

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Seiring dengan perkembangan teknologi yang sudah sangat maju ini dalam bidang komputer kita tidak perlu lagi menyimpan dokumen secara hardcopy dan disimpan dalam lemari lalu termakan oleh waktu, sekarang semua file sudah berubah menjadi softcopy yang mana bisa diakses lewat komputer ataupun lewat smartphone masing-masing. Namun entah dikarenakan kesibukan ataupun kelalaian dan mungkin karena terlalu teknologi terlalu praktis maka plagiarisme inipun terjadi, banyak yang melakukan penjiplakan dan tidak dicantumkan sumbernya sehingga terjadi pelanggaran hak cipta. baik dalam akademik yang mana banyak mahasiswa ataupun para siswa yang mengerjakan tugas dengan hanya mengcopy tugas dari temannya tanpa adanya perubahan. Oleh karena itu untuk memudahkan para guru dan dosen dalam bidang akademik, dibutuhkan pengidentifikasian pada dokumen dokumen tersebut agar diketahui apakah dokumen tersebut termasuk plagiat dan tidak memakan waktu yang lama dalam pemeriksaan dan penilaian. Dengan adanya sistem identifikasi tingkat kemiripan dokumen teks ini, keaslian dari dokumen tetap terjaga. Karena saat ini sangat mudah dilakukan *plagiarisme*.

Plagiarisme dapat dipahami sebagai tindakan mengambil pernyataan atau mencuri ide orang lain, dalam hukum positif, hak paten dan kekayaan intelektual sudah ada diatur dalam undang-undang, sedangkan dalam hukum islam sendiri hak kekayaan intelektual termasuk kedalam (huquq maliyyah) yang mendapatkan perlindungan hukum (mashun). MUI menegaskan kalau segala bentuk dalam pelanggaran atas hak kekayaan intelektual merupakan kezaliman dan haram hukumnya. Allah berfirman dalam sudah Asy-Syu'ara' Ayat 183:

وَلَا تَبْخَسُوا النَّاسَ أَشْيَاءَهُمْ وَلَا تَعْثَوْا فِي الْأَرْضِ مُفْسِدِينَ

Artinya: “Dan janganlah kamu merugikan manusia pada hak-haknya dan janganlah kamu merajalela dimuka bumi dengan membuat kerusakan”.

Oleh karena itu, dibutuhkan adanya pendeteksian dalam dokumen atau karya tulis agar diketahui apakah termasuk plagiat atau tidak, dengan mendeteksi kemiripan antar dokumen maka keaslian dalam tiap dokumen atau karya tulis bisa tetap terjaga keasliannya, karena saat ini mudah sekali dilakukan plagiarisme.

Dalam penelitian ini akan menggunakan dua algoritma yaitu Winnowing dan Ratcliff/Obershelp. yang mana Algoritma winnowing adalah algoritma yang digunakan untuk melakukan pemeriksaan kemiripan kata (document fingerprinting) untuk mengidentifikasi plagiarisme. Algoritma yang digunakan untuk mencari hash pada Winnowing adalah Rolling Hash. Nilai hash adalah nilai numerik yang terbentuk dari perhitungan ASCII dari setiap karakter atau angka yang ada. Pada penelitian ini akan dilakukan penentuan kemiripan dokumen dengan menggunakan fungsi hash pada algoritma Winnowing. Sedangkan algoritma Ratcliff/Obershelp juga untuk mendeteksi adanya kemiripan namun dalam prinsip kerjanya dilakukan pengembalian nilai yang dapat digunakan sebagai persentase dalam menunjukkan kesamaan dua string dengan memerhitungkan jumlah karakter yang terdapat pada kedua string tersebut.

Adapun beberapa penelitian terdahulu yang telah dilakukan dan berkaitan dengan dengan penelitian ini adalah penelitian yang dilakukan oleh Yudhana dkk (2018). Dilakukan penelitian mendeteksi plagiarisme menggunakan algoritma Winnowing dan didapatkan kesimpulan bahwa metode yang dilakukan dapat mencapai persentase sampai 85% - 100% yang disimpulkan hasil ini cukup andal dalam pendeteksian plagiarisme. Pendeteksian menggunakan metode n-gram dan jaccard similarity pada algoritma Winnowing, cukup baik digunakan dalam untuk membandingkan kesamaan antara dua dokumen dan juga cukup baik digunakan untuk meminimalisir tindakan plagiarisme. Lalu juga penelitian yang dilakukan oleh Yudhi. L. J, dkk (2017) dilakukan pengujian pendeteksian kemiripan dokumen pada 11 dokumen dengan menggunakan algoritma Rafclif/Obershelp dan didapati hasil dari nilai yang didapat mendekati nilai yang diprediksi dan pada posisi yang berbeda dengan kesamaan yang sama nilai yang didapatkan berbeda, dapat disimpulkan bahwa banyaknya tugas yang terkumpul dan banyaknya karakter didalam dokumen tugas pada suatu tugas mempengaruhi waktu eksekusi

dari proses pengidentifikasian kesamaan dokumen. Dan posisi kalimat yang sama mempengaruhi nilai kesamaan dari pendeteksian dokumen dokumen yang diuji.

Berdasarkan kedua terdahulu yang telah dicantumkan di atas, penulis mengambil kesimpulan bahwa algoritma Winnowing dan Ratcliff/Obershelp cocok untuk dikombinasikan karena kedua algoritma ini saling melengkapi satu sama lain yang mana Winnowing mengidentifikasi kemiripan dokumen teks dengan fungsi hash yang diubah menjadi fingerprint, sedangkan algoritma Ratcliff/Obershelp mengidentifikasi kemiripan dokumen teks berdasarkan keseluruhan teks pada dokumen yang akan diuji.

Berdasarkan analisa dari masalah diatas, maka penulis mengambil judul penelitian " Identifikasi tingkat kemiripan dokumen teks menggunakan fungsi hash pada algoritma Winnowing dan Pattern Recognition Ratcliff/Obershelp". sehingga diharapkan penerapan algoritma Winnowing dapat menemukan tingkat kemiripan antar teks pada dokumen.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang masalah dan identifikasi masalah diatas, rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritma Winnowing dan Ratcliff/Obershelp dalam mencari tingkat kemiripan dokumen teks?
2. Bagaimana menentukan persentase tingkat kemiripan dokumen antar dua dokumen menggunakan algoritma Winnowing dan Ratcliff/Obershelp?

## **1.3 Batasan Masalah**

Untuk menghindari adanya cakupan materi yang meluas nantinya pada tugas akhir ini, maka penulis membatasi dalam tugas akhir ini hanya mencakup hal-hal sebagai berikut:

1. Pengujian dilakukan dengan menggunakan format file docx.
2. Dalam pengujian ini hanya mencakup teks yang terdapat pada dokumen, dan tidak mencakup hal lain seperti gambar, video dan suara.

3. Tidak memperlakukan adanya kesalahan dalam penulisan kata ataupun kalimat.
4. Penelitian dilakukan dengan 10 dokumen docx pendahuluan skripsi.

#### **1.4 Tujuan Penelitian**

Adapun tujuan yang didapat dari hasil penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma Winnowing dan Ratcliff/Obershelp untuk mencari tingkat kemiripan antar teks.
2. Mencari persentase tingkat kemiripan antar dua dokumen menggunakan algoritma Winnowing dan algoritma Ratcliff/Obershelp.

#### **1.5 Manfaat Penelitian**

Adapun manfaat yang didapatkan dari hasil penelitian ini adalah sebagai berikut:

1. Untuk mendeteksi tingkat kemiripan pada beberapa dokumen teks menggunakan waktu yang singkat dan mengidentifikasi apakah terdapat dokumen tersebut adalah hasil plagiarisme.
2. Untuk mengurangi dan menghindari adanya tindakan plagiarisme pada dokumen juga semakin memahami bagaimana prosen pendeteksian adanya plagiarime dengan adanya program ini.
3. Untuk lebih memahami bagaimana cara pendeteksian plagiarism pada dokumen teks.