

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi informasi saat ini semakin maju membuat dokumen yang tadinya bersifat fisik kini telah banyak dibuat dalam bentuk digital sehingga dapat dengan mudah dilakukan penyalinan digital yang dapat menimbulkan plagiarisme. Rawannya tindakan *plagiarisme* pada dokumen digital mendorong para peneliti untuk mengembangkan piranti lunak *plagiarism checker* untuk mendeteksi plagiarisme dengan cara mengukur tingkat kemiripan dokumen tersebut dengan dokumen-dokumen lainnya dengan menggunakan berbagai algoritma. Dalam mendeteksi plagiarisme dokumen digital terdapat beberapa metode yang dapat digunakan untuk mengukur tingkat similaritas sebuah dokumen yaitu metode perbandingan teks lengkap, metode kesamaan kata kunci dan metode *fingerprinting* (Hermawan, 2015). Pada penelitian ini akan dilakukan pengukuran similaritas menggunakan metode *fingerprinting*. Metode *fingerprinting* adalah metode yang menelusuri karakter satu persatu pada deret karakter. Prinsip kerja dari metode dokumen *fingerprinting* ini adalah dengan menggunakan teknik *hashing*. Kelebihan dari metode *fingerprinting* adalah waktu proses lebih cepat dibandingkan dari metode perbandingan teks lengkap dan metode kesamaan kata kunci (Hermawan, 2015).

Beberapa algoritma yang termasuk dalam metode *fingerprinting* yaitu Algoritma Rabin Karp, Algoritma Winnowing dan Algoritma Manber. Pada penelitian ini algoritma yang digunakan adalah algoritma Winnowing, karena

menurut penelitian sebelumnya yang dilakukan oleh (Hermawan, 2015) mengenai analisis performansi algoritma Winnowing dan algoritma Manber untuk deteksi kesamaan dokumen teks berbahasa Indonesia menyimpulkan bahwa algoritma Winnowing mempunyai akurasi yang lebih baik. Kemudian penelitian yang dilakukan oleh (Alamsyah, 2017) mengenai perbandingan algoritma Winnowing dengan algoritma Rabin-karp untuk mendeteksi plagiarisme pada kemiripan teks judul skripsi menyimpulkan bahwa pendekatan algoritma Winnowing lebih baik karena menghasilkan tingkat persentase yang lebih kecil dan waktu proses yang lebih cepat.

Kemiripan dokumen dapat ditingkatkan dengan mengubah kata-kata yang mengandung imbuhan menjadi kata dasarnya. Pada ilmu komputer, teknik ini dikenal sebagai *stemming* yaitu suatu teknik untuk mengekstraksi kata dasar dari suatu kata (Utomo and Winarko, 2011). Penelitian sebelumnya yang menerapkan *stemming* pada algoritma Winnowing dilakukan oleh (Ferari, 2014) mengenai implementasi algoritma winnowing dan porter *stemmer* mendeteksi kemiripan dua dokumen berbasis web, hasil penelitiannya menghasilkan bahwa algoritma *stemmer* porter membantu mempercepat algoritma winnowing dalam menentukan nilai *fingerprint* dari suatu teks. Selanjutnya, pengaruh algoritma *stemming* Nazief-adriani terhadap kinerja algoritma Winnowing untuk mendeteksi plagiarisme bahasa Indonesia oleh (Nugroho, 2017), hasil penelitiannya menghasilkan proses *stemming* pada algoritma Winnowing cenderung menurunkan tingkat *similarity* yang dicapai, namun mempercepat *processing time* kurang lebih sebesar 30%.

Sedangkan untuk perbandingan pengaruh *stemmer* porter dan Nazief-Adriani terhadap kinerja algoritma Winnowing belum pernah dilakukan.

Pada penelitian ini akan dilakukan perbandingan pengaruh dari *stemmer* Porter dan Nazief-Adriani terhadap kinerja algoritma Winnowing. Untuk pengimplementasian penelitian ini yaitu menggunakan aplikasi berbasis web yang dibangun menggunakan Bahasa pemrograman PHP (*Hypertext Preprocessor*), Basis Data menggunakan Mysql, Web Servernya yaitu Apache dan Permodelan sistem yang digunakan yaitu UML (*Unified Modeling Language*).

