BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Plagiarisme adalah perbuatan yang mengambil tulisan, karangan, atau pendapat dari orang lain tanpa adanya rujukan dan menjadikannya seolah-olah kutipan atau tulisan tersebut sebagai tulisan milik sendiri. Kemajuan teknologi saat ini ada yang berdampak negatif, yaitu dengan mudah mendapat dokumen secara online. Kemudahan itu disalahgunakan dengan seringnya mengutip tulisan dari orang lain tanpa menuliskan rujukan terhadap tulisan yang diambil.

Penelitian yang berkaitan dengan plagiarisme sudah pernah dilakukan sebelumnya. Penelitian yang dilakukan oleh Taufiq M Isa, dkk (Isa dkk, 2013) dengan judul ”Mengukur Tingkat Kesamaan Paragraf Menggunakan Vector Space Model untuk Mendeteksi Plagiarisme” menunjukkan bahwa algoritma yang diusulkan dapat mendeteksi dengan baik kesamaan antar dua dokumen. Proses yang dilakukan penelitian tersebut memasukkan data berupa dokumen PDF lalu di konversi kedalam bentuk XML dan nantinya dokumen dirangking dengan cara mengambil paragraf-paragraf yang kemudian dipecah-pecah, setelah didapatkan nilai IDF, TF-IDF dan VSM nya, didapatkan hasil akhir yaitu nilai similaritas antar paragraf. Dalam penelitian tersebut pada tahap *preprocessing* tidak digunakan *stemming* dan hasil pengujian tersebut peneliti menggunakan beberapa jenis query, yaitu query satu kata dimana hasilnya menunjukkan pasangan paragraf dalam kelompok similaritas tinggi lebih banyak dibandingkan dengan pasangan paragraf similaritas rendah dan sedang, query dua kata memperlihatkan hasil query juga mirip dengan hasil query satu kata, dan untuk query tiga kata memperlihatkan bahwa pasangan dengan similaritas tinggi dapat dideteksi dengan baik, dan untuk persentase nilai similaritas rendah sebesar 50-65,99%, sedang 66-80,99%, dan tinggi 81-100%.

Pada penelitian ini dilakukan tahap *stemming* dalam mendeteksi plagiarisme, dan beberapa penelitian yang terkait dengan *stemming* pernah dilakukan sebelumnya. Penelitian oleh Winda Ayu Iriyanto (Iriyanto, 2013) dengan judul

"Penentuan Tingkat Plagiarisme Dokumen Penelitian Menggunakan Centroid Linkage Hierarchical Method (CLHM)" mampu membandingkan kemiripan dokumen terhadap sekelompok dokumen dan penelitian tersebut melalui tahapan *stemming* dimana pemotongan kata secara secara *random* berpengaruh pada nilai *similarity*, serta hasil akurasi sistem menghasilkan rata-rata f-measure sebesar 0,984%. Penggunaan *stemming* bertujuan untuk memperkecil tindakan penjiplakan tanpa mengabaikan perubahan posisi kalimat, pemotongan kata ataupun perubahan kata aktif menjadi kata pasif atau sebaliknya. Penelitian yang dilakukan oleh Milani Winangga (Winangga, 2012) dengan judul "Deteksi Plagiarisme Pada Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Winnowing Dengan Stemming" penggunaan *stemming* menghasilkan persentase kemiripan yang lebih baik pada kasus pengubahan jenis kata kerja berimbuhan dari aktif ke pasif ataupun sebaliknya tanpa mengubah susunan objek, predikat, ataupun objeknya, dari dua penelitian tersebut penggunaan *stemming* berpengaruh pada persentase kemiripan dan perubahan kata dari aktif ke pasif. Dalam penelitian yang dilakukan tersebut masih diperlukan pengembangan dari batasan yang telah ada sebelumnya dalam mendeteksi kemiripan atau plagiarisme suatu dokumen.

Berdasarkan permasalahan tersebut maka akan dirancang dan dibangun suatu aplikasi untuk pendeteksi plagiarisme, dan penggunaan metode *vector space model* diharapkan bisa membantu aplikasi ini dalam mendeteksi plagiarisme dengan baik dan memberikan persentase tingkat kemiripan dokumen lebih baik. Oleh karena itu, penulis tertarik mengambil pokok bahasan skripsi dengan judul "Pendeteksian Plagiarisme Abstrak Paper Menggunakan Metode Vector Space Model".

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka dapat ditarik identifikasi dan masalah yang timbul, yaitu : mengetahui tingkat kemiripan atau plagiarisme suatu dokumen dengan penggunaan *stemming* pada proses *preprocessing*.

1.3 Maksud dan Tujuan

Berdasarkan permasalahan yang ada, maksud dari penelitian tugas akhir ini adalah untuk membangun aplikasi pendeteksi plagiarisme menggunakan metode Vector Space Model.

