

# 1. Pendahuluan

## 1.1 Latar Belakang

Tugas Akhir (TA) adalah tugas yang harus dilakukan oleh seorang mahasiswa di tahap akhir perkuliahannya untuk membuktikan pemahamannya atas seluruh materi yang telah dipelajari sesuai dengan jurusan mahasiswa tersebut. Selain itu, Tugas Akhir juga dapat dilakukan penelitian terbaru dan pengembangan lebih lanjut terkait bidang studi mahasiswa tersebut. Seiring perkembangan teknologi informasi yang semakin berkembang pesat saat ini sangat mudah dilakukan penjiplakan (plagiarisme) terhadap beberapa Tugas Akhir yang sudah ada. Karya orang lain dicap sebagai karya sendiri, hal tersebut merupakan tindakan kejahatan akademik yang bisa memperburuk dunia pendidikan di negara kita. Karena banyaknya penyalahgunaan oleh mahasiswa yang sedang melakukan Tugas Akhir seperti menjiplak karya Tugas Akhir yang lain (plagiarisme).

Oleh karena itu, diperlukan sebuah sistem yang bisa memberikan keluaran nilai keunikan sebuah proposal Tugas Akhir berdasarkan judul dan abstraknya sehingga mahasiswa bisa memperkirakan TA yang diajukan tidak sama dan belum pernah ada yang menggunakan sebelumnya dengan nilai keunