

BAB I

PENDAHULUAN

1.1. Latar Belakang

Plagiarisme menurut Peraturan Menteri Pendidikan RI Nomor 17 Tahun 2010 dikatakan bahwa: "Plagiat adalah perbuatan 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"(Purwani Istiana et al, 2010). Plagiarisme sering dianggap sebagai tindak pidana karena mencuri hak kekayaan intelektual orang lain. Di dunia lembaga pendidikan, pelaku plagiarisme atau dapat disebut plagiator dapat mendapat hukuman berat seperti dikeluarkan dari sekolah/universitas. Kasus tindakan plagiarisme atau penjiplakan ini sudah sering terjadi khususnya pada kalangan akademisi baik lingkungan sekolah menengah maupun perguruan tinggi (Firdaus et al, 2008).

Klasifikasi plagiarisme terbagi berdasarkan proporsi atau persentase kata, kalimat, atau paragraf yang dibajak yaitu : (1) Plagiarisme ringan, plagiarisme yang jumlah proporsi atau persentase kata, kalimat, maupun kalimat yang dibajak tidak melebihi 30 persen (< 30%), (2) plagiarisme sedang, plagiarisme yang jumlah proporsi atau persentase kata, kalimat, maupun paragraf yang dibajak antara 30-70 persen, (3) plagiarisme berat, plagiarisme yang jumlah proporsi atau persentase kata, kalimat, paragraf yang dibajak lebih dari 70 persen (>70%) (Afdhal et al, 2014).

Oleh karena tindakan plagiarisme secara perlahan harus ditekan sejak dini, maka untuk mengatasi hal tersebut perlu adanya suatu sistem untuk melakukan deteksi terhadap karya tulis akademik yang ada di universitas atau Lembaga Pendidikan semisalnya. Dengan melakukan perbandingan sumber karya tulis asli, suatu karya tulis dapat diperiksa apakah karya tulis itu plagiat atau tidak serta tingkat plagiasinya menggunakan metode *string matching pattern*. Terdapat banyak algoritma yang tersedia untuk melakukan hal tersebut, diantaranya adalah *Naive string matching algorithm*, *Brute Force algorithm*, *Rabin-Karp algorithm*, *Boyer-Moore algorithm*, *Knuth-Morris-Pratt algorithm*, *Aho-Corasick Algorithm* and *Commentz Walter algorithm* (Akhtar Rasool et al,2014). Tentunya algoritma-algoritma ini memiliki

kelebihan dan kekurangannya masing-masing dalam proses *string matching pattern*. *String matching pattern* itu sendiri dibagi menjadi dua dilihat dari banyaknya dokumen atau teks pembanding, yaitu *single string matching pattern* dan *multiple string matching pattern*. Untuk membuat suatu sistem pendeteksi plagiat yang baik kita memerlukan algoritma yang baik untuk jenis *multiple string matching pattern*. Algoritma yang cocok untuk permasalahan *multiple string matching pattern* yaitu algoritma Rabin-Karp dan Knuth-Morris-Pratt.

Algoritma Rabin-Karp adalah salah satu algoritma pencarian *string multi pattern* yang dikembangkan oleh Michael O. Rabin dan Richard M. Karp pada tahun 1987 menggunakan fungsi *hashing* untuk menemukan *pattern* di dalam *string* teks (N. Bansal, 2018). Algoritma Rabin-Karp bekerja dengan menggunakan fungsi *hash* sebagai pembanding antara *string* yang dicari (m) dengan *substring* pada teks (n). Apabila hasil *hash* value keduanya sama maka akan dilakukan proses perbandingan sekali lagi terhadap karakter. Algoritma Rabin-Karp berhasil menghasilkan efisiensi waktu yang baik dalam mendeteksi *string* yang memiliki lebih dari satu pola. Hal ini yang membuat algoritma Rabin-Karp umum digunakan dalam melakukan pendeteksian terhadap tindak plagiat dokumen.

Upaya untuk mengatasi masalah kemiripan tidak cukup dengan mengingatkan bahwa tindakan tersebut melanggar hak kekayaan intelektual. Pencegahan dan pendeteksian merupakan salah satu cara yang paling efektif untuk meningkatkan kualitas pendidikan dan mengurangi kemiripan dokumen satu dengan yang lain. Akan tetapi, pendeteksian sangat sulit jika dilakukan secara manual, sehingga diperlukan suatu sistem cerdas yang mampu mendeteksi kemiripan kata. Pada penelitian Surahman, disarankan untuk mengembangkan algoritma *Rabin-Karp* berbasis web (Surahman, 2013). Sehingga diperlukan sistem untuk mendeteksi karya lebih mudah dengan menggunakan sistem berbasis web. Agar mempermudah pengguna untuk mengecek karya tersebut melalui internet.

