mudahnya menyusun suatu karya dari sebuah dokumen teks, memicu tindak plagiat yang seharusnya dihindari menjadi semakin mudah dilakukan. Plagiarisme dalam dokumen teks sangat luas cakupannya. Salah satu yang bisa diindikasikan adalah jika terdapat kesamaan isi, lebih jauh lagi kesamaan unsur pembangunnya, yaitu sebuah kata yang merupakan unsur terkecil pembangun dokumen teks. Dalam bidang bioinformatika, terdapat algoritma Smith-Waterman yang digunakan untuk mengetahui kemiripan yang signifikan (*local alignment*) dari suatu sekuens gen makhluk hidup dengan cara *sequence alignment* (penyejajaran sekuens) yaitu menyusun pasangan rantai basa hidrogen pembentuk sekuens tersebut dengan sekuens lain berdasarkan kesamaan strukturnya [14].

Dengan menerapkan konsep algoritma Smith-Waterman, suatu kalimat dalam dokumen teks akan diibaratkan sebagai sebuah sekuens gen yang dibentuk oleh rangkaian kata dan akan dibandingkan dengan kalimat di dokumen teks lain. Sebelumnya, dokumen teks akan melalui tahapan penghilangan *noise*. Setelah itu diproses menggunakan algoritma Smith-Waterman sehingga diperoleh suatu prosentase kesamaan dokumen yang menandakan seberapa besar kemiripan dokumen tersebut dengan dokumen lain.

Kata kunci : plagiarisme, dokumen teks, penyejajaran sekuens, algoritma Smith-Waterman.