

BAB I PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi saat ini sudah sangat pesat tidak hanya memberikan efek positif akan tetapi ada dampak negatif juga yang tidak dapat dihindari [1], salah satu kasusnya adalah teknologi mesin pencari. Teknologi mesin pencari yang semakin berkembang sangat membantu pengguna internet untuk mendapatkan informasi secara *online*. Dengan menggunakan situs mesin pencarian seperti Google, Bing, dan yahoo informasi yang dibutuhkan hampir semuanya tersedia, mulai dari berita, ilmu pengetahuan, atau bahkan *paper* dari penelitian orang lain. Komputer dan internet merupakan penyebab utama meningkatnya aksi plagiasi [2].

Plagiarisme adalah tindakan yang disengaja atau tidak disengaja untuk mendapatkan nilai untuk studi ilmiah. Selain itu, plagiarisme juga digambarkan sebagai tindakan pencurian karya orang lain dan membuat pekerjaan seolah-olah itu adalah pekerjaan sendiri, sebagian peneliti sering melakukan kecurangan ketika mengerjakan karya ilmiah atau tugas. Hal ini dapat terjadi dengan mengutip sebagian atau keseluruhan dari karya orang lain tanpa menyertakan sumbernya [3]. Menurut penelitian yang dilakukan oleh Pusat Penelitian Pew pada tahun 2011, bahwa ada peningkatan plagiarisme *cyber* di kalangan pelajar, sekitar 55% pelajar mengaku melakukan plagiasi dalam mengerjakan tugas. Tidak hanya dilakukan oleh kalangan pelajar saja, tetapi kalangan peneliti juga melakukan hal serupa dalam mengerjakan karya tulis ilmiah [2].

Mendeteksi plagiarisme sangatlah penting dalam dunia pendidikan, karena kredibilitas lembaga ini sebagian besar didasarkan pada orisinalitas. Dapat mendeteksi plagiarisme telah menjadi salah satu tantangan tersendiri bagi dunia pendidikan, terlebih sebagian besar dari pelajar atau peneliti sering melakukan kecurangan saat melakukan pekerjaannya. Dengan memanfaatkan mesin pencari di internet, pelajar atau peneliti bisa mendapatkan topik tertentu dan digunakan untuk melakukan kecurangan berupa plagiarisme [3].

Dari penjelasan masalah sebelumnya yang sudah dipaparkan, bahwa penting untuk membuat aplikasi pendeteksi plagiasi khususnya dokumen teks. Metode untuk mendeteksi plagiarisme dapat diklasifikasikan menjadi tiga metode yaitu metode perbandingan teks lengkap, metode dokumen *fingerprinting* dan metode kesamaan kata kunci [4]. Untuk saat ini sudah ada beberapa aplikasi gratis dan juga berbayar untuk mendeteksi plagiasi seperti *Turnitin.com*, *Plagiserve.com*, *CopyCatch Gold*, *Eve2 – Essay Verification Engine*, dan *WordCheck Keyword DP* [3]. Y. arifin dkk pernah melakukan penelitian pada tahun 2018 dengan menggunakan algoritma *Winnowing (fingerprint)* untuk mendeteksi plagiasi dokumen teks berbahasa Indonesia dengan menerapkan proses paralel untuk meningkatkan kecepatan proses pendeteksian plagiasi. Dari penelitian yang dilakukan didapatkan hasil dengan hematnya waktu proses sebesar 74,66% [2].

Pada penelitian sebelumnya Muhammad Alif menggunakan *stemming Porter Stemmer* bahasa Indonesia dokumen uji coba diambil dari penelitian Tala dengan uji coba 300 dokumen porter hanya mampu menghasilkan 61%[5]. Untuk melakukan perhitungan persamaan dokumen, digunakanlah metode *cosine similarity* dan pembobotan menggunakan TF-IDF. Metode *cosine similarity* digunakan karena metode ini mempunyai nilai akurasi yang tinggi dimana kelebihan utamanya adalah tidak terpengaruh pada panjang pendeknya suatu dokumen dan pembobotan TF-IDF digunakan karena metode ini merupakan metode yang paling umum digunakan pada *information retrieval* untuk perhitungan bobot suatu kata. Metode ini juga terkenal efisien, mudah, dan memiliki hasil yang tepat. [6][7].

Algoritma yang digunakan oleh Kusnawan dalam aplikasi pendeteksian plagiarisme di Indonesia hanya menggunakan cara sederhana pra-pemrosesan teks seperti menghilangkan spasi, dan karakter yang tidak dibutuhkan. Metode semacam ini hanya bisa mendeteksi plagiarisme dengan cara sederhana tetapi tidak dapat mendeteksi bentuk yang lebih kompleks [8]. Metode semacam ini hanya bisa mendeteksi plagiarisme dengan cara sederhana tetapi tidak dapat mendeteksi bentuk yang lebih kompleks. Dengan tambahan proses *stemming Enhanced Confix Stripping* dalam pra-pemrosesan. Pendekatan yang diusulkan menggunakan algoritma *Rabin-Karp* sebagai tahap membandingkan nilai *hash* antara *string*

masuk dan *substring* dalam teks, jika nilai *hash* adalah sama, maka perbandingan dilakukan sekali lagi untuk karakter, tetapi jika nilai *hash* tidak sama, *substring* bergeser ke kanan [9].

