

# BAB I. PENDAHULUAN

## 1.1 Latar Belakang

Saat ini banyak sekali dokumen teks seperti jurnal maupun karya ilmiah yang tersebar di internet dan semua orang dapat dengan mudah mengaksesnya. Dengan kondisi seperti itu menyebabkan semakin meningkatnya tindakan plagiarisme. Plagiarisme merupakan tindakan tidak jujur yang disebabkan karena kurangnya kreatifitas dan inovasi dalam menciptakan karya yang asli atau orisinal (Leman et al., 2019). Terdapat plagiarisme lintas bahasa di mana dokumen asli diterjemahkan ke bahasa lain dan menyalin isi dari dokumen tersebut (Roostae et al., 2020).

Plagiarisme lintas bahasa adalah menyalin suatu karya milik orang lain dari bahasa yang berbeda dan diterjemahkan ke bahasa lain (Agarwal, 2019). Dalam mendeteksi plagiarisme lintas bahasa membutuhkan komponen tambahan untuk menerjemahkan setiap kalimat. Hasil terjemahan harus akurat agar hasil akhir deteksi kemiripan antar dokumen optimal. Seperti penelitian yang dilakukan oleh Putri Ratna yang menggunakan Googletrans API untuk menerjemahkan dokumen teks bahasa Inggris ke dalam bahasa Indonesia (Putri Ratna et al., 2019).

Dalam mendeteksi plagiarisme karya ilmiah biasanya dilakukan dengan cara manual yaitu mencocokkan karya ilmiah yang baru dengan karya ilmiah yang sudah ada sehingga hal tersebut sangat tidak efisien (Pardede & Alvian, 2015). Oleh karena itu dibutuhkan sebuah sistem pendeteksi plagiarisme dokumen teks lintas bahasa. Salah satu metode yang dapat digunakan untuk mendeteksi kemiripan dokumen teks yaitu metode *Winnowing* (Alamsyah et al., 2019).

Metode *Winnowing* merupakan metode di mana inputan teks akan diproses sehingga menghasilkan kumpulan nilai *hash* yang disebut dengan *fingerprint* (Ilham & Pasnur, 2017). Nilai *hash* merupakan nilai numerik yang didapatkan dari perhitungan ASCII (Priantara et al., 2016). Hasil akurasi metode *Winnowing* lebih baik dibandingkan dengan metode rabin karp yaitu menghasilkan tingkat akurasi 88,89% dan waktu pemrosesan 0,13 detik sedangkan pada metode rabin karp menghasilkan tingkat akurasi 37,50% dengan waktu pemrosesan 0,19 detik (Sugiono et al., 2018).

Sebelum mendapatkan hasil *fingerprint*, terdapat proses pembentukan *n-gram* dan *w-gram*. *N-gram* merupakan hasil potongan karakter huruf sebanyak *n* dari sebuah kalimat yang diinputkan. Sedangkan *w-gram* digunakan untuk membentuk *substring* dari nilai *hash* (Wibowo & Hastuti, 2016). Besar nilai *n-gram* dan *w-gram* yang digunakan untuk mendeteksi kemiripan dokumen teks sangat berpengaruh pada hasil akurasi kemiripan. Seperti pada penelitian yang dilakukan oleh Sunardi menunjukkan bahwa jika semakin besar nilai *n-gram* dan *w-gram* maka nilai akurasi kemiripannya semakin rendah dan proses yang dihasilkan lebih cepat. Sedangkan jika nilai *n-gram* dan *w-gram* semakin rendah maka nilai akurasi kemiripan akan semakin tinggi dan proses yang diperlukan semakin lama (Sunardi, A. Yudhana, 2017).

