BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi informasi saat ini membuat dokumen yang tadinya bersifat fisik kini telah banyak dibuat dalam bentuk digital sehingga dapat dengan mudah dilakukan penyalinan digital yang dapat menimbulkan plagiarisme. Plagiarisme merupakan tindakan menjiplak, mencuri atau mengambil ide, hasil karya atau tulisan orang lain, baik seluruh, sebagian besar maupun sebagian kecil, untuk jadi ide atau karya tulisan sendiri tanpa menyebutkan nama penulis dan sumber aslinya [1]. Dalam mendeteksi plagiarisme dokumen digital terdapat beberapa metode yang dapat digunakan untuk mengukur tingkat similaritas sebuah dokumen yaitu metode perbandingan teks lengkap, metode kesamaan kata kunci dan metode *fingerprinting* [2,3].

Dalam penelitian ini akan dilakukan analisis pengukuran similaritas antar dokumen menggunakan metode *fingerprinting*. Metode *fingerprinting* adalah metode yang menelusuri karakter satu persatu pada deret karakter. Prinsip kerja dari metode dokumen *fingerprinting* ini adalah dengan menggunakan teknik *hashing*. Teknik *hashing* adalah sebuah fungsi yang mengkonversi setiap *string* menjadi bilangan. Bilangan-bilangan tersebut menghasilkan nilai-nilai *fingerprints* sebagai acuan dalam perhitungan kesamaan dokumen. Kelebihan dari metode *fingerprinting* adalah waktu proses lebih cepat dibandingkan dari metode perbandingan teks lengkap dan metode kesamaan kata kunci [2]. Pengujian similaritas dokumen *fingerprinting* memiliki tahapan pencarian nilai hashing dari setiap kata, pengambilan nilai *fingerprints* dan pembobotan persentasi kemiripan.

Beberapa algoritma yang termasuk dalam metode *fingerprinting* yaitu Algoritma Rabin Karp, Algoritma Winnowing dan Algoritma Manber. Dalam penelitian ini akan dilakukan analisis performansi dari dua buah algoritma yang termasuk dalam metode

fingerprinting yaitu Algoritma Winnowing dan Algoritma Manber yang dapat diterapkan untuk mendeteksi similaritas pada dokumen teks. Analisis performansi dilakukan dengan tujuan mendapatkan informasi dari kedua algoritma yaitu berupa informasi kecepatan, ketepatan, dan jumlah langkah dari Algoritma Winnowing dan Algoritma Manber dalam penerapannya pada sebuah sistem.

Dari sedikitnya penelitian dalam lingkup analisis performansi kedua algoritma ini maka menjadi dasar penelitian ini. Terdapat penelitian sebelumnya mengenai perbandingan pendekatan kedua algoritma ini yang berjudul "Perbandingan Pendekatan Deteksi Plagiarisme Dokumen Dalam Bahasa Inggris" oleh Ana Kurniawati dan I Wayan Simri Wicaksana dalam jurnalnya yang menganalisis perbandingan dan penerapannya pada dokumen berbahasa inggris. Berdasarkan hasil penelitian tersebut, disimpulkan bahwa algoritma winnowing lebih baik dari algoritma manber karena memberikan jaminan terdeteksinya dokumen sama [2]. Penerapan penelitian tersebut dilakukan pada sistem yang memproses teks berbahasa inggris. Berdasarkan hal tersebut maka akan dilakukan analisis yang nantinya akan diterapkan pada dokumen berbahasa Indonesia.

Penelitian lainnya dari kedua algoritma yaitu berjudul "Implementasi Algoritma Winnwoing untuk Mendeteksi Kemiripan Pada Dokumen Teks" [4], hanya berfokus pada implementasi algoritma sehingga informasi dari parameter-parameter yang digunakan tidak memiliki standar optimal untuk diterapkan pada rumus yang dipakai oleh kedua algoritma tersebut. Ketika ditemukan jumlah parameter optimal diharapkan dapat mendapatkan akurasi yang lebih tepat dalam mendeteksi similaritas dari dokumen yang dibandingkan. Parameter yang perlu diuji yaitu jumlah gram, jumlah window, basis bilangan prima yang di gunakan dan proses pemilihan nilai *hash*ing yang berpengaruh pada persentasi kemiripan dari dokumen yang dibandingkan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang diuraikan, maka dapat diidentifikasi masalah dalam penelitian ini yaitu:

1. Bagaimana mendeteksi kesamaan antar dokumen teks sebagai dugaan plagiarisme menggunakan metode *fingerprinting.*
2. Bagaimana performansi Algoritma Winnowing dan Algoritma Manber dalam mendeteksi similaritas dokumen.
3. Bagaimana pengaruh jumlah parameter input terhadap persentasi kemiripan dokumen yang dibandingkan.

1.3 Maksud dan Tujuan

1.3.1 Maksud

Maksud dari penelitian ini adalah untuk menganalisis Algoritma Winnowing dan Algoritma Manber dalam mendeteksi kesamaan dokumen teks berbahasa Indonesia

1.3.2 Tujuan

Sedangkan tujuan yang akan dicapai yaitu mendapatkan informasi dari Algoritma Winnowing dan Algoritma Manber berupa informasi perbandingan kecepatan, ketepatan, keefektifan langkah dan pengaruh jumlah parameter input dalam proses mendeteksi kesamaan dokumen.

