

ABSTRAKSI

Plagiarisme berasal dari bahasa latin yaitu plagiarus, yang artinya pengambilan karya orang lain, berupa tulisan, makalah, atau pendapat dan sebagainya, untuk diakui sebagai hasil karya si pelaku. Praktek ini sudah sering terjadi khususnya pada kalangan akademisi. Untuk mengantisipasinya, dibutuhkan suatu cara yang dapat menganalisis teknik-teknik plagiat yang dilakukan. Ada beberapa pendekatan atau teknik yang dapat diambil, salah satunya adalah dengan pencocokan *string/term*. Untuk pencocokan *string/term* terdapat beberapa algoritma yang ada, salah satunya adalah algoritma pencarian *string* Rabin-Karp. Algoritma Rabin-Karp menggunakan fungsi *hash* dalam proses *string matching*. Fungsi hash inilah yang menjadi kunci algoritma Rabin-Karp. Algoritma Rabin-Karp sangat cocok untuk diterapkan dalam pendeteksian plagiarisme dokumen, karena algoritma ini cocok untuk pola pencarian jamak(*multiple pattern search*). Untuk melakukan deteksi plagiarisme dokumen, khususnya dokumen berbahasa Indonesia, perlu adanya proses *preprocessing* sebelum suatu dokumen siap untuk masuk ke dalam proses inti yaitu *string matching*. Di dalam proses *preprocessing* terdapat tahapan yaitu *case folding*, *tokenization*, *number removal*, *filtering* dan terakhir yaitu *stemming* atau pengembalian suatu kata ke bentuk asalnya (kata dasar). Untuk melakukan tahapan *stemming* terdapat beberapa algoritma salah satunya adalah algoritma *stemming* Nazief-Adriani. Algoritma *stemming* Nazief-Adriani mempunyai persentase keakuratan (presisi) yang lebih besar dibandingkan dengan algoritma-algoritma *stemming* lainnya. Hasil keakuratan algoritma *stemming* Nazief-Adriani sangat dipengaruhi kelengkapan dari kamus kata dasar yang ada. Semakin lengkap kamus kata dasar maka akan semakin akurat pula hasil *stemming*. Sesudah *preprocessing* dan proses *string matching*, langkah terakhir adalah menghitung *similarity* antar dokumen.

Kata kunci : plagiarisme, *string matching*, Rabin-Karp, fungsi *hash*, *preprocessing*, *case folding*, *tokenization*, *number removal*, *filtering*, *stemming*, Nazief-Adriani, *hashing*, *similarity*, dokumen.