Dari penjelasan di atas, maka mengusulkan sebuah sistem deteksi kemiripan antar dokumen teks lintas bahasa menggunakan metode *Winnowing* karena metode *Winnowing* memiliki tingkat akurasi lebih baik dibandingkan dengan metode rabin karp. Bahasa yang dapat dideteksi pada dokumen teks adalah bahasa Inggris dan Indonesia. Dokumen teks yang berbahasa Inggris akan diterjemahkan ke dalam bahasa Indonesia dan selanjutnya akan dilakukan *preprocessing*, membentuk nilai *n-gram*, menghitung nilai *hash*, membentuk *window* dari *hash*, membentuk *fingerprint* dan menghitung nilai persentase kesamaan menggunakan *Jaccard Coefficient*. Oleh karena itu dibuatlah sebuah skripsi berjudul “Sistem Pendeteksi Plagiarisme Dokumen Teks Lintas Bahasa Menggunakan Metode *Winnowing*”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang dapat diambil adalah sebagai berikut:

- a. Bagaimana membangun sistem pendeteksi plagiarisme dokumen teks yang dapat memproses teks berbeda bahasa?
- b. Bagaimana tingkat keberhasilan metode *Winnowing* dalam penerapan sistem pendeteksi plagiarisme dokumen teks lintas bahasa?

### 1.3 Tujuan

Tujuan dari dilakukannya skripsi dengan judul “Sistem Pendeteksi Plagiarisme Dokumen Teks Lintas Bahasa Menggunakan Metode *Winnowing*”, adalah sebagai berikut:

- a. Membangun sistem pendeteksi plagiarisme dokumen teks yang dapat memproses teks berbeda bahasa yaitu bahasa Inggris dan bahasa Indonesia
- b. Mengetahui tingkat keberhasilan metode *Winnowing* saat diterapkan dalam sistem pendeteksi plagiarisme dokumen teks lintas bahasa

### 1.4 Batasan Masalah

Permasalahan yang dibahas pada penelitian ini memiliki beberapa batasan sebagai berikut:

- a. Dokumen teks yang diuji berfokus pada jurnal dan prosiding atau seminar
- b. Dokumen teks yang diuji berupa inputan teks dan *file* PDF
- c. Ukuran maksimum *file* PDF yang dapat diunggah adalah 2 MB
- d. Jumlah halaman *file* PDF maksimum yang dapat diproses adalah 45 halaman
- e. Kalimat yang berada pada tanda petik, sitasi dan referensi tidak diikutsertakan dalam pengujian kemiripan dokumen
- f. Bahasa yang dapat dideteksi oleh sistem adalah bahasa Inggris dan bahasa Indonesia
- g. Bahasa yang digunakan adalah bahasa Inggris dan bahasa Indonesia
- h. Dokumen yang berbahasa Inggris akan diterjemahkan terlebih dahulu ke dalam bahasa Indonesia
- i. Proses pendeteksi plagiarisme dilakukan dalam ranah bahasa Indonesia
- j. Hasil kemiripan teks ditampilkan dalam bentuk persentase

### 1.5 Sistematika Penulisan

#### **BAB 1                   PENDAHULUAN**

Bab ini membahas latar belakang, rumusan masalah, tujuan, batasan masalah dan sistematika penulisan.

#### **BAB II                  LANDASAN TEORI**

Bab ini berisi tentang dasar teori yang relevan yang melengkapi latar belakang serta tinjauan pustaka yang telah dibaca selama pencarian solusi terhadap masalah yang diangkat dalam tugas akhir.

### **BAB III           METODOLOGI PENELITIAN**

Bab ini berisi penjelasan mengenai waktu dan tempat penelitian, metode pengumpulan data, metode pengolahan data, desain sistem serta uji coba sistem.

### **BAB IV           ANALISIS DAN PERANCANGAN SISTEM**

Bab ini berisi penjelasan mengenai deskripsi sistem, analisis pengguna, analisis fungsional dan perancangan detail sistem. Pada perancangan sistem terdapat diagram yang dibutuhkan untuk membangun sistem pendeteksi plagiarisme dokumen teks lintas bahasa.

### **BAB V            IMPLEMENTASI DAN PENGUJIAN**

Bab ini berisi penjelasan implementasi yang dijabarkan pada bab sebelumnya dan pengujian yang dilakukan menggunakan metode *Winnowing*.

### **BAB VI           HASIL DAN PEMBAHASAN**

Bab ini berisi penjelasan hasil dan pembahasan dari pelaksanaan pengujian yang sudah dilakukan pada bab analisis dan pengujian.

### **BAB VII          KESIMPULAN DAN SARAN**

Bab ini berisi tentang kesimpulan dan saran berdasarkan hasil dan pembahasan.

### **DAFTAR PUSTAKA**

Pada bagian ini berisikan sumber-sumber yang digunakan sebagai pendukung pada kajian literatur.