1.4 Batasan Masalah

Batasan masalah dari penelitian ini yaitu:

1. Metode yang digunakan adalah metode *fingerprinting.*
2. Penerapan algoritma pada sistem memproses dokumen berbahasa Indonesia.
3. Penerapan algoritma pada sistem tidak membedakan arti atau sinonim dari kata yang diseleksi.

4. Penerapan algoritma menggunakan metode *Jaccard's Similarity Coefficient* sebagai perhitungan kesamaan.
5. Parameter yang akan diukur adalah kecepatan, ketepatan dan jumlah langkah dari kedua algoritma.
6. Sistem tidak melihat acuan referensi yang digunakan dokumen.
7. Dokumen yang akan dideteksi berextensi .txt
8. Jumlah gram dibatasi 2 sampai 10

1.5 Metodologi Penelitian

Metodologi penelitian yang dimaksud adalah segala hal yang berhubungan dengan metode-metode yang digunakan dalam melakukan penelitian ini dengan cara melakukan pendekatan terhadap metode-metode yang telah ada. Dalam penelitian ini menggunakan metode-metode yang diuraikan di bawah ini.

1.5.1 Metode Penelitian

Metode penelitian yang dipakai adalah metode penelitian deskriptif. Dalam metode penelitian ini digunakan teknik-teknik analisis, klasifikasi masalah, studi literatur terhadap masalah-masalah yang berhubungan dengan skripsi yang disusun, dan teknik *test* terhadap objek penelitian yang telah ada..

Dengan metode deskriptif, data yang telah dikumpulkan mula-mula disusun, dijelaskan, dianalisis, dan kemudian diimplementasikan dalam sebuah perangkat lunak.

1.5.2 Tahap Pengumpulan Data

Pengumpulan data adalah cara atau prosedur yang sistematis dan standar untuk memperoleh data yang diperlukan. Adapun metode pengumpulan data pada penelitian ini yaitu studi pustaka dengan cara mengkaji buku, media, atau hasil penelitian orang lain yang memiliki keterkaitan dengan penelitian ini.

1.5.3 Tahap Pengembangan Perangkat Lunak

Teknik analisis data dalam pembuatan perangkat lunak menggunakan paradigma perangkat lunak secara *prototyping*, yang meliputi beberapa proses diantaranya:

1. Pengumpulan kebutuhan

   Pendefinisian format seluruh perangkat lunak, mengidentifikasikan semua kebutuhan, dan garis besar sistem yang akan dibuat.

2. Membangun prototyping

   Membangun prototyping dengan membuat perancangan sementara yang berfokus pada penyajian kepada pengguna (misalnya dengan membuat input dan format output)

3. Evaluasi protoptyping

   Evaluasi ini dilakukan oleh pengguna apakah prototyping yang sudah dibangun sudah sesuai dengan keinginann pengguna. Jika sudah sesuai maka langkah 4 akan diambil. Jika tidak prototyping direvisi dengan mengulangi langkah 1, 2 , dan 3.

4. Mengkodekan sistem

   Dalam tahap ini prototyping yang sudah di sepakati diterjemahkan ke dalam bahasa pemrograman yang sesuai.

5. Menguji sistem

   Setelah sistem sudah menjadi suatu perangkat lunak yang siap pakai, harus dites dahulu sebelum digunakan. Pengujian ini dilakukan dengan metode pengujian Black Box yaitu pengujian luar sistem.

6. Evaluasi Sistem

   Pengguna mengevaluasi apakah sistem yang sudah jadi sudah sesuai dengan yang diharapkan . Jika ya, langkah 7 dilakukan; jika tidak, ulangi langkah 4 dan 5.

7. Menggunakan sistem

   Perangkat lunak yang telah diuji dan diterima pengguna siap untuk digunakan .

1.6 Sistematika Penulisan

Penyusunan skripsi ini dibagi ke dalam beberapa bab secara sistematis sesuai dengan pokok-pokok permasalahan yang dibahas. Adapun sistematika penulisan secara umum adalah sebagai berikut:

BAB I . PENDAHULUAN

Bab ini menjelaskan secara singkat mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, metodologi penelitian, batasan masalah, serta sistematika penulisan.

BAB II . LANDASAN TEORI

Pada bab ini berisi tentang analisis kebutuhan dalam membangun aplikasi ini yang sesuai dengan metode pembangunan perangkat lunak yang digunakan. Selain itu terdapat juga teori-teori yang berkaitan dengan penelitian yang dijalankan.

BAB III . ANALISIS DAN PERANCANGAN SISTEM

Bab ini menerangkan analisis yang dilakukan terhadap cara kerja aplikasi yang dibuat. Bab ini juga membahas mengenai lingkungan bahasa pemrograman yang digunakan dan perancangan dalam tahapan-tahapan yang sistematis.

BAB IV . IMPLEMENTASI SISTEM

Pada bab ini berisi tentang analisis kebutuhan dalam membangun aplikasi ini yang sesuai dengan metode pembangunan perangkat lunak yang digunakan. Selain itu terdapat juga perancangan antarmuka dan pengujian untuk aplikasi yang akan dibangun sesuai dengan hasil analisis.

BAB V . KESIMPULAN DAN SARAN

Bab ini terdiri dari kesimpulan dan saran yang berisikan hal-hal terpenting yang penulis bahas dan kemudian dibuat kesimpulan. Bab ini juga berisi saran-saran yang penulis berikan untuk pengembangan selanjutnya.