Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan perkembangan teknologi dan informasi yang begitu pesat, pencarian informasi menjadi semakin mudah dan cepat. Hal tersebut memberikan dampak positif bagi berbagai kalangan masyarakat dalam memperluas wawasan dan pengetahuan. Namun tak lepas dari itu, kemudahan dalam pencarian informasi pun juga memiliki dampak negatif bagi orang yang menyalahgunakan kemudahan tersebut. Penyalahgunaan hal tersebut seringkali terjadi di dalam dunia pendidikan khususnya dalam penyusunan tugas ataupun karya ilmiah yang dilakukan oleh para pelajar, seperti penjiplakan karya orang lain tanpa menyatakan sumber. Hal tersebut banyak dikenal masyarakat dengan istilah plagiarisme.

Menurut Republik Indonesia (2010), plagiat didefinisikan sebagai "perbuatan secara sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya dan/atau karya ilmiah pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyatakan sumber secara tepat dan memadai." Selain itu, plagiarisme juga dijelaskan sebagai tindakan-tindakan pencurian hasil karya orang lain dan menjadikan hasil karya tersebut seolah-olah hasil karyanya sendiri (Setiyati, 2015).

Plagiarisme menurut Tudesman, dkk. (2014) dapat diatasi dengan cara mencegah dan mendeteksi. Mencegah berarti menjaga atau menghalangi agar plagiarisme tidak dilakukan. Pencegahan harus dilakukan sedini mungkin terutama pada sistem pendidikan dan moral masyarakat, sedangkan mendeteksi berarti

Implementasi Algoritma Rabin..., Deardo Dibrianto Sinaga, FTI, 2017

melakukan usaha untuk menemukan tindakan plagiat yang telah dilakukan. Oleh karena itu, diperlukan sistem pendeteksi plagiarisme pada dokumen teks yang dilakukan secara terkomputerisasi.

Telah banyak *software* dan *website* yang dapat digunakan untuk memeriksa terjadinya plagiarisme pada dokumen teks, tetapi *software* maupun *website* tersebut kurang cocok bagi makalah atau tulisan karya ilmiah dalam Bahasa Indonesia karena dirancang untuk teks dalam Bahasa Inggris (Wibowo, 2012). Oleh karena itu, diperlukan sistem pendeteksi kesamaan pada dokumen teks yang lebih cocok untuk struktur penulisan dalam Bahasa Indonesia.

Dalam penelitian ini, tentunya tak lepas dari penelitian yang telah ada. Penelitian ini mengacu pada karya ilmiah yang ditulis oleh Wicaksono, dkk. (2012) serta Lede, dkk. (2014) yang telah berhasil dalam mengimplementasikan algoritma Rabin-Karp serta algoritma *stemming* Nazief-Adriani untuk mendeteksi kemiripan pada dokumen teks berbahasa Indonesia. Berdasarkan penelitian-penelitian tersebut, maka dalam penelitian ini dilakukan pengembangan lebih lanjut dengan menggunakan algoritma *stemming Confix-Stripping*. Algoritma *stemming Confix-Stripping* sendiri merupakan algoritma pengembangan, atau modifikasi dari algoritma *stemming* Nazief-Adriani yang dilakukan oleh Asian, dkk. (2005). *Stemming* merupakan sebuah cara untuk memperoleh kata dasar dari sebuah kata yang berimbuhan. Hasil penelitian yang dilakukan oleh Asian, dkk. (2005) menyatakan bahwa algoritma Nazief-Adriani memiliki tingkat kesalahan *stemming* di bawah 1 dari 21 kata. Setelah dilakukan modifikasi pada algoritma tersebut dengan menggunakan algoritma *Confix-Stripping*, tingkat kesalahan *stemming* menjadi di bawah 1 dari 38 kata. Dengan demikian penggunaan algoritma *Confix-*

Implementasi Algoritma Rabin..., Deardo Dibrianto Sinaga, FTI, 2017

Stripping terbukti lebih baik dalam melakukan proses *stemming* pada teks berbahasa Indonesia.