Sedangkan tujuan dari penelitian tugas akhir ini adalah untuk mengetahui tingkat akurasi pendeteksian kemiripan atau plagiarisme suatu dokumen dengan penggunaan *stemming* pada proses *preprocessing*.

1.4 Batasan Masalah

Adapun batasan masalah yang diberikan pada penelitian ini sebagai berikut :

1. Data Input :
   a. Bentuk teks paper masukan dalam Bahasa Indonesia.
   b. Format dokumen : *.pdf dengan 30 data.
   c. Bila terdapat rumus, gambar dalam teks maka tidak akan dimasukkan, hanya huruf saja yang diambil.
   d. Aplikasi ini tidak memperhatikan kesalahan penulisan (typo) pada dokumen input.
2. Data Output yang ditampilkan berupa status persentase plagiarisme.

1.5 Metodologi Penelitian

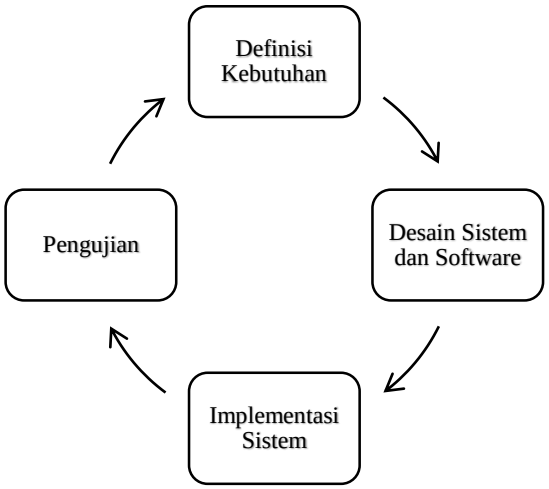
Untuk mempermudah dalam pengumpulan dan pencarian informasi dalam pengerjaan penelitian ini, ada beberapa metode penelitian yang dilakukan diantaranya:

1.5.1 Metode Pengumpulan Data

Metode yang dilakukan dalam pengumpulan data adalah Studi Literatur. Tahapan ini melakukan studi kepustakaan dari hasil penelitian yang telah dilakukan sebelumnya oleh orang lain, artikel–artikel yang terkait dengan pembuatan pendeteksi plagiarisme, membaca paper dan jurnal mengenai plagiarisme, serta mempelajari teknik dan algoritma yang tepat untuk dapat meningkatkan kemampuan aplikasi pendeteksi plagiarisme yang sudah dilakukan oleh peneliti sebelumnya.

1.5.2 Metode Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan adalah dengan menggunakan metode prototype yang dapat dilihat pada Gambar 1.1.



Gambar 1.1 Metode Pembangunan Perangkat Lunak (Roger, 2001)

1. Definisi Kebutuhan

   Tahap ini memperoleh informasi melalui paper, jurnal, dan buku. Informasi tersebut dianalisis untuk mendapatkan dokumentasi kebutuhan pengguna untuk digunakan pada tahap selanjutnya.
2. Desain Sistem dan Software

   Tahap ini dilakukan sebelum melakukan coding. Dalam tahap ini membuat gambaran yang dikerjakan dan bagaimana tampilan dari software yang dibangun.
3. Implementasi Sistem

   Tahap ini dilakukan pemrograman (*coding*). Pembuatan software dipecah menjadi modul-modul kecil yang nantinya akan digabungkan dalam sistem.
4. Pengujian

   Tahap ini dilakukan pengujian terhadap sistem melakukan pemeriksaaan terhadap modul yang dibuat, apakah sudah memenuhi fungsi yang diinginkan atau belum.

1.6 Sistematika Penulisan

Sistematika dalam penulisan penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan laporan ini adalah sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, sistematika penulisan dari pembangunan aplikasi deteksi plagiarisme.

BAB II LANDASAN TEORI

Bab ini berisikan teori-teori mengenai plagiarisme, pembobotan dengan algoritma TF/IDF, metode Vector Space Model(VSM), Software, Model Perangkat Lunak.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini berisikan tentang menganalisis masalah dari model penelitian untuk memperlihatkan keterkaitan antar variabel yang diteliti serta model matematis untuk analisisnya. Serta dalam bab ini akan dibahas tentang analisis sistem, analisis masalah, gambaran umum mengenai sistem yang akan dibangun, pemodelan analisis terstruktur dan desain antarmuka.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini berisikan tahapan yang dilakukan dalam penelitian secara garis besar sejak dari tahap persiapan sampai penarikan kesimpulan, metode dan kaidah yang diterapkan dalam penelitian. perancangan aplikasi untuk mengetahui keberhasilan algoritma dalam menyelesaikan aplikasi pendeteksi plagiarisme. Pengujian terhadap aplikasi deteksi plagiarisme dilakukan dengan menggunakan metode Black Box.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian mengenai deteksi plgiarisme yang dilakukan oleh penulis dan saran untuk pengembangan selanjutnya yang terkait dengan deteksi plagiarisme.