

Abstrak

Untuk mendeteksi tindakan plagiarisme telah banyak dikembangkan berbagai cara dan algoritma, khususnya pada dokumen teks. Algoritma yang paling populer digunakan oleh para peneliti adalah jenis algoritma *local string matching* yang dapat memberikan hasil deteksi yang cukup akurat, seperti *Longest Common Subsequence*, *Smith Waterman*, dan *Edit Distance*. Namun, kendala yang muncul pada metode tersebut adalah semakin banyak dokumen diperiksa, maka waktu yang dibutuhkan akan semakin lama. Hal tersebut disebabkan karena satu buah dokumen harus dibandingkan dengan seluruh dokumen yang ada. Oleh karena itu, dibutuhkan pengembangan sistem yang dapat mengurangi jumlah dokumen yang dideteksi, terutama dokumen yang tidak relevan. Untuk mencapai tujuan di atas diusulkan metode untuk memberikan nilai kemiripan antar dokumen berdasarkan *term frequency*. Untuk mempercepat proses tersebut, seluruh dokumen pembandingan diindeks berdasarkan *term*-nya. Pengujian dilakukan terhadap 3 buah rancangan sistem. Dengan menggunakan 50 dokumen *query* yang diuji terhadap lebih dari 10000 dokumen, sistem diukur dari akurasi dan kecepatan pemrosesannya. Dari hasil pengujian, didapatkan sistem dengan gabungan antara *winnowing*, indeks, dan LCS yang memberikan akurasi dan kecepatan yang terbaik. Akurasi yang didapat mencapai 89,78% sama dengan sistem sebelumnya (*winnowing* dan LCS) serta mengurangi waktu yang dibutuhkan untuk pemrosesan hingga 10 kali lipat.

Kata Kunci: Deteksi Plagiarisme, *AVL Tree*, *Term Frequency*, *Information Retrieval*, *Longest Common Subsequence*, *Winnowing*