

1. Pendahuluan

1.1 Latar belakang

Pengumpulan tugas dalam jumlah yang banyak seringkali dihadapi oleh pemateri yang memberikan perkuliahan umum. Dengan kondisi ini, hampir tidak mungkin manusia dapat mendeteksi tindakan plagiat dari pengumpul tugas. Akan sangat membantu jika ada piranti yang dapat membantu mendeteksi tindak plagiat.

Sistem pendeteksi plagiat yang sudah ada seperti Sherlock [1], Jplag [2], dan MOSS [3] dapat menangani kasus diatas. Ketiga sistem tersebut melakukan perbandingan banyak ke banyak untuk semua dokumen, lalu memberikan daftar nilai similaritas yang selanjutnya digunakan untuk menentukan dokumen-dokumen mana saja yang berplagiat dengan dokumen lain [4].

Namun terjadi kendala ketika jumlah dokumen yang masuk ke sistem sudah sangat banyak. Contoh pada sistem Sherlock dan Jplag, keduanya memiliki kompleksitas waktu $O(f(n)) = N^2$ dimana N adalah jumlah dokumen, dan $f(n)$ waktu yang dibutuhkan untuk melakukan perbandingan dengan panjang dokumen n [4].

Solusi yang ditawarkan pada tugas akhir ini adalah pemanfaatan indeks untuk memetakan dokumen dan pengeliminasian dokumen berdasarkan pemetaan tersebut. Selain itu, proses pencocokkan untuk deteksi plagiat yang digunakan adalah teknik graph, sehingga pencocokkan dapat dilakukan secara paralel. Dengan kombinasi seperti ini, diharapkan waktu pemrosesan dapat diminimalisasi yang menghasilkan sistem dengan waktu pemrosesan cepat.

1.2 Perumusan masalah

Pada penelitian tugas akhir ini, rumusan masalah yang akan diteliti adalah bagaimana peningkatan performansi dalam hal kecepatan yang dihasilkan dalam implementasi *scapegoat tree indexing* pada sistem pendeteksi plagiat?

1.3 Tujuan

Tujuan dari Tugas Akhir ini adalah untuk mengukur efek penggunaan *scapegoat tree indexing* dalam mempercepat sistem pendeteksi plagiat.

1.4 Batasan Masalah

Batasan masalah yang dibahas dalam penelitan tugas akhir ini sebagai berikut :

1. Jumlah dokumen yang terlibat dalam Tugas Akhir ini berjumlah 10.000 dokumen
2. Ekstensi dokumen yang digunakan :

- a. Txt, yaitu ekstensi dokumen yang hanya berisi teks
- b. Doc, yaitu ekstensi dokumen yang dihasilkan dari *Microsoft Word*
3. Dokumen hanya berisi tulisan saja, selain itu tidak dihiraukan.
4. Dokumen berbahasa Indonesia

1.5 Metodologi penyelesaian masalah

Metodologi yang digunakan penulis untuk mencari penyelesaian masalah adalah :

1. Studi Literatur

Pada tahap ini, dilakukan pembelajaran mengenai topik-topik yang berhubungan dengan penelitian. Topik yang dipelajari adalah mengenai indexing, *Scapegoat Tree*, dan *Document Index Graph*. Sumber bacaan berupa paper, *e-book*, jurnal, atau artikel.

2. Pengumpulan data uji

Pada tahap ini, dilakukan pengumpulan data uji untuk membangun sistem, dan analisa dari performansi sistem. Data uji didapat dari berbagai sumber, diantaranya tugas mata kuliah, tugas akhir, dan hasil *crawling*. Total data yang berhasil didapatkan berjumlah 10.000 dokumen.

Untuk data uji berupa pengelompokkan dokumen plagiat, dilakukan survei kepada 10 responden, lalu dilihat suara terbanyak apakah pasangan dokumen yang disajikan plagiat atau tidak.

3. Perancangan sistem

Pada tahap ini, dilakukan analisa sistem apa saja yang perlu dibangun. Analisa diawali dengan melihat data uji yang didapat memiliki variasi bentuk dan panjang. Oleh karena itu, diperlukan adanya preprocessing terhadap data uji sehingga sistem akan memberikan hasil yang optimal.

4. Pengujian

Tahap pengujian dimulai dengan memasukkan dokumen dengan jumlah yang bervariasi. Dari setiap pengujian, dicatat waktu yang dibutuhkan untuk memproses setiap dokumen, sehingga dapat dianalisis kestabilan waktu pemrosesannya.

5. Analisa hasil

Tahap analisa dilakukan setelah tahap pengujian. Analisa dilakukan terhadap data-data yang didapat dari hasil pengujian berupa log file yang berisi pencatatan waktu, dan data-data lainnya. Pada akhir tahap, didapatkan kesimpulan.

6. Pelaporan

Semua tahap yang sudah dilalui dituangkan dalam bentuk laporan penelitian. Laporan ditujukan untuk mendokumentasikan penelitian, dan diharapkan dapat dikembangkan menjadi lebih baik lagi.

1.6 Sistematika Penulisan

Berikut sistematika penulisan pada laporan penelitian tugas akhir ini :

1. Bab 1 menjelaskan mengenai latar belakang, perumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah, dan sistematika penulisan.
2. Bab 2 berisi tinjauan pustaka yang menguraikan mengenai indexing, *winnowing*, *scapegoat tree*, *f-measure* dan *Document Index Graph (DIG)*
3. Bab 3 berisikan tentang perancangan sistem, mulai dari gambaran umum sistem, dan perancangan sistem yang terdiri preprocessing, indexing dengan menggunakan *scapegoat tree*, pencocokkan dokumen dengan DIG, dan skenario pengujian.
4. Bab 4 menjelaskan mengenai pengujian dan analisis berdasarkan skenario pada bab sebelumnya. Bab 4 terdiri dari tujuan, strategi pengujian, dan analisis hasil pengujian.
5. Bab 5 menjelaskan mengenai kesimpulan dan saran yang didapatkan setelah melakukan pengujian dan analisis