BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Penjiplakan atau *plagiarisme* berarti mencontoh atau meniru tulisan dan karya orang lain yang kemudian diakui sebagai karanganya sendiri dengan ataupun tanpa seizin penulisnya. Penjiplakan dokumen digital bukanlah hal yang susah, cukup dengan menggunakan teknik *copy-paste-modify* pada sebagian isi dokumen dan bahkan keseluruhan isi dokumen sudah bisa dikatakan bahwa dokumen tersebut merupakan hasil duplikasi dari dokumen lain [2]. *Plagiarisme* juga dapat dilakukan dengan dengan cara menambah atau menghilangkan kata [1]. Terkadang *plagiarisme* juga dilakukan dengan mengganti kata-kata yang mengandung kata - kata sinonim, dengan maksud agar terlihat berbeda dari dokumen aslinya. Dimana hanya merubah kata – kata yang mengandung sinonim yang terdapat di dokumen asli dengan merubah nya sesuai dengan kata yang ada di Kamus bahasa indonesia , tanpa memperhatikan struktur kalimat dari dokumen aslinya[3]. Pendeteksian plagiarism dapat dilakukan dengan cara pencocokan string (*string matching*). Metode ini dapat digunakan untuk menghitung kemiripan teks antara satu dokumen dengan dokumen lainnya.

Beberapa penelitian mengenai deteksi *plagiarisme* telah dilakukan sebelumnya diantaranya oleh Sandi Dewanto [2] dimana penelitian ini menggunakan algoritma *Rabin-Karp* varian Chillar-Kochar yang mampu mendeteksi kesamaan string dengan maksimal karena menghindari terjadinya spurious hit sehingga tidak perlu pencocokan karakter setelah string ditemukan, penggunaan fitur *synonim recognition* pada penelitin ini dengan cara menyocokan kata pada dokumen dengan kamus di database kemudian mengganti kata yang didokumen dengan kamus yang di database. Penelitian berikut nya dilakukan oleh Devi Dwi Purwanto [3] dimana penelitian ini proses pengenalan sinonim yang digunakan menggunakan algoritma Lesk *Word Sense* Disambiguation serta pemanfaatan Wordnet Bahasa Indonesia , dan untuk pendeteksian kesamaan kata atau *carbon copy* Menggukana Algoriam Winowing sedangkan menurut penelitian Hari Bagus Firdaus [4] Algoritma Rabin-

Karp menghasilkan efisiensi waktu yang baik dalam mendeteksi string yang memiliki lebih dari satu pola. Hal ini membuat algoritma Rabin-Karp dimanfaatkan dalam melakukan pendeteksian terhadap tindak plagiat dokumen.

Berdasarkan uraian di atas, pada penelitian ini akan mengombinasikan Algoritma *Lesk* dalam Pengenalan sinonim dengan memanfaatkan wordnet indonesia dan Rabin Karp untuk mendeteksi tingkat similiarity plagriarsme dokumen.

1.1 Identifikasi Masalah

Berdasarkan penjelasan dari latar belakang ditemukan permasalahan yang dirumuskan ke dalam suatu rumusan masalah adalah bagaimana mengkombinasikan Algoritma *Lesk* dalam Pengenalan sinonim dengan memanfaatkan wordnet indonesia dan Rabin Karp untuk mendeteksi tingkat similiarity plagriarsme dokumen.

1.2 Maksud Dan Tujuan

Maksud dari penelitian yang akan dilakukan adalah menerapkan algoritma *Lesk* Dan Rabin Karp untuk mendeteksi plagiarisme dua dokumen.

Sedangkan tujuan yang ingin dicapai adalah mengetahui besarnya tingkat similiarity dari kombinasi Algoritma Lesk dalam Pengenalan sinonim dengan memanfaatkan wordnet indonesia dan *Rabin Karp* yang akan diterapkan untuk mendeteksi plagiarism dua dokumen.

1.3 Batasan Masalah

Agar penelitian yang dilakukan lebih terarah dan mencapai sasaran yang ditentukan, maka perlukan sebuah pembatasan masalah atau ruang lingkup kajian, yaitu sebagai berikut:

1. Data dokumen yang bisa di upload oleh user hanya berextensi doc dan .docx
2. Kasus uji yang digunakan adalah latar belakang skripsi informatika.
3. Menggunakan medota k-gram kata , dengan nilai *k-gram* secara dinamis yaitu bisa di inputkan oleh user.
4. Kalimat yang mengandung reference pada akhir kalimat tidak akan di proses lebih lanjut.

