

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kemajuan teknologi informasi dan komunikasi tidak hanya membawa dampak positif, tetapi juga membawa dampak negatif, salah satunya adalah tindakan plagiarisme (Kharisman, 2012). Plagiarisme dalam Kamus Besar Bahasa Indonesia (KBBI) adalah penjiplakan atau pengambilan karangan (pendapat dan sebagainya) orang lain dan menjadikannya seolah-olah karangan (pendapat dan sebagainya) sendiri (KBBI, 2002).

Plagiarisme adalah suatu kejahatan akademik (*academic criminal*) dan aib yang sangat tidak mudah terhapuskan (Sugiyanto, 2010). Di kalangan mahasiswa, kegiatan plagiarisme sering ditemukan dalam tugas, makalah bahkan skripsi. Ini terjadi karena mahasiswa salah memanfaatkan fasilitas *copy paste* dari suatu aplikasi komputer. Untuk mengatasi hal tersebut, maka perlu adanya suatu aplikasi komputer yang dapat mendeteksi tindakan plagiarisme agar dapat mengurangi kegiatan plagiarisme.

Sejumlah penelitian telah dilakukan dalam membangun sebuah aplikasi pendeteksian plagiarisme dokumen diantaranya penelitian tentang *Perancangan Sistem Deteksi Plagiarisme Dokumen Teks dengan Menggunakan Algoritma Rabin-Karp* oleh Eko Nugroho (2011), *Deteksi Kemiripan Isi Dokumen Teks Menggunakan Algoritma Lavenstein Distance* oleh Hendri Winoto (2012) dan *Rancang Bangun Aplikasi Pendeteksi Penjiplakan Dokumen Menggunakan Algoritma Biword Winnowing* oleh Muhammad Ridho (2013).

Pada penelitian Eko Nugroho (2011) dilakukan penelitian bagaimana membuat suatu sistem yang dapat melakukan pendeteksian teks. Algoritma yang digunakan adalah algoritma *Rabin-Karp*. Algoritma *Rabin-Karp* adalah algoritma pencocokan string yang menggunakan fungsi *hash* sebagai pembanding antara *string* yang dicari ( $m$ ) dengan *substring* pada teks ( $n$ ). Algoritma *Rabin-Karp* pada penelitian tersebut dibagi menjadi dua, yaitu algoritma *Rabin-Karp* yang asli dan algoritma *Rabin-Karp* modifikasi. Penelitian ini bertujuan untuk mengetahui

perbandingan hasil nilai *similarity* dan waktu proses menggunakan algoritma *Rabin-Karp* asli dan *Rabin-Karp* modifikasi. Modifikasi yang dilakukan pada algoritma *Rabin-Karp* ini adalah dengan menyisipkan metode *stemming* dengan menggunakan algoritma Arifin-Setiono pada tahap *preprocessing*-nya dan melakukan modifikasi pada saat proses *hashing* serta perubahan pada proses *string-matching*. Hasil yang diperoleh bahwa aplikasi tersebut bisa digunakan untuk pendeteksi plagiarisme, algoritma *Rabin-Karp* asli dan algoritma *Rabin-Karp* modifikasi mempunyai akurasi nilai *similarity* yang relatif sama, tetapi algoritma *Rabin-Karp* modifikasi mempunyai rata-rata waktu proses yang lebih baik, terutama dokumen teks yang mempunyai *size/ukuran file* yang besar.

Pada penelitian Hendri Winoto (2012) dilakukan penelitian bagaimana menerapkan algoritma *levenshtein distance* dalam mendeteksi kemiripan isi dokumen teks. *Levenshtein distance* merupakan metrik yang digunakan untuk mengukur keterbedaan jarak antara dua sekuens. *Levenshtein distance* antara dua *string* ditentukan berdasarkan jumlah minimum perubahan/pengeditan yang diperlukan untuk melakukan transformasi dari satu bentuk *string* ke bentuk *string* yang lain. Algoritma *levenshtein distance* yang digunakan dibagi menjadi dua yaitu algoritma *levenshtein distance standart* (asli) dan algoritma *levenshtein distance preprocessing (filtering, stemming dan sorting)*. Penelitian ini dilakukan untuk mengetahui perbandingan kemiripan dokumen yang diperoleh dari *levenshtein distance standart* dan *levenshtein distance preprocessing*. Hasil yang diperoleh aplikasi tersebut dapat digunakan untuk deteksi plagiarisme dan algoritma *levenshtein distance preprocessing* memiliki nilai kemiripan tinggi dibandingkan *levenshtein distance standart*.

