

Rina Amelia. Koefisien Jaccard dan Stemming Nazief dan Adriani dalam Pendeteksian Kemiripan Dokumen. Di bawah bimbingan Dr. Awang Hendrianto Pratomo, S.T. M.T. dan Wilis Kaswidjanti, S.Si., M.Kom.

ABSTRAK

Plagiarisme adalah tindakan menjiplak karya orang lain dan mengaku sebagai hasil karya pribadinya. Saat ini sudah banyak algoritma yang membahas cara mendeteksi plagiarisme dokumen teks seperti *Cosine*, *Levenshtein Distance*, *winnowing* dan *Jaccard coefficient*. Hasil penelitian sebelumnya menyatakan bahwa metode Koefisien Jaccard memiliki keakurasian yang rendah, sehingga pada penelitian ini dilakukan pengembangan dari Algoritma Koefisien Jaccard. Proses pendeteksian kemiripan dokumen pertama-tama dilakukan pendekatan n-grams dalam membentuk fingerprint dokumen. Untuk meningkatkan akurasi metode Koefisien Jaccard dalam pendeteksian kemiripan dokumen dilakukan *preprocessing* Algoritma Stemming Nazief & Adriani.

Dari uji coba yang dilakukan penambahan *preprocessing* yaitu *stemming* mempengaruhi nilai keakurasian dan waktu proses. Penggunaan *stemming* Algoritma Nazief & Adriani pada dokumen pasif-aktif memberikan nilai keakurasian yang lebih baik sebesar 48,44% namun memberikan efek terhadap lamanya proses pendeteksianmeskipun menambah waktu untuk pemrosesan.

Kata kunci: plagiarisme dokumen, Koefisien Jaccard, Winnowing, *stemming*