Salah satu metode dalam sistem cerdas yang dapat digunakan yaitu metode *string matching* dengan menggunakan Algoritma *Rabin-Karp*. Algoritma *Rabin-Karp* adalah *multiple pattern search* yang sangat efisien untuk mencari *string* dengan pola banyak dan melakukan proses pendeteksian kemiripan kata dengan menghasilkan nilai persentase kemiripan (Nugroho, 2011). Sistem cerdas dengan algoritma *Rabin-Karp* dalam penelitian Nugroho, mampu menampilkan hasil akurasi kemiripan yang sama

dan rata-rata proses lebih baik, terutama dokumen teks yang mempunyai *size*/ ukuran file yang besar (Nugroho, 2011). Namun dalam penelitian Nugroho, belum menunjukkan keberadaan kata yang memiliki kemiripan. Dalam penelitian yang lain, Purwitasari mampu mengindikasi kemiripan kalimat yang sama menggunakan Algoritma Winowwing (Purwitasari et al., 2011). Sehingga pada penelitian ini, mengutip penelitian yang dilakukan Nugroho ditambahkan dengan mengindikasi kalimat sama.

Berdasarkan uraian di atas, dilakukan penelitian tugas akhir berupa pembuatan sebuah sistem yang berguna untuk mendeteksi tingkat kemiripan kata pada karya tulis ilmiah menggunakan algoritma *Rabin-Karp*.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, dapat dirumuskan permasalahan yang dihadapi yaitu bagaimana mengimplementasikan algoritma *Rabin-Karp* untuk sistem deteksi kemiripan kata pada dua dokumen.

1. Bagaimana mengukur kemiripan karya tulis ilmiah menggunakan algoritma rabin-karp?
2. Bagaimana performa algoritma rabin-karp mengukur kemiripan karya tulis ilmiah?

1.3. Tujuan dan Manfaat

Tujuan dari penelitian tugas akhir ini adalah mengimplementasikan algoritma *Rabin-Karp* untuk sistem deteksi kemiripan kata pada karya tulis ilmiah.

Manfaat dari penelitian tugas akhir ini adalah untuk mempermudah dan mempercepat pengecekan terhadap karya tulis yang kemungkinan memiliki kemiripan kata dengan karya tulis dari hasil penelitian yang sudah pernah ada, dengan ketelitian yang lebih tinggi dan usaha yang lebih kecil dibandingkan pendeteksian manual.

1.4. Batasan Masalah

Ruang lingkup Batasan masalah dalam tugas akhir ini, antara lain :

1. Sistem dapat mendeteksi kemiripan kata antara dua dokumen skripsi, terdiri dari abstrak dan pendahuluan.
2. Bagian dokumen teks yang diproses adalah file teks digital yang bersifat plain teks, yaitu *file* teks yang terdiri dari huruf dan angka, tanpa mencakup gambar,

dan tabel.

3. Dokumen *input* adalah dokumen yang berformat (*.doc, *.docx).
4. Sistem hanya memproses dokumen teks berbahasa Indonesia.
5. Nilai *k* pada *parsing k-gram* ditentukan sebanyak 4 ukuran perkata.
6. Sistem ini diimplementasikan berbasis *web*, dengan bahasa pemrograman PHP (*Hypertext Preprocessor*) dan sistem basis data MySQL.

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam tugas akhir ini terbagi menjadi beberapa pokok bahasan, yaitu :

BAB I PENDAHULUAN

Bab ini berisi uraian tentang latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan tugas akhir mengenai sistem deteksi kemiripan kata pada dua dokumen menggunakan Algoritma *Rabin-Karp*.

BAB II DASAR TEORI

Bab ini berisi penjelasan singkat konsep-konsep yang mendukung pengembangan sistem, meliputi Plagiarisme, Metode Pencocokkan *String*, *Rolling Hash*, Tahap *Text Preprocessing*, *Parsing K-Gram*, Persentase Kemiripan, Tahap Pengembangan Sistem, *Data Flow Diagram* (DFD),

Entity Relationship Diagram (ERD), Flowchart, MySQL, PHP, dan Pengukuran dan Kesalahan.

BAB III SPESIFIKASI, ANALISIS, DAN PERANCANGAN

Bab ini membahas proses pengembangan sistem pada tahap gambaran umum sistem, spesifikasi dan analisis kebutuhan sistem, dengan hasilnya berupa desain dan rancangan sistem yang dikembangkan.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas hasil pengembangan sistem pada tahap implementasi dan menerangkan rincian pengujian sistem yang dibangun dengan metode *black box*.

BAB V PENUTUP

Bab ini berisi kesimpulan yang diambil berkaitan dengan sistem yang dibangun dan saran untuk pengembangan lebih lanjut.