Dalam penelitian ini akan diterapkan metode N-Gram yang akan disisipkan dalam proses Rabin-Karp guna meningkatkan nilai *similarity* dari pengujian dokumen. Teknik *hashing* yang digunakan dalam penelitian ini adalah *Rolling Hash*, sedangkan rumus yang digunakan untuk mengukur tingkat *similarity* dari dua dokumen setelah dilakukannya proses Rabin-Karp adalah *Dice's Similarity Coefficient*. Pada penelitian ini juga akan dilakukan pengukuran akurasi dari sistem yang dibangun menggunakan *Confusion Matrix*.

1.2 Rumusan Masalah

Berdasarkan topik yang diambil dalam penelitian ini, didapatkan rumusan permasalahan sebagai berikut.

a. Bagaimana mengimplementasikan algoritma Rabin-Karp dan *stemming Confix-Stripping* pada sistem pendeteksi kemiripan dokumen teks?
b. Bagaimana perbedaan hasil yang didapatkan dalam segi waktu maupun tingkat *similarity* bila sistem pendeteksi kesamaan dokumen menggunakan atau tanpa menggunakan proses *stemming* dengan algoritma *Confix-Stripping*?

1.3 Batasan Masalah

Penelitian ini memiliki batasan masalah sebagai berikut.

a. Sistem tidak memperbaiki atau memperhatikan kesalahan ejaan kata pada dokumen teks yang diuji.
b. Format dokumen yang digunakan dalam sistem pendeteksi kemiripan dokumen teks ini adalah .pdf, .docx, dan .txt.

Implementasi Algoritma Rabin..., Deardo Dibrianto Sinaga, FTI, 2017

c. Kesamaan kalimat atau kata-kata yang terdeteksi pada sistem hanyalah indikasi bahwa adanya plagiat pada dokumen tersebut.

d. Dokumen yang digunakan dalam penelitian adalah dokumen yang berisi teks berbahasa Indonesia.

1.4 Tujuan Penelitian

Tujuan dalam penelitian ini adalah mengimplementasikan algoritma Rabin-Karp dalam mendeteksi tingkat kemiripan dokumen pada teks berbahasa Indonesia, serta mendapatkan hasil analisis saat sistem menggunakan atau tidak menggunakan algoritma *stemming Confix-Stripping*.

1.5 Manfaat Penelitian

Manfaat yang didapatkan dari penelitian ini adalah mengetahui perbandingan performa sistem baik dalam segi waktu atau tingkat *similarity* dokumen, dengan menggunakan atau tanpa menggunakan algoritma *stemming Confix-Stripping*. Selain itu sistem ini juga diharapkan dapat digunakan untuk membantu mendeteksi tingkat kemiripan pada dokumen teks.

1.6 Sistematika Penulisan Laporan Penelitian

Laporan penelitian ini memiliki sistematika yang terdiri dari lima bagian utama, yaitu sebagai berikut.

1. BAB I PENDAHULUAN

Bab ini berisi penjelasan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan penelitian.

Implementasi Algoritma Rabin..., Deardo Dibrianto Sinaga, FTI, 2017

2. BAB II LANDASAN TEORI

Bab ini berisi penjelasan tentang dasar teori yang digunakan dalam penelitian. Adapun dasar teori yang digunakan dalam penelitian ini adalah plagiarisme, algoritma Rabin-Karp, teknik *hashing Rolling Hash,* metode N-Gram, tingkat pengukuran *Dice's Similarity Coefficient,* serta algoritma *stemming Confix-Stripping,* dan juga *Confusion Matrix* untuk mengukur tingkat akurasi sistem.

3. BAB III METODOLOGI DAN PERANCANGAN SISTEM

Bab ini berisi penjelasan tentang metodologi serta perancangan sistem yang digunakan dalam pembuatan aplikasi, seperti *flowchart* dan desain *user interface.*

4. BAB IV IMPLEMENTASI DAN UJI COBA

Bab ini berisi penjelasan tentang implementasi dan pengujian yang dilakukan terhadap sistem yang telah dibuat.

5. BAB V SIMPULAN DAN SARAN

Bab ini berisi kesimpulan akhir yang didapat dari penelitian dan saran untuk penelitian selanjutnya.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA