

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi berkembang dengan sangat pesat sehingga memberikan kemudahan dalam hal bertukar informasi atau dalam menyusun suatu laporan. Dikutip melalui Wikipedia laporan adalah suatu bentuk penyampaian berita, keterangan, pemberitahuan atau pertanggungjawaban baik secara lisan atau tulisan. Laporan banyak dimanfaatkan oleh berbagai instansi antara lain sekolah, kampus, perkantoran, dan pemerintahan dengan kebutuhan isi penyampaian yang berbeda-beda. Salah satu contoh fungsi laporan dalam lingkup kampus dalam pengumpulan tugas mahasiswa yang diberikan oleh dosen pengampu. Dalam hal ini terkadang terdapat beberapa permasalahan dari lingkungan akademis dalam pengumpulan tugas yang dilakukan mahasiswa serta proses beberapa pengajar dalam mengoreksi tugas mahasiswa.

Menurut Azhar Firdaus, Ernawati, dan Arie Vatesia(2014) berdasarkan Jurnal Teknologi Informasi, terdapat beberapa penyalahgunaan yang terkadang terjadi di lingkungan akademis terutama oleh mahasiswa. Salah satu contoh tindakan penyalahgunaan yang terjadi adalah melakukan penjiplakan terhadap seluruh atau beberapa teks dari satu atau beberapa berkas dokumen teks sumber ke berkas dokumen teks lain. Praktik penyalahgunaan ini sering terjadi pada proses pembuatan tugas mata kuliah, tugas, praktikum, dan tugas akhir. Plagiat adalah pengambilan karangan (pendapat) orang lain dan disiarkan sebagai karangan (pendapat) sendiri dengan standart nilai threshold tertentu . Tindakan pengambilan karangan atau ide bukan suatu hal yang baru bagi mahasiswa. Adanya alat bantu seperti komputer dan laptop, memberikan kemudahan untuk melakukan tindakan penjiplakan. Hal tersebut berdampak pada perilaku yang terkesan malas dalam mengerjakan tugas mata kuliah, tugas praktikum dan tugas akhir. Hal tersebut juga berpengaruh pada tahap evaluasi hasil pembelajaran. Peserta mata kuliah yang tidak sedikit

memberikan kesulitan bagi para pengajar untuk memberikan hasil yang objektif. Hal tersebut menimbulkan kecurigaan terhadap tugas-tugas yang nantinya akan diakumulasikan ke tahap nilai akhir. Sehingga para pengajar akan direpotkan untuk menganalisis satu per satu tugas mahasiswa dikarenakan jumlah peserta mata kuliah tidak sedikit. Sedangkan beberapa pengajar masih menggunakan proses manual untuk mengoreksi tugas mahasiswa dengan membandingkan dua atau lebih berkas tugas dan melakukan penelusuran per paragraph, dan per kalimat untuk menentukan apakah tugas yang telah dikerjakan mahasiswa memiliki kemiripan dengan mahasiswa lainnya. Proses tersebut merupakan cara yang kurang efektif dan efisien. Oleh sebab itu diperlukan adanya aplikasi yang mampu mendeteksi kemiripan pada dokumen teks.

Aplikasi pendeteksi kemiripan pada dokumen teks, dibangun dengan menggunakan suatu ilmu yang mengacu pada pengolahan data teks. *Text Mining* adalah proses menganalisa teks untuk mengekstrak informasi yang berguna untuk tujuan tertentu. Tahap-tahap pada *text mining* secara umum terdiri dari *case folding*, *tokenizing*, *filtering*, *stemming*, *tagging*, *analyzing*. Terdapat beberapa metode dalam melakukan tahapan terakhir *text mining* yaitu *analyzing* atau biasa disebut analisa data untuk menentukan kemiripan dua objek yang sedang dibandingkan. *Cosine Similarity* adalah salah satu metode dalam menentukan nilai kemiripan antar dua objek. Salah satu contoh penerapan adalah penentuan kemiripan pada sidik jari manusia. *Cosine Similarity* dapat diterapkan dalam menentukan nilai kemiripan pada dua berkas dokumen teks. Parameter yang digunakan adalah jumlah kata-kata pada dua dokumen teks yang dibandingkan. *Cosine Similarity* menggunakan dua vektor yang mempresentasikan dua dokumen teks dimana nilai sudut kosinus dari kedua vektor tersebut adalah nilai kemiripan dari dua dokumen teks tersebut. Batasan nilai yang dihasilkan mulai dari 0 sampai dengan 1 atau biasa disebut *Binary Term*.