Pada penelitian Muhammad Ridho (2013) dilakukan penelitian tentang bagaimana merancang dan membangun sebuah aplikasi pendeteksi plagiarisme dokumen menggunakan algoritma *winnowing* dengan pendekatan *biword*. Algoritma *winnowing* merupakan algoritma yang digunakan dalam deteksi penjiplakan termasuk bagian-bagian kecil yang mirip dalam dokumen yang berjumlah banyak. Input dari algoritma ini adalah dokumen teks yang diproses sehingga menghasilkan *output* berupa kumpulan nilai-nilai *hash*, kumpulan-kumpulan nilai *hash* tersebut selanjutnya disebut *fingerprint*. *Fingerprint* inilah

yang dijadikan dasar pembandingan antara *file-file* teks yang telah dimasukkan dan digunakan dalam deteksi penjiplakan. Algoritma *winnowing* pada penelitian ini dilakukan modifikasi, yaitu penggunaan *biword* yang menggantikan metode *k-gram* dan adanya penggunaan fungsi *MD5* untuk proses enkripsi frasa sebelum dilakukan *rolling hash*. Hasil yang diperoleh bahwa aplikasi ini dapat mendeteksi kemiripan antar dokumen dan informatif karena dapat menampilkan teks yang sama antar dokumen berdasarkan nilai *fingerprint* yang dihasilkan.

Dari penelitian Eko Nugroho (2011), Hendri Winoto (2012), dan Muhammad Ridho (2013) telah berhasil membangun aplikasi pendeteksi plagiarisme. Namun aplikasi pendeteksi plagiarisme tersebut hanya sebatas mendeteksi kemiripan dari dua buah dokumen saja. Belum termasuk mendeteksi kemiripan dari satu dokumen dengan banyak dokumen dan mendeteksi dimana dokumen sumber dari plagiarisme tersebut. Untuk mendeteksi dokumen sumber tersebut, pada dasarnya kita bisa memanfaatkan sistem *information retrieval* (IR). Sistem IR adalah *system* yang digunakan untuk menemukan kembali (*retrieve*) informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis (Bunyamin dan Negara, 2008). Isi dokumen yang diduga plagiarisme dijadikan sebagai *query* yang nanti akan diproses oleh sistem IR dan akan dikembalikan dokumen-dokumen sumber yang relevan sesuai dengan isi dokumen yang diduga plagiarisme tersebut. Dari cara tersebut, timbul permasalahan jika isi teks dalam dokumen tersebut panjang, tentu *query* yang akan digunakan panjang pula. Hal ini tentu akan berpengaruh terhadap hasil pencarian yang diperoleh. Untuk itu perlu adanya algoritma yang dapat melakukan penyeleksian terhadap isi dokumen, mencari bagian mana yang cocok dan tepat mewakili isi dokumen untuk dijadikan *query* pencarian.

Pada penelitian kali ini akan dilakukan riset bagaimana membentuk *query* dari isi teks dokumen yang panjang diduga plagiarisme dengan menggunakan metode *fingerprint biword winnowing*. *Query* yang terbentuk akan dicobakan pada sistem IR dengan model Okapi BM25. Model Okapi BM25 merupakan salah satu model IR yang dinilai memiliki kemampuan yang baik dalam *me-retrieve* dokumen yang relevan (Bintana, 2012). Kemudian dokumen yang diduga plagiarisme tersebut akan dilakukan perbandingan isi dengan banyak dokumen

sumber dari hasil pencarian yang diperoleh untuk mengetahui tingkat kemiripannya dengan menggunakan algoritma *biword winnowing*. Algoritma *biword winnowing* dipilih karena memiliki hasil yang informatif, yakni dapat menampilkan teks yang sama antar dokumen berdasarkan nilai *fingerprint* yang dihasilkan.

