

1. Pendahuluan

1.1 Latar belakang

Plagiarisme merupakan sebuah tindakan menggunakan sebagian atau seluruh hasil karya orang lain tanpa mencantumkan identitas pengarangnya baik secara sengaja maupun tidak. Pada era teknologi sekarang ini, mudahnya pencarian informasi melalui internet dan penggunaan proses *copy-paste-modify* menjadikan tindakan ini banyak dilakukan sebagai solusi pengerjaan tugas-tugas akademik. Untuk itulah diperlukan sebuah sistem pendeteksi plagiarisme. Dengan sistem ini diharapkan sebuah tindakan plagiarisme dapat dideteksi dengan cepat dan tepat sehingga plagiat mendapat sanksi.

Cara naif untuk menentukan sebuah dokumen plagiat atau tidak adalah dengan melakukan perbandingan antara satu dokumen *query* dengan semua dokumen lain yang ada pada *database*. Cara ini telah digunakan pada modern detektor plagiarisme seperti Sherlock [13,23], Jplag [19,23] dan MOSS [20,23]. Ketiga sistem tersebut memiliki ide yang sama, yaitu perbandingan terhadap satu *query* ke semua dokumen yang ada. Proses ini akan memakan waktu yang lama.

Dengan adanya permasalahan tersebut, diajukan usulan agar sistem dapat melakukan proses pengeliminasian dokumen pada *database* terlebih dahulu agar tidak seluruh dokumen dicocokkan dengan dokumen *query*. Dokumen-dokumen yang telah terfilter karena terindikasi memiliki kesamaan selanjutnya akan dilakukan pencocokan. Hal ini akan mempercepat waktu komputasi karena proses pencocokan akan menjadi lebih sedikit. Proses pengeliminasian diusulkan dengan menggunakan *indexing*. Untuk menangani masalah pencarian indeks sekuensial, digunakan salah satu struktur data *balance tree*, yaitu *2-3 tree* guna mempercepat proses pencocokan antara suatu kata yang ada pada dokumen uji dan dokumen pembanding.

1.2 Perumusan masalah

Masalah yang difokuskan pada penelitian ini adalah bagaimana cara mempercepat waktu pemrosesan sistem indikasi deteksi plagiarisme tanpa mengurangi keakuratannya dengan proses *indexing* struktur data *2-3 tree*.

1.3 Tujuan

Tujuan dari penelitian ini adalah:

1. Merancang sistem yang lebih cepat dengan mengimplementasikan proses *indexing* struktur data *2-3 tree*.
2. Menghitung performansi (kecepatan dan akurasi) sistem saat menggunakan proses *indexing 2-3 tree*.

3. Membandingkan performansi sistem dengan menggunakan proses pencocokan ke semua dokumen pada *database* dan proses *indexing 2-3 tree*.

1.4 Batasan Masalah

Dalam pengerjaannya penelitian ini memiliki beberapa batasan masalah sebagai berikut:

1. Data dokumen yang berisi teks saja. Tidak dapat menangani dokumen gambar, suara maupun multimedia lainnya. Extensi file teks yang dapat ditangani hanya: .txt (dokumen yang hanya berisi teks), .doc (ekstensi yang dihasilkan dari microsoft word), .docx (ekstensi yang dihasilkan dari microsoft word).
2. Bahasa yang digunakan adalah bahasa Indonesia dengan jumlah data yang akan digunakan adalah dokumen sebanyak 10000 buah.
3. Proses *stemming* menggunakan library Lucene.
4. Sistem tidak menangani kesamaan makna kata dan sinonim, perubahan kalimat aktif ke pasif maupun sebaliknya, pendeteksian quote (“...”) sebagai bentuk pencegahan plagiarisme, kesalahan eja pada dokumen yang plagiat.
5. Menggunakan dokumen *query* yang redaksinya disusun secara mandiri.

1.5 Metodologi penyelesaian masalah

Pembuatan sistem pendeteksian plagiarisme ini menggunakan metodologi sebagai berikut:

1. Identifikasi masalah
Masalah yang ada pada dunia nyata diidentifikasi, lebih dispesifikkan dan dicari tahu apakah dapat dapat diselesaikan atau tidak. Masalah yang melandasi pengerjaan penelitian ini adalah karena diperlukannya sebuah sistem pendeteksian plagiarisme yang cepat dan akurat.
2. Studi Literatur
Pada tahap ini dipelajari sejumlah literatur yang berkaitan mengenai masalah yang dihadapi, beserta teknik/proses yang dapat digunakan untuk dapat menyelesaikan masalah yang telah ditentukan sebelumnya. Literatur yang digunakan meliputi paper, jurnal, buku dan artikel-artikel yang terkait dengan permasalahan. Topik yang dipelajari adalah mengenai deteksi plagiat, *stemming*, *winnowing*, penghapusan *stopword*, struktur data *2-3 tree*, *indexing* dan LCS.
3. Pembangunan Sistem
 - 3.1 Analisis Kebutuhan Sistem
Pada kasus pendeteksian plagiarisme ini pada proses analisis kebutuhan meliputi teknologi apa yang akan dipakai, bahasa

pemrograman apa yang akan digunakan, penentuan dataset yang akan digunakan, penentuan cara mendapatkan dataset dan lain-lain.

3.2 Perancangan Sistem

Pada tahap ini dirancang fungsionalitas yang dapat dilakukan oleh sistem, modul yang saling berinteraksi, diagram alur proses dan lain-lain.

3.3 Pembangunan Sistem

Pada tahap ini sistem mulai dibangun sesuai dengan hasil rancangan yang telah dibuat sebelumnya.

3.4 Pengumpulan Dataset

Pada tahap ini pengumpulan dataset untuk dokumen pembanding dilakukan. Dokumen pembanding ditargetkan sebanyak 10000 buah. Dataset didapat dari hasil tugas-tugas mahasiswa dan hasil *crawling* dengan *query* tertentu.

3.5 Pengujian Sistem

Pengujian sistem dilakukan untuk menguji coba kebenaran sistem. Beberapa sampel data uji dimasukkan dan diproses oleh sistem. Beberapa skenario dibentuk sedemikian rupa sebagai bentuk pengukuran terhadap suatu aspek.

3.6 Analisis Hasil

Pada tahap ini hasil pengujian sistem akan di analisis terhadap hasil dari sistem terhadap berbagai macam skenario pengujian. Pada penelitian ini hasil sistem juga akan dibandingkan dengan hasil survey.

3.7 Pembuatan Laporan

Pada tahap ini dilakukan penyusunan laporan yang berisi dasar teori yang digunakan, hal-hal yang dilakukan selama pembuatan sistem (dokumentasi sistem), analisis hasil dan kesimpulan.

1.6 Sistematika Penulisan

Bab I: Pendahuluan

Bab ini berisi tentang informasi umum dari laporan penelitian ini. Pendahuluan terbagi kedalam beberapa bagian: Latar Belakang, Perumusan Masalah, Tujuan, Batasan Masalah, Metodologi Penyelesaian Masalah dan Sistematika Penulisan.

Bab II: Landasan Teori

Bab ini berisi mengenai teori-teori yang berkaitan dengan penelitian yang dikerjakan. Definisi, teori dan konsep diambil dari berbagai macam jurnal dan buku. Landasan Teori terbagi kedalam beberapa *sub*-bab: Plagiarisme,

Data *Pre-processing*, *Winnowing*, *Inverted index*, struktur data *2-3 tree* dan *LCS*.

Bab III: Analisis Perancangan dan Implementasi

Bab ini berisi mengenai rancangan akan sistem yang akan dibuat, mulai dari use-case diagram, diagram blok dan lain-lain, beserta implementasinya. Pada bab ini terbagi kedalam beberapa *sub*-bab: Deskripsi Dan Analisis Sistem, Perancangan Sistem dan Implementasi.

Bab IV: Pengujian dan Analisis

Bab ini bertujuan untuk mnguji sistem berdasarkan skenario yang telah dibuat sebelumnya. Hasil dari pengujian akan dianalisis.

Bab V: Kesimpulan dan Saran

Bab ini berisi kesimpulan dari hasil pengujian dan analisis beserta saran untuk penelitian selanjutnya.