

Abstrak

Plagiarisme merupakan perbuatan secara sengaja atau tidak sengaja dalam memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya ilmiah orang lain, tanpa menyatakan sumber. Tindakan plagiarisme ini sangatlah buruk, tindakan ini tidak hanya memberikan dampak negatif ke orang yang dijiplak namun terhadap orang yang melakukan plagiat juga. Oleh karena itu, sudah banyak diciptakan algoritma untuk mendeteksi adanya kejadian plagiarisme. Beberapa contoh algoritma pendeteksian plagiarisme, diantaranya *Longest Common Subsequence (LCS)*, *Edit Distance*, *Fingerprinting Document*, dan *Winnowing*.

Pada algoritma LCS, bila diterapkan pada dokumen D1 dengan jumlah token m , dan dokumen D2 dengan jumlah token n , maka kompleksitas pencocokan mencapai $O(m*n)$. Kompleksitas ini membutuhkan waktu pemrosesan lebih lama apabila dokumen harus dicocokkan pada *corpus* yang besar. Kompleksitas waktu ini dapat diatasi dengan mengurangi jumlah token yang dibandingkan. Pada tugas akhir ini dibangun sistem pendeteksian plagiarisme menggunakan metode *Co-Occurrence Statistical Information* dan LCS. *Fingerprint* yang mengandung kata kunci hasil ekstraksi dokumen dengan metode *Co-Occurrence Statistical Information* dibandingkan dengan *fingerprint* dokumen lain menggunakan algoritma LCS.

Proses pendeteksian menggunakan metode *Co-Occurrence Statistical Information* dan LCS, memerlukan waktu komputasi yang lebih cepat dibandingkan hanya dengan LCS saja, sedangkan untuk nilai *similarity* yang dihasilkan relatif sama.

Kata Kunci: Plagiarisme, *Fingerprint*, *Co-Occurrence Statistical Information*, *Longest Common Subsequence*.