Dengan adanya penelitian “*Source Detection* pada Kasus *Plagiarisme* Dokumen Menggunakan Metode *Biword Winnowing* dan *Retrieval* Berbasis *Okapi BM25*” ini, diharapkan mampu membentuk *query* dari dokumen yang kemudian dapat digunakan untuk mendeteksi kegiatan plagiarisme dan mencari sumber dokumennya sehingga dapat mengurangi kegiatan plagiarisme.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang permasalahan di atas, maka dirumuskan masalah penelitian ini, yaitu : “Bagaimana membentuk *query* dari dokumen yang memiliki isi teks panjang yang diduga melakukan plagiarisme dengan menggunakan *fingerprint* algoritma *biword winnowing* yang kemudian *query* tersebut akan dicari dokumen sumbernya menggunakan Sistem IR model Okapi BM25”.

## **1.3 Batasan Masalah**

Adapun yang menjadi batasan masalah dalam laporan tugas akhir ini adalah:

1. Dokumen yang digunakan merupakan dokumen digital yang berisi teks, tidak termasuk gambar.
2. Dokumen yang digunakan menggunakan bahasa Indonesia.
3. Sinonim diabaikan.
4. Menggunakan algoritma *stemming* nazief dan adriani

## 1.4 Tujuan

Adapun tujuan yang ingin dicapai oleh penulis dari penelitian dan penyusunan tugas akhir ini adalah:

1. Menghasilkan suatu aplikasi yang dapat membentuk *query* dari dokumen yang memiliki teks panjang yang diduga melakukan plagiarisme dengan menggunakan *fingerprint* algoritma *biword winnowing* yang kemudian *query* tersebut dapat digunakan untuk mencari dokumen sumber menggunakan sistem IR model Okapi BM25.
2. Dapat melakukan perbandingan kemiripan isi dokumen yang diduga melakukan plagiarisme dengan banyak dokumen sumber yang diperoleh menggunakan algoritma *biword winnowing*.

## 1.5 Sistematika Pembahasan

Sistematika penulisan laporan Tugas Akhir ini terdiri dari pokok-pokok permasalahan yang dibahas pada masing-masing yang diuraikan menjadi beberapa bagian :

### Bab I. Pendahuluan

Bab ini membahas tentang gambaran umum isi tugas akhir yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan.

### Bab II. Landasan Teori

Bab ini menjelaskan tentang teori-teori yang berhubungan dengan penelitian yang terdiri dari penjelasan mengenai konsep plagiarisme, konsep *information retrieval*, model Okapi BM25 dan algoritma *biword winnowing*.

### Bab III. Metodologi Penelitian

Bab ini membahas langkah-langkah yang dilaksanakan dalam proses penelitian, yaitu identifikasi masalah, merumuskan masalah, *study literatur*, hipotesa, analisa sistem, perancangan sistem, implementasi, pengujian sistem dan kesimpulan dan saran.

### Bab IV. Analisa Dan Perancangan Sistem

Berisi pembahasan mengenai analisa sistem meliputi analisa sistem dan perancangan pada aplikasi pendeteksi dokumen sumber.

#### Bab V. Implementasi

Dalam bab ini membahas tentang implementasi aplikasi pendeteksi dokumen sumber, serta pengujian dan evaluasi

#### Bab VI. Penutup

Dalam bab ini akan dijelaskan mengenai beberapa kesimpulan yang didapatkan dari pembahasan pembuatan aplikasi pendeteksi dokumen sumber dari kasus plagiarisme dengan metode *biword winnowing* dan *retrieval* berbasis Okapi BM25 disertai beberapa saran sebagai hasil akhir dari penelitian yang telah dilakukan.