

1. Pendahuluan

1.1 Latar Belakang

Seiring semakin berkembangnya teknologi, kemudahan pengaksesan informasi dari berbagai macam bahasa disediakan melalui World Web Wide. Dampak negatif dari akses bebas ini salah satunya adalah *plagiarism* suatu informasi dari satu bahasa ke bahasa lain.

Plagiarism (plagiatisme) adalah menyalin sepenuhnya suatu paper milik orang lain tanpa menyertakan sumbernya (penulis aslinya) dan kemudian diakui sebagai paper miliknya. *Plagiarism* ini akan sangat merugikan bagi pemilik jika sampai dipublikasikan oleh orang yang tidak berhak.

Deteksi otomatis suatu plagiatisme adalah sesuatu yang sangat dibutuhkan dalam suatu lingkup akademisi. *Plagiarism* ini akan menjadi lebih kompleks jika dibuat dalam bentuk multilingual, yaitu teks asli ditulis ke dalam bahasa yang berbeda. Dari sebuah *cross-lingual perspective*, suatu fragmen teks dalam satu bahasa dianggap plagiat dari suatu teks dalam bahasa lain jika isi teks tersebut secara semantik sama.

Dalam Tugas Akhir ini, analisis *Cross-lingual Plagiarism* berdasarkan pada eksploitasi dari kamus bilingual statistical. Kamus ini dibentuk pada basis corpus paralel yang berisi original fragmen ditulis dalam satu bahasa dan versi plagiat dari fragmen tersebut ditulis dalam bahasa lain.

1.2 Perumusan Masalah

Masalah yang dikedepankan dalam tugas akhir ini adalah bagaimana menerapkan model statistical untuk menganalisa *Cross-lingual Plagiarism*.

Batasan Masalah

Batasan masalah yang didefinisikan dalam penelitian ini adalah :

1. Bahasa yang digunakan adalah bahasa Inggris dan bahasa Indonesia, dengan bahasa Inggris sebagai teks asli.

2. Permasalahan ambiguitas akan ditangani hanya dalam level semantik saja dan tidak menangani sampai level sintatik yaitu hanya menangani sampai pada penerjemahan makna saja dan tidak menangani sampai level struktur dari kata.
3. Dokumen yang digunakan pada sistem temu balik informasi ini dalam bentuk teks.
4. Tidak menangani masalah thesaurus dan frase kalimat.
5. Tidak melibatkan kutipan atau pemberian credit yang dilakukan kepada penulis asli.
6. Mekanisme penerjemahan dari bahasa Inggris ke bahasa Indonesia yaitu dengan menggunakan beberapa *machine translation*, antara lain translate.google.co.id, www.sederet.com, www.toggletext.com.

1.3 Tujuan

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menilai tingkat kemiripan suatu fragmen atau teks antar lintas bahasa.
2. Mengkaji penggunaan IBM Model-1 dalam cross-lingual plagiarism.
3. Mencari nilai threshold yang tepat untuk menentukan mirip tidaknya suatu fragmen atau teks.

1.4 Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah yang akan digunakan adalah :

- a. Studi literatur
Merupakan tahapan dalam mempelajari konsep dan teori pendukung untuk memecahkan permasalahan. Dalam penelitian ini, studi literatur meliputi pembelajaran konsep cross-lingual plagiarism, metode statistical, serta informasi lainnya yang menunjang pembuatan penelitian ini.
- b. Pengumpulan data
Pada tahap ini, dilakukan pencarian dokumen plagiarism, sekaligus dilakukan proses preprocessing pada data tersebut.
- c. Pemodelan sistem
Sistem yang akan dibangun adalah sistem berbasis object-oriented.
- d. Testing dan analisa hasil

1.5 Sistematika Penulisan

Struktur Pembahasan Tugas Akhir ini disusun sebagai berikut :

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan pembahasan, metodologi pemecahan masalah dan sistematika penulisan.

BAB II LANDASAN TEORI

Membahas dasar teori yang berhubungan dengan pengertian umum plagiarism, pemodelan statistical yang terdiri dari IBM Model-1 dan algoritma Expectation Maximization.

BAB III PERANCANGAN PERANGKAT LUNAK

Bab ini akan membahas proses perancangan aplikasi analisis Cross-lingual Plagiarism yang menggunakan model statistical.

BAB IV IMPLEMENTASI DAN ANALISIS HASIL UJI COBA

Membahas tentang analisis dari hasil pengujian ataupun percobaan pada implementasi Cross-lingual Plagiarism.

BAB V KESIMPULAN & SARAN

Pada bab ini akan menjelaskan kesimpulan dan saran sebagai hasil dari analisa dan implementasi Tugas Akhir.