5. Steming pada penelitian ini menggunakan algoritma porter stemmer.
6. Menggunakan Algoritma *Rabin Karp* varian Chillar-Kochar untuk mendeteksi plagiarism dokumen.
7. Mengunakan algoritma *lesk* untuk pengenalan dan pemilihan persamaan kata dengan teknik WSD (*Word Disambiguition Sense*) menggunakan kamus sinonim dari database.
8. Makna kata tiap kalimat diambil dari website kbbi.web.id
9. Sinonim kata tiap kalimat diambil dari website sinonimkata.com
10. Pendekatan pembangunan perangkat lunak yang digunakan pada penelitian ini adalah pendekatan terstruktur.

1.4 Metodologi Penelitian

Metodologi penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif. Metode yang digunakan dalam penulisan laporan penelitian ini menggunakan dua metode, yaitu metode pengumpulan data dan metode pembangunan perangkat lunak.

1.4.1 Metoda Pengumpulan Data

Metode yang digunakan dalam pengumpulan data pada penelitian ini adalah sebagai berikut:

a. Studi Literatur

   Pengumpulan data dengan cara mengumpulkan referensi seperti jurnal, *paper*, buku referensi dan bacaan-bacaan yang ada kaitannya dengan judul penelitian.

b. Pengumpulan Dokumen

   Pengumpulan dokumen disini yaitu pengumpulan dokumen – dokumen asli yang akan dijadikan objek untuk dideteksi dengan dokumen yang merupakan hasil *modify*.

1.5 Metode Pembangunan Perangkat Lunak

Metode yang dilakukan dalam proses pembangunan aplikasi ini adalah model *waterfall*. Terdapat beberapa tahapan dalam pembangunan perangkat lunak pada model *waterfall* yaitu sebagai berikut:

1. Analisis dan definisi persyaratan.

   Analisis dan definisi persyaratan meruapakan tahap pengumpulan kebutuhan secara lengkap kemudian dianalisis dan didefinisikan secara rinci yang selanjutnya akan berfungsi sebagai spesifikasi sistem.
2. Perancangan sistem dan perangkat lunak.

   Tahap perancangan sistem dan perangkat lunak meruapakan tahap mendesain perangkat lunak yang akan dibangun.
3. Implementasi dan pengujian unit.

   Implementasi dan pengujian unit meruapakan tahap perancangan perangkat lunak yang direalisasikan sebagai serangkaian program atau unit program dengan bahasa pemrograman yang sudah ditentukan. Pengujian dilakukan secara unit untuk memastikan setiap unit dari program sudah berfungsi.
4. Integrasi dan pengujian sistem.

   Integrasi dan pengujian sistem merupakan tahap penyatuan unit-unit program menjadi sebuah program utuh. Dilakukan pengujian perangkat lunak secara keseluruhan, sehingga perangkat lunak siap untuk digunakan.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum mengenai penelitian yang dikerjakan. Sistematika penulisan penelitian sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini berisi penjelasan mengenai latar belakang permasalahan, identifikasi masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, serta sistematika.

BAB II LANDASAN TEORI

Pada bab ini berisi teori-teori atau tinjauan dari literatur pendukung yang akan digunakan untuk penerapan algoritma untuk mendeteksi plagiarism dokumen.

BAB III ANALISIS DAN PERANCANGAN

Pada bab ini dijelaskan secara rinci tahapan-tahapan dari mulai melakukan tahap *preprocessing*, menerapkan metode k-gram, menerapkan metoda *synonym*

recognition. Selain itu juga dijelaskan analisis kebutuhan fungsional dan non-fungsional untuk membangun aplikasi yang akan digunakan sebagai untuk penerapan dan pengujian dari tahap-tahap yang dijelaskan di atas.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Pada bab ini ditampilkan implementasi tahap-tahap hasil dari analisis dan perancangan yang sudah dilakukan sebelumnya menjadi sebuah aplikasi. Pada aplikasi yang sudah dibuat dilakukan pengujian alpa yaitu pengujian terhadap fungsionalnya. Lalu aplikasi digunakan untuk melakukan pengujian 2 dokumen dan menghitung jumlah nilai similiarity kesamaan dokumen tersebut.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari semua hal yang di bahas pada bab sebelumnya dan saran untuk pengembangan selanjutnya.