Metode *stemming* Bahasa Indonesia sudah banyak dikembangkan seperti *Porter Stemmer*, *Confix Stripping Stemmer*, Nazief & Andriani dan *Enhanced Confix Stripping Stemmer* [10]. Pada penelitian sebelumnya Dibrianto Sinaga menggunakan *Confix Stripping Stemmer* untuk mengetahui pengaruh stemming terhadap algoritma *Rabin-karp* dengan hasil waktu proses rata-rata 34,4881 detik dan tingkat kesamaan rata-rata 74,5% data yang digunakan merupakan potongan-potongan teks bahasa Indonesia, dokumen yang diuji berjumlah 31. Pada penelitian tersebut masih memiliki masalah karena menggunakan algoritma stemming yang belum dikembangkan hal tersebut bisa mempengaruhi kecepatan proses, kemiripan dokumen maupun akurasi dari sistem yang dibangun [9]. Algoritma *Enhanced Confix Stripping Stemmer* merupakan perbaikan dari algoritma sebelumnya yaitu, algoritma *Confix Stripping Stemmer*. Setelah dilakukan beberapa percobaan dan analisis, ditemukan beberapa kata yang tidak dapat di-stemming menggunakan *Confix Stripping Stemmer*. Pada 2013 Muhamad Alif melakukan penelitian metode *ECS* dengan hasil Uji coba 300 dokumen, *ECS* telah menghasilkan 81% dengan kecepatan 11,8311 detik [11].

Untuk mengatasi permasalahan yang ada pada aplikasi pendeteksi plagiasi sebelumnya, Di penelitian ini ingin membandingkan metode *Winnowing* dengan *Cosine Similarity* yang dikombinasikan dengan *Enhanced Confix Stripping* untuk mengetahui metode apakah yang efektif dalam mendeteksi plagiasi yang lebih tepat untuk meminimalisir aksi plagiarisme bahasa Indonesia. [2] [12].

1.2 Perumusan Masalah

1.2.1 Permasalahan

Tingkat kemiripan dari suatu dokumen dapat dideteksi menggunakan metode *Cosine Similarity* dan metode *Winnowing*. Metode *cosine similarity* merupakan metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah vektor dengan menggunakan *keywords* (kata kunci), sedangkan metode *Winnowing* merupakan metode untuk mengidentifikasi plagiasi

(penjiplakan) berdasarkan proses pengecekan kesamaan kata (*document fingerprinting*). *Enhanced Confix Stripping* adalah algoritma dalam proses stemming pada Bahasa Indonesia dengan tingkat keakuratan lebih baik. Maka dari itu dilakukan penelitian untuk membandingkan manakah yang lebih baik dari kedua metode tersebut jika dikombinasikan menggunakan *stemming Enhanced Confix Stripping* untuk mendeteksi kemiripan dokumen.

1.2.2 Metode Usulan

Metode yang di usulkan yaitu penerapan metode *Winnowing* dan metode *Cosine Similarity* untuk mengecek kemiripan suatu dokumen, serta menerapkan metode *stemming Enhanced Confix Stripping* untuk meningkatkan proses *pre-processing*.

1.2.3 Pertanyaan Penelitian

1. Seberapa tepat algoritma *Winnowing* dengan *Cosine Similarity* yang sama-sama dikombinasikan dengan *Enhanced Confix Stripping* untuk mendeteksi plagiasi dokumen teks Bahasa Indonesia?
2. Seberapa cepat proses algoritma *Winnowing* dengan *Cosine Similarity* yang sama-sama dikombinasikan dengan *Enhanced Confix Stripping* untuk mendeteksi plagiasi dokumen teks Bahasa Indonesia?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Penerapan *Enhanced Confix Stripping* di algoritma *Winnowing* dengan *Cosine Similarity* bertujuan untuk mengetahui hasil yang lebih tepat dan efisien dalam segi waktu untuk mendeteksi kemiripan dokumen bahasa Indonesia.

1.3.2 Manfaat

Dapat mengetahui metode yang lebih baik dan efisien jika *Enhanced Confix Stripping* dikombinasikan dengan *Winnowing* dengan *Cosine similarity*.

1.4 Batasan Masalah

Batasan masalah dari penelitian ini yaitu :

1. Dokumen yang diuji merupakan dokumen teks dengan format dokumen yang digunakan dalam penelitian ini merupakan .txt

2. Teks yang digunakan sebagai dokumen uji merupakan dokumen yang menggunakan bahasa Indonesia yang baku.
3. Uji coba yang dilakukan adalah abstrak jurnal
4. Sistem yang dibangun menggunakan bahasa pemrograman PHP.

1.5 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Pada bab ini dijelaskan mengenai materi-materi yang menjadi landasan dalam penyusunan penelitian ini. Materi-materi tersebut diantaranya adalah penelitian-penelitian terkait sistem deteksi plagiasi, algoritma *Winnowing*, algoritma *Cosine Similarity*, dan *Enhanced Confix Stripping*.

BAB III METODE USULAN

Pada bab ini dijelaskan mengenai rancangan dari sistem yang akan dibuat setelah melakukan pengumpulan data dan analisa masalah mulai dari kebutuhan software dan hardware, serta alur dari sistem yang dikembangkan.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini diuraikan hasil uji coba dari sistem yang sudah dikerjakan, seberapa tepat deteksi yang dihasilkan, dan analisis tingkat ketepatan hasil uji coba.

BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang disesuaikan dengan pertanyaan penelitian dan saran mengenai algoritma yang digunakan.

REFERENSI

Berisi tentang daftar jurnal-jurnal atau referensi mengenai Kemiripan Dokumen, *Winnowing*, *Cosine similarity*, dan *Enhanced Confix Stripping* pada beberapa topik yang terkait.