



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Seiring dengan kemajuan zaman, teknologi tidak dapat dipisahkan dari dampak negatifnya. Ancaman plagiarisme berlaku terutama di dunia dimana teknologi informasi telah mempermudah melakukan penyalinan informasi (Cebrian, Dkk., 2007). Dokumen elektronik yang dapat diunduh dari internet dengan mudah dapat dilakukan perubahan. Dokumen elektronik kemudian dapat disimpan dengan menggunakan nama baru dari pengguna yang baru. Modifikasi dari informasi tanpa mencantumkan sumbernya melanggar aturan dan disebut sebagai plagiarisme (Putri & Siahaan, 2017). Plagiarisme adalah penjiplakan yang melanggar hak cipta (KBBI, 2019). Plagiarisme dapat membunuh kreativitas seseorang dalam mengembangkan ide baru.

Salah satu cara untuk mencegah plagiarisme adalah dengan bantuan dari metode pencocokan string. Metode pencocokan string dapat dimodifikasi untuk menganalisa teks, gambar bahkan suara. Penelitian ini bertujuan untuk mencocokkan sebuah dokumen elektronik tertulis dengan menggunakan algoritma Rabin-Karp. Algoritma Rabin-Karp didasari dengan teknik *hashing*. Kata-kata diubah menjadi sebuah nilai hash. Nilai-nilai tersebut, dilakukan pencocokan *string*. Berdasarkan jurnal Fuyao & Qingwei (2009) menyatakan bahwa algoritma Rabin-Karp menyediakan kemampuan untuk melakukan pencocokan string yang mengungguli algoritma dengan pendekatan berbasis prefiks. Ketika jumlah data terbilang cukup besar algoritma Rabin-Karp jauh mengungguli algoritma Knuth-Morris-Pratt dalam

hal efisiensi dan kecepatan dalam melakukan pencocokkan (Chengguo & Hui, 2012).

Data berupa *string* yang diperoleh tidak dapat langsung dicocokkan menggunakan Rabin-Karp. Pada dasarnya data yang belum diproses bersifat tidak beraturan, tidak konsisten dan tidak relevan sehingga data tidak dapat langsung digunakan untuk tujuan *mining* atau analisis maka dari itu data perlu melalui tahap *preprocessing* (Dwivedi & Rawat, 2015). Data *preprocessing* mengacu pada serangkaian aktifitas yang dilakukan untuk membuat data mentah menjadi sesuai untuk proses selanjutnya (Dwivedi & Rawat, 2015). Oleh karena itu, penelitian ini mengimplementasikan proses *preprocessing* data sebelum melakukan pencocokan *string* menggunakan algoritma Rabin-Karp. *Preprocessing* data memastikan kualitas dari sumber data, sangat meningkatkan akurasi dan performa dari data (Ang, Dkk., 2010). *Preprocessing* yang dilakukan berupa, penghilangan karakter unik dan spasi berlebih, pengubahan menjadi huruf kecil, *tokenization*, dan *stopwords removal*. Diharapkan pada akhir penelitian ini hasil jadi program dapat membantu dalam memberikan pertimbangan bagi pengguna untuk menentukan tingkat kesamaan dokumen yang dibandingkan berdasarkan nilai persentase kesamaan yang diperoleh oleh program.

## 1.2 Rumusan Masalah

Rumusan masalah penelitian ini adalah

1. Bagaimana mengimplementasikan algoritma Rabin-Karp untuk mendeteksi plagiarisme pada dokumen elektronik tertulis Bahasa Indonesia?

2. Bagaimana nilai *precision*, *accuracy* dan *recall* program hasil implementasi algoritma Rabin-Karp?

### 1.3 Batasan Masalah

Berdasarkan rumusan masalah, batasan masalah dijabarkan sebagai berikut.

1. Informasi yang dihasilkan status plagiarisme antara 2 dokumen yang dicocokkan.
2. Dokumen yang dicocokkan antara 500 hingga 1000 kata.
3. Extensi file yang digunakan adalah .txt, .pdf, .doc, dan .docx.

### 1.4 Tujuan Penelitian

Sesuai dengan rumusan masalah di atas, tujuan dari penelitian ini adalah

1. Algoritma Rabin-Karp dapat diimplementasikan untuk mendeteksi plagiarisme pada dokumen tertulis Bahasa Indonesia.
2. Nilai *precision*, *accuracy* dan *recall* program hasil implementasi algoritma Rabin-Karp dapat dihitung.

### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk mendeteksi plagiat melalui hasil persentase yang dihasilkan dalam dokumen teks elektronik Bahasa Indonesia.

### 1.6 Sistematika Penulisan

Sistematika penulisan laporan skripsi sebagai berikut.

## BAB I Pendahuluan

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

## BAB II Tinjauan Pustaka

Berisi teori mengenai plagiarisme, Rabin-Karp, dan *Confusion Matrix*.

## BAB III Metodologi dan Perancangan Aplikasi

Berisi mengenai metodologi penelitian seperti telaah literatur, analisis kebutuhan system, desain sistem, proses pemrograman sistem, *testing* dan *debugging*, uji coba system dan evaluasi serta konsultasi penulisan dengan dosen pembimbing. Perancangan aplikasi berupa *flowchart* dan desain antarmuka.

## BAB IV Implementasi dan Uji Coba

Berisi mengenai Implementasi dari perancangan pada bab 3 dan hasil uji coba program. Implementasi terbagi menjadi implementasi algoritma Rabin-Karp beserta *preprocessing* dan implementasi antar muka program.

## BAB V Simpulan dan Saran

Berisi mengenai kesimpulan dari keseluruhan penelitian ini berdasarkan tujuan yang ingin dicapai beserta saran untuk penelitian selanjutnya.

U M N  
U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A