

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan internet menyebabkan semakin banyaknya informasi yang tersedia. Hal ini memudahkan seseorang dalam melakukan penjiplakan suatu karya. Penjiplakan menurut Kamus Besar Bahasa Indonesia (KBBI) berarti menggambar atau menulis garis-garis gambaran atau tulisan yang telah tersedia (dengan menempelkan kertas kosong pada gambar atau tulisan yang akan ditiru), mencontoh atau meniru tulisan atau pekerjaan orang lain, mencuri karangan orang lain dan mengakui sebagai karangan sendiri, mengutip karangan orang lain tanpa seizin penulisnya [1].

Plagiarisme dapat dikelompokkan berdasarkan proporsi atau persentase kata, kalimat, atau paragraf yang dibajak, yaitu plagiarisme ringan (<30%), plagiarisme sedang (30 – 70%) dan plagiarisme besar atau total (>70%) [2]. Zulkarnain menyatakan bahwa cara untuk menghindari plagiarisme adalah senantiasa taat pada gaya selingkung, melakukan pengutipan (menyitir) secara langsung, dan melakukan parafrasa terhadap kutipan yang dirujuk [3].

Tindakan plagiarisme dapat membuat seseorang menjadi malas berkreatifitas karena tidak mau berfikir [4]. Ada dua cara untuk mengatasi permasalahan plagiarisme, yaitu dengan mencegah dan mendeteksi. Mencegah berarti menjaga atau menghalangi agar plagiarisme tidak dilakukan. Mendeteksi tindakan plagiarisme

berarti melakukan pencocokan kemiripan dokumen dengan cara membaca satu persatu dokumen, namun jika dilakukan secara manual merupakan suatu yang sulit karena membutuhkan waktu dan tenaga yang ekstra.

Oleh karena itu dibutuhkan sebuah aplikasi yang mampu membantu untuk mendeteksi kesamaan antar dokumen. Saat ini sudah banyak algoritma yang berfungsi sebagai pendeteksi kemiripan antar dokumen teks seperti algoritma *Jaro-Winkler Distance* sebuah algoritma untuk mengukur kesamaan antara dua *string*, biasanya algoritma ini digunakan didalam pendeteksian duplikat. Semakin tinggi *Jaro-Winkler Distance* untuk dua *string*, semakin mirip dengan *string* tersebut. Nilai normalnya yaitu 0 untuk menandakan tidak ada kesamaan, dan 1 untuk menandakan adanya kesamaan [5], algoritma Rabin Karp yang menggunakan metode hash dalam mencari suatu kata. Teori ini jarang digunakan untuk mencari kata tunggal, namun cukup penting dan sangat efektif bila digunakan untuk pencarian jamak [6], Algoritma N-Gram dan *Winnowing*, Algoritma N-gram digunakan untuk mengambil potongan-potongan karakter huruf sejumlah n dari sebuah kata yang secara kontinuitas dibaca dari teks sumber sampai akhir dari sebuah dokumen, sedangkan *Winnowing* membuang seluruh karakter-karakter yang tidak relevan misalnya tanda baca dan spasi [7], algoritma Smith Waterman yaitu metode klasik untuk membandingkan dua string (rangkaian huruf) dengan sebuah pandangan untuk mengidentifikasi bagian yang sangat mirip di dalamnya. Algoritma ini digunakan secara luas dalam pencarian *near-matches* (kecocokan yang dekat) yang baik, atau yang sering disebut dengan *local alignments*, di dalam jajaran/rentetan biologis [8]

dan algoritma SCAM. Diantara algoritma yang sudah disebutkan, algoritma SCAM baik dalam mendeteksi dokumen yang *overlap* tanpa bergantung pada *length* dokumen tersebut dan juga SCAM mampu mengenali dengan baik untuk dokumen yang merupakan *subset* atau superset dari dokumen lainnya [7].

Berdasarkan latar di atas, akan dibuat aplikasi deteksi plagiat pada dokumen teks menggunakan algoritma SCAM.

1.2 Rumusan Masalah

Bagaimana mengimplementasikan algoritma SCAM untuk mendeteksi plagiarisme pada dokumen teks.

1.3 Tujuan dan Manfaat

Tujuan dari tugas akhir ini adalah :

1. Mengimplemenetasikan algoritma SCAM untuk mendeteksi plagiat pada dokumen teks.
2. Mengetahui tingkat kemiripan (similaritas) pada dokumen.
3. Mengetahui proses deteksi kemiripan (similaritas) pada algoritma SCAM
4. Mengetahui durasi waktu proses algoritma SCAM dalam melakukan proses deteksi kemiripan (similaritas).

Adapun manfaat yang diperoleh dari tugas akhir ini yaitu memberikan kemudahan dalam mengetahui sebuah dokumen teks terindikasi plagiat atau tidak.

1.4 Ruang Lingkup

Agar pengerjaan tugas akhir ini menjadi lebih terarah dan mendapatkan hasil yang lebih spesifik, maka aplikasi yang dirancang dibatasi pada ruang lingkup pembahasan sebagai berikut:

1. Algoritma yang digunakan dalam pembuatan aplikasi ini menggunakan algoritma SCAM.
2. Aplikasi ini dirancang untuk menentukan seberapa besar kemiripan sebuah dokumen teks yang berformat *.pdf.

1.5 Metodologi

Metodologi yang digunakan dalam tugas akhir ini meliputi :

1. Pengumpulan Data

Pengumpulan data dengan studi kepustakaan dilakukan dengan cara mempelajari jurnal, buku-buku, dan sumber terkait lainnya yang menjadi acuan dalam membangun aplikasi deteksi plagiat menggunakan algoritma SCAM.

2. *Preprocessing*

Pada tahap ini data disiapkan menjadi data yang siap diolah menggunakan algoritma SCAM. Proses yang perlu disiapkan diantaranya adalah sebagai berikut :

- a) *Case Folding*, mengubah huruf menjadi huruf kecil.

- b) *Cleaning*, menghapus noise seperti tanda baca yaitu titik(.), koma(,), dan lain-lain
- c) *Filtering*, menghapus kata yang tidak memiliki makna.
- d) Tokenisasi, memecah kalimat menjadi kata.
- e) *Stemming*, mengubah kata-kata menjadi kata dasar.

3. Analisis Algoritma

Pada tahap ini dilakukan analisis cara kerja perhitungan kemiripan menggunakan algoritma SCAM terhadap dokumen teks.

4. Implementasi

Algoritma SCAM diimplementasikan dengan menggunakan bahasa pemrograman PHP.

1.6 Sistematika Penulisan

Secara garis besar materi laporan tugas akhir ini terbagi dalam beberapa bab yang tersusun sebagai berikut :

BAB I PENDAHULUAN

Pada bab ini akan menjelaskan mengenai latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, metodologi dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini akan menjelaskan mengenai dasar-dasar pengetahuan yang dapat membantu merancang dan memperkuat aplikasi.

BAB III ANALISIS DAN PERANCANGAN

Bab ini memuat tentang analisis kebutuhan serta perancangan aplikasi deteksi plagiat yang akan dibuat.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas implementasi algoritma SCAM untuk mendeteksi kemiripan dokumen dalam bentuk tampilan antarmuka yang telah dirancang serta pengujian terhadap perangkat lunak yang meliputi tingkat akurasi deteksi, performa aplikasi dan kecepatan eksekusi.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran yang berkaitan dengan pembuatan aplikasi dan penulisan laporan berdasarkan apa yang telah diuraikan pada bab - bab sebelumnya.