

1. Pendahuluan

1.1 Latar Belakang

Perkembangan teknologi informasi di era informasi sangatlah pesat. Pertukaran informasi menjadi hal yang terpenting dalam segala lini kehidupan. Namun, ada beberapa pihak yang memanfaatkan kemudahan pertukaran informasi untuk tindak kecurangan, seperti plagiarisme. Di dalam dunia akademik, plagiarisme merupakan tindak kecurangan yang paling berat, sehingga dirasakan perlu adanya sebuah sistem yang dapat mendeteksi plagiarisme.

Plagiarisme atau yang sering disebut dengan plagiat merupakan perbuatan secara sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya ilmiah orang lain, tanpa menyatakan sumber secara tepat dan memadai [11]. Tindakan plagiarisme ini sangatlah buruk, tindakan ini tidak hanya memberikan dampak negatif ke orang yang diplak namun terhadap orang yang melakukan plagiat juga. Plagiarisme ini dapat mematahkan kreatifitas seseorang dan cenderung membuat seseorang menjadi malas untuk berfikir.

Proses pendeteksian plagiarisme dilakukan dengan cara membandingkan kemiripan dokumen asli dengan dokumen yang ingin diuji. Kemiripan dari dokumen bisa didapat melalui pengecekan dokumen pada level konteks leksik maupun pada konteks parafrase kalimat. Sebuah sistem pendeteksi plagiarisme level konteks leksik dapat dikembangkan dengan tiga pendekatan pendeteksian plagiarisme, yaitu pendekatan perbandingan teks lengkap, pendekatan dokumen *fingerprint*, dan pendekatan kesamaan kata kunci[3]. Berdasarkan permasalahan di atas, sudah mulai banyak dikembangkan aplikasi untuk mendeteksi plagiarisme seperti menggunakan pendekatan *fingerprint* dan *Longest Common Subsequence* (LCS).

Pada umumnya pendekatan *fingerprint* pada suatu dokumen dapat dilakukan dengan memanfaatkan suatu fungsi *hash* pada *k-grams* dari suatu dokumen[14]. *K-grams* adalah *substring* dengan panjang *k* dari penggabungan simbol/token pada suatu dokumen. Dalam proses pendeteksian plagiarisme pada aplikasi yang menggunakan pendekatan *fingerprint* terdapat kelemahan yang dapat mengurangi ketepatan dalam mendeteksi plagiarisme. Salah satu kelemahan yang terdapat pada metode *fingerprint* adalah kemungkinan mendeteksi bagian dokumen yang bukan merupakan topik yang dibahas oleh dokumen tersebut.

Selain metode *fingerprint*, metode yang umum digunakan untuk membangun aplikasi pendeteksian plagiarisme adalah metode LCS[14]. Metode ini merupakan sebuah metode pencocokan *string* yang mencari sub-rangkaian umum (*common subsequence*) terpanjang dari beberapa (pada umumnya dua) rangkaian *string*. Dalam implementasinya LCS ini juga memiliki kelemahan. Kelemahan yang terdapat pada metode ini adalah dari segi kecepatan pendeteksian atau pencocokannya. Dalam melakukan pencocokan *string* atau isi dari dokumen, metode ini harus melakukan pengecekan ke seluruh isi dari dokumen.

1.2 Perumusan masalah

Dari penjelasan diatas maka masalah yang dicari penyelesaiannya dalam Tugas akhir ini adalah :

1. Bagaimana membuat sebuah sistem yang dapat mempercepat proses pendeteksian plagiarisme terhadap dokumen teks menggunakan metode *Longest Common Subsequence*?
2. Bagaimana memilih kata kunci sebagai *fingerprint* dalam proses pendeteksian plagiarisme?

Dalam tugas akhir ini yang dibahas adalah Implementasi Deteksi Plagiarisme dengan Metode *Co-Occurrence Statistical Information* dan LCS dengan batasan masalahnya sebagai berikut :

1. Hanya menguji data teks, tidak data berupa gambar maupun suara.
2. Hanya menangani plagiarisme dengan jenis plagiarisme *copy paste*, plagiarisme parafrase, dan plagiarisme aktif-pasif.
3. Sistem tidak memperhatikan sinonim / persamaan kata.
4. Sistem tidak memperhatikan makna kata.
5. Sistem tidak memperhatikan kesalahan ejaan / penulisan pada dokumen.
6. Data yang diuji bertipe .doc atau .txt.
7. Data yang diuji menggunakan bahasa Indonesia.

1.3 Tujuan

Berdasarkan pada masalah yang telah diidentifikasi diatas, maka tujuan yang ingin dicapai dalam tugas akhir ini adalah :

1. Merancang aplikasi yang mampu mempercepat proses pendeteksian plagiarisme dengan menggunakan metode *Longest Common Subsequence*.
2. Melakukan verifikasi terhadap kata kunci yang digunakan sebagai *fingerprint*.
3. Menghitung performansi sistem dalam hal lama waktu komputasi dan akurasi yang dihasilkan.

1.4 Metodologi penyelesaian masalah

Metode yang digunakan untuk menyelesaikan Tugas Akhir ini, yaitu :

1. Studi Literatur dan Pengumpulan Data
Pada tahap ini dilakukan pencarian sumber-sumber bacaan yang berhubungan dengan penelitian yang dilakukan. Topik – topik yang dipelajari adalah mengenai deteksi plagiat, algoritma *Co-Occurrence Statistical Information*, dan LCS. Sumber bacaan berupa e-book, jurnal, paper atau artikel yang diperoleh dari internet dan perpustakaan IT Telkom. Pengumpulan data didapatkan dari membangun dataset.
2. Analisis Kebutuhan Sistem dan Perancangan Perangkat Lunak
Pada tahap ini dilakukan analisis, perancangan dan penerapan dari metode *co-occurrence Statistical Information* dan LCS dalam pendeteksian plagiarisme, inputan sistem berupa dokumen, kemudian dokumen diringkas berdasarkan kata kunci yang diekstrak, selanjutnya hasil ringkasan tersebut digunakan sebagai identitas dokumen yang nantinya

dibandingkan dengan dokumen lain menggunakan metode LCS. Keluaran akhir dari sistem ini berupa nilai kemiripan dari dokumen yang diuji.

3. Rancangan dan Implementasi Perangkat Lunak

Untuk menggunakan *Co-Occurrence Statistical Information* dan LCS ini, maka perlu dilakukan tahapan *preprocessing*. Pada *preprocessing* ini dilakukan persiapan atau penyeragaman bentuk dari dokumen inputan. Setelah *preprocessing* pada dokumen inputan, proses selanjutnya peringkasan dokumen dengan *Co-Occurrence Statistical Information* dan terakhir perbandingan dokumen dengan LCS. Secara teknis, implementasi dilakukan dengan spesifikasi *software* dan *hardware* sebagai berikut:

- Java SDK 1.6
- Eclipse sebagai java editor
- Processor AMD Phenom(tm) II X4 955
- RAM 3.25GB

4. Pengujian dan Analisis hasil

Pada tahap ini dilakukan pengujian terhadap sistem yang telah dibangun dan sekaligus melakukan analisis terhadap hasil pengujian sistem tersebut. Analisis terhadap hasil pengujian sistem ini dilakukan dengan menggunakan analisis parameter *threshold* dan jumlah kata kunci, untuk melihat bagaimana performansi dari segi akurasi dan lama waktu komputasi dari sistem yang sudah dibuat.

5. Penyusunan Laporan Tugas Akhir

Membuat kesimpulan dari hasil analisis dan pegujian yang dilakukan. Kemudian dilakukan dokumentasi semua tahapan proses diatas berupa laporan yang berisi tentang dasar teori dan hasil Tugas Akhir ini ke dalam sebuah buku tugas akhir