Penelitian ini diharapkan dapat menghasilkan aplikasi yang dapat mendeteksi kemiripan dokumen yang sedang dibandingkan dan memiliki

efektifitas dalam penganalisisan data dari penelitian yang sudah dilakukan terdahulu. Serta bermanfaat bagi pengguna dalam memberikan kemudahan mencari kemiripan antar dokumen teks dengan praktis dan cepat. Dengan melakukan beberapa penerapan metode yang berbeda dalam tahapan *text mining*. Pada sistem ini akan diaplikasikan *text mining* untuk tahap *preprocessing* dan *cosine similarity* sebagai *string matching*. Dengan mengekstrak dokumen yang berisi teks kemudian dilakukan suatu perhitungan dengan metode *cosine similarity* agar menghasilkan suatu keluaran *similarity* dokumen dari jumlah *string* kata dengan batasan *threshold* tertentu untuk mengetahui kemiripan dokumen tersebut. Selanjutnya akan dibahas bagaimana *text mining* dan *cosine similarity* bekerja dan sekaligus implementasi pada sebuah aplikasi dalam mendeteksi kemiripan

1.2 Rumusan Masalah

Bagaimana cara melakukan deteksi kemiripan isi dokumen?

1.3 Tujuan Penelitian

Adapun penelitian ini bertujuan untuk merancang dan membangun aplikasi deteksi kemiripan dokumen menggunakan *text mining* dengan metode *cosine similarity* dengan fungsi sebagai berikut :

1. Melakukan pengecekan plagiasi dengan standart *threshold* tertentu.
2. Memberikan informasi letak kata yang sama dari dua dokumen

1.4 Batasan Masalah

Dalam perancangan aplikasi ini memiliki beberapa batasan masalah sebagai berikut:

1. Berkas dokumen yang akan dibandingkan berupa *file* dengan ekstensi antara lain: *.pdf, *.doc.
2. Berkas dokumen seluruhnya atau sebagian besar menggunakan Bahasa Indonesia.
3. Berkas dokumen memiliki *format* penulisan ejaan yang dibenarkan (EYD).

1.5 Manfaat Penelitian

Adapun manfaat yang ingin dicapai dari penelitian ini adalah :

Bagi pengguna aplikasi dapat memberikan kemudahan dalam mencari kemiripan antar dokumen dengan praktis dan membantu pengguna dalam menganalisa kemiripan dokumen melalui nilai similarity dan dibuktikan dengan letak kemiripan yang sama antar dokumen.

1.6 Metodologi Penelitian

Metodologi dalam penelitian skripsi ini sebagai berikut:

1. Studi Literatur

Dalam tahap ini dilakukan pembelajaran sejumlah literatur dalam hal konsep dan teknologi yang akan digunakan. Melalui suatu literature meliputi refrensi paper, buku dan dokumentasi internet.

2. Tahap Pengumpulan Data

Pada Tahap ini dilakukan pengumpulan data untuk dokumen yang akan dibandingkan.

3. Tahap Preprocessing Data

Dalam Tahap ini data yang sudah diperoleh akan diproses untuk dipilah kembali sesuai kebutuhan sistem yang nantinya akan dijadikan sebagai dokumen teks sumber.

4. Analisis Sistem

Pada tahap ini dilakukan sebuah analisis sistem untuk menentukan konsep dan teknologi yang digunakan untuk menyelesaikan permasalahan terkait.

5. Implementasi

Dalam tahap ini dilakukan sebuah proses penerjemah dari perancangan sistem yang telah di analisis sebelumnya untuk dilakukan pengkodean menggunakan Bahasa pemrograman PHP dan Mysql.

6. Pengujian

Tahap selanjutnya merupakan uji coba terhadap program yang telah dibuat untuk mengetahui sejauh mana kinerja sistem dalam melakukan perhitungan kemiripan dokumen yang dibandingkan.

7. Tahap Penyusunan Laporan

Melakukan penyusunan laporan berdasarkan sistematika penulisan.

1.7 Sistematika Penulisan

Adapun sistematika penulisan pada tugas akhir ini sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, metodologi penelitian, sistematika penulisan dan jadwal kegiatan yang direncanakan.

BAB II LANDASAN TEORI

Bab ini membahas tentang teori serta konsep dari permasalahan terkait kemiripan dokumen menggunakan *Text Mining* dengan metode *Cosine Similarity*.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini menjelaskan tentang penganalisaan kebutuhan dan perancangan sistem. Meliputi dari analisa kebutuhan sistem dan desain perancangan sistem

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas tentang uji coba sistem yang telah dibuat dari hasil implementasi melalui pengkodean dengan Bahasa pemrograman terkait.

BAB V PENUTUP

Bab ini terdiri dari kesimpulan dan saran dari hasil penelitian yang telah dilakukan berguna untuk pengembangan sistem selanjutnya.

