

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi dari tahun ke tahun selalu berkembang secara signifikan. Jumlah pengguna internet yang besar dan semakin berkembang mempunyai pengaruh yang besar atas ilmu dan pandangan dunia (Darma, dkk., 2009). Google merupakan salah satu mesin pencari yang memberikan kemudahan dalam mencari sebuah dokumen maupun sumber ilmiah (Sriyati, 2009). Dalam hal ini pengguna dipermudah dalam pencarian dokumen, baik sebagai sumber informasi maupun referensi dalam membuat sebuah karya ilmiah. Hal tersebut merupakan salah satu dampak positif dalam kemajuan teknologi. Namun semakin banyak kemudahan yang diberikan kepada pengguna, pastinya ada dampak negatifnya, salah satunya ialah *plagiarism* (Ulum, 2014).

Plagiarisme adalah salah satu bentuk kecurangan dalam bentuk penyalinan atau penjiplakan suatu tulisan. Neville (2010) dalam buku “The Complete Guide to Referencing and Avoiding Plagiarism” mendefinisikan *plagiarisme* sebagai tindakan atau praktek yang dianggap oleh universitas merupakan suatu kecurangan dengan cara mengambil ide atau tulisan orang lain tanpa menyebutkan rujukan dan diklaim sebagai miliknya. Oleh karena itu penulisan sebuah rujukan dan sumber merupakan hal yang mutlak agar sebuah karya tidak dikatakan sebuah plagiat. Sastroasmoro (2007) klasifikasi *plagiarisme* berdasarkan proporsi atau persentasi kata, kalimat, *paragraph* yang dibajak, terbagi menjadi tiga yaitu *plagiarisme* ringan: <30%, *plagiarisme*

sedang: 30% - 70%, dan *plagiarisme* berat atau total: >70%.

Dalam sebuah dokumen terdapat kumpulan *string* yang dirangkai menjadi satu kesatuan kata atau kalimat. Terdapat banyak macam algoritma yang dapat digunakan dalam *string matching* dan setiap algoritma memiliki kompleksitasnya masing-masing. Dengan menggunakan teknik *string matching*, maka antar dokumen dapat dibandingkan apakah terdapat indikasi *plagiarisme* atau tidak.

Dalam pembuatan penelitian ini tentunya tidak lepas dari referensi lainnya, baik itu dari skripsi maupun jurnal-jurnal pendukung dari peneliti lainnya. Penelitian sejenis yang berhubungan dengan penelitian ini diantaranya penelitian tugas akhir yang disusun oleh Kornain, dkk. (2014) dengan judul “Penerapan Algoritma Jaro Winkler Distance untuk Sistem Pendeteksi Plagiarisme pada Dokumen Teks Berbahasa Indonesia”, Nugroho (2011) dengan judul “Perancangan Sistem Deteksi Plagiarisme Dokumen Teks dengan Menggunakan Algoritma Rabin-Karp”, Faranika, dkk. (2013) dengan judul “Sistem Pengukur Kemiripan Dokumen Menggunakan Algoritma Jaro Winkler Distance”. Dari ketiga penelitian tersebut diperoleh kesimpulan bahwa kedua algoritma dapat digunakan untuk mendeteksi kesamaan dokumen, seperti deteksi plagiat pada dokumen.

Penelitian ini merupakan pengembangan berdasarkan penelitian dari Kornain, dkk. (2014) yang menyarankan agar aplikasi dapat diterapkan berbasis *website*. Selain itu berdasarkan penelitian dari Nugroho (2011), dalam penelitian tugas akhirnya peneliti menyarankan agar data yang diuji lebih bervariasi dan bentuk kalimat yang digunakan lebih banyak. Berdasarkan hasil uji coba yang dilakukan oleh Kornain, dkk. (2014) membuktikan bahwa Algoritma Jaro Winkler Distance berhasil untuk mendeteksi *plagiarisme* terhadap suatu dokumen. Sama seperti penelitian dengan

menggunakan Algoritma Jaro Winkler Distance, hasil uji coba dengan menggunakan Algoritma Rabin-Karp yang dilakukan oleh Nugroho (2011) membuktikan bahwa algoritma ini dapat digunakan untuk mendeteksi *plagiarisme* meskipun dokumen yang diuji berukuran besar.

Menurut Wicaksono, dkk. (2012) untuk membuat suatu sistem pendeteksi plagiat diperlukan algoritma yang baik untuk jenis *multiple string matching pattern*. Salah satu algoritma yang cocok untuk permasalahan *multiple string matching pattern* adalah Rabin-Karp algorithm. Kelebihan dari algoritma Rabin-Karp dibandingkan algoritma pencocokan *string* lainnya adalah kemampuan dalam mencari banyak pola *string* (Cormen, dkk., 2003).

Algoritma Jaro Winkler Distance adalah algoritma yang menggunakan pendekatan *string metric*, yaitu melakukan perbandingan *string* dengan memasukkannya ke dalam fungsi matematis tertentu. Beberapa algoritma yang berdasarkan kepada string metric diantaranya adalah *Levenshtein distance*, *TF/IDF*, *Needleman-Wunsch distance*, *Jaro-Winkler distance*, dan sebagainya. Dari algoritma yang telah disebutkan, Jaro Winkler Distance memiliki ketepatan yang baik di dalam pencocokan string yang relatif pendek (Kurniawati, dkk., 2010). Menurut Kurniawati, dkk. (2010), algoritma ini memiliki *quadratic runtime complexity* yang sangat efektif pada *string* pendek dan dapat bekerja lebih cepat.

Berdasarkan penelitian yang telah dilakukan serta sumber-sumber yang ada diperoleh kesimpulan dasar bahwa algoritma Rabin-Karp dan Jaro Winkler Distance dapat digunakan untuk mendeteksi *plagiarisme*. Selain itu, kedua algoritma tersebut memiliki keunggulan satu sama lain dengan kompleksitas waktu yang sama pada

tahapan face *preprocessing* dengan yaitu $O(m)$ meskipun kompleksitas fase pencariannya berbeda, sehingga menjadi sebuah inspirasi dalam melakukan penelitian untuk membandingkan kinerja dari kedua algoritma tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka permasalahan penelitian ini dapat dirumuskan menjadi manakah algoritma yang dapat mendeteksi *plagiarisme* dokumen teks paling cepat dan efektif, diantara Algoritma Rabin-Karp dan Algoritma Jaro Winkler Distance?

1.3 Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan agar penelitian ini tetap fokus sebagai berikut.

1. Hanya menguji dokumen yang murni berisi teks saja dan berbahasa Indonesia.
2. Hasil kemiripan dokumen dengan dokumen pembanding disimpan dalam *database* hanya dalam bentuk persentase kemiripan.
3. Kesalahan ejaan atau *typo* serta kesamaan makna dalam penulisan pada dokumen yang diuji tidak langsung dikategorikan plagiat, namun dilihat persentase kemiripan.
4. Hasil akhir dari penelitian untuk mengukur tingkat efektivitas dari kedua algoritma.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai dalam penelitian ini adalah mendapatkan algoritma yang efektif dan paling cepat antara Algoritma Rabin Karp dan Algoritma Jaro Winkler Distance dalam mendeteksi *plagiarisme* pada dokumen teks.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian dan pembuatan aplikasi ini antara lain:

1. Membantu memberikan pertimbangan mengenai aspek pengukuran sebuah dokumen dikatakan plagiat atau tidak bagi masyarakat.
2. Mengetahui tingkat efektivitas algoritma yang cocok untuk pendeteksian *plagiarisme* pada dokumen teks agar peneliti lainnya dapat membuat sistem pendeteksian *plagiarisme* menggunakan salah satu algoritma.

1.6 Sistematika Penulisan Laporan Penelitian

Sistematika penulisan dalam laporan penelitian ini disajikan sebagai berikut.

BAB I PENDAHULUAN

Dalam bab ini terdapat uraian mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan penelitian ini.

BAB II LANDASAN TEORI

Bab ini menjabarkan berbagai konsep dasar dan teori-teori yang digunakan untuk mendukung penelitian terkait dari permasalahan yang diuraikan, meliputi *plagiarisme*, algoritma, kompleksitas algoritma, performa algoritma, *string matching*, algoritma

Rabin Karp, konsep algoritma Rabin Karp, algoritma Jaro Winkler Distance, dan konsep algoritma Jaro Winkler Distance.

BAB III METODE DAN PERANCANGAN APLIKASI

Pada bab ini, terdapat penjabaran secara komprehensif dari metode penelitian dan rancangan dari aplikasi yang dibuat.

BAB IV IMPLEMENTASI DAN HASIL PENELITIAN

Bab ini berisi mengenai implementasi sistem, disertai dengan data dari hasil penelitian yang sudah dilakukan beserta hasil analisa data tersebut.

BAB V SIMPULAN DAN SARAN

Dalam bab ini, menjabarkan simpulan dari penelitian yang telah dilakukan untuk mencapai tujuan dari penelitian ini dan saran terhadap pengembangan penelitian lebih lanjut