Berdasarkan studi ini, diharapkan dapat diketahui perbandingan pengaruh *stemmer* terhadap similaritas, kecepatan proses dan *fingerprints* yang dihasilkan algoritma Winnowing dalam mengukur plagiarisme, sehingga dapat menentukan pengaruh *stemmer* mana yang lebih baik untuk algoritma Winnowing.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang permasalahan, permasalahan yang akan diselesaikan dalam penelitian ini yaitu bagaimana membandingkan pengaruh dari *stemmer* Porter dan Nazief-Adriani terhadap kinerja algoritma Winnowing berdasarkan similaritas, kecepatan proses dan *fingerprint* yang dihasilkan dalam mengukur *plagiarisme*?

## **1.3 Batasan Masalah**

Batasan masalah dalam penelitian ini adalah:

1. Dokumen yang digunakan pada penelitian ini adalah dokumen teks tidak meliputi dokumen yang berisi gambar.

2. Dokumen yang di gunakan pada penelitian kali ini adalah dokumen teks bahasa indonesia dan berformat pdf.
3. *Stemming* yang digunakan adalah untuk Bahasa Indonesia.

#### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini yaitu mengetahui informasi hasil perbandingan pengaruh *stemmer* Porter dan Nazief-Adriani terhadap kinerja algoritma Winnowing berupa similaritas, kecepatan proses dan *fingerprints* yang dihasilkan.

#### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah untuk mengetahui hasil perbandingan pengaruh dari *stemmer* Porter dan Nazief-Adriani terhadap kinerja algoritma Winnowing. Sehingga dapat menentukan *stemmer* mana yang baik untuk algoritma Winnowing.

#### **1.6 Metode Penelitian**

Metode yang dilakukan adalah metodologi studi komparasi, dimana setiap algoritma di uji dari segi kinerja algoritma. Kemudian hasil dari pengujian dibandingkan berdasarkan similaritas, waktu proses dan *fingerprints* yang dihasilkan. Tahap-tahap yang digunakan yaitu:

- a. Studi literatur

Pengumpulan referensi yang dijadikan acuan dalam melakukan penelitian.

- b. Perancangan Perangkat Lunak

Pada tahap ini dilakukan analisis kebutuhan perangkat lunak, permodelan dari perangkat lunak dan rancangan tampilan antarmuka perangkat lunak.

c. Implementasi

Pada tahap ini mengimplementasikan tampilan antarmuka sistem, proses *preprocessing* dan penerapan algoritma pada *source code* program.

d. Pengujian dan Perbandingan

Pada tahap ini melakukan pengujian terhadap parameter similaritas, akurasi, kecepatan proses dan fingerprint yang dihasilkan, kemudian melakukan perbandingan dari hasil pengujiannya.

e. Menarik Kesimpulan

Pada tahap ini dilakukan penarikan kesimpulan dari hasil perbandingan.

## 1.7 Sistematika Penulisan

Adapun sistematika penulisan tugas akhir ini adalah sebagai berikut

### **BAB I PENDAHULUAN**

Bab ini berisi pembahasan masalah umum, yang merupakan gambaran dan garis besar tentang isi laporan. Yang didalamnya memuat latar belakang dilakukannya penelitian, rumusan masalah, batasan permasalahan pada penelitian, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan laporan tugas akhir.

### **BAB II LANDASAN TEORI**

Bab ini berisi pembahasan teori - teori yang saling berhubungan dengan penelitian dan pembuatan sistem. Mulai dari penjabaran mengenai metode yang digunakan dan teori lainnya yang saling berhubungan serta ulasan mengenai penelitian – penelitian sebelumnya.

**BAB III METODOLOGI**

Berisi metodologi dan langkah-langkah selama mengerjakan penelitian

**BAB IV HASIL DAN PEMBAHASAN**

Bab ini berisi hasil dan pembahasan dari penelitian (missal: rancangan aplikasi, implementasi algoritma, hasil pengujian, hasil perbandingan, dll)

**BAB V KESIMPULAN DAN SARAN**

Bab ini merupakan bab akhir dari penulisan laporan yang berisi mengenai simpulan-simpulan yang merupakan hasil analisis kinerja pada bagian sebelumnya serta saran yang perlu diperhatikan berdasarkan keterbatasan yang ditemukan dan asumsi-asumsi yang dibuat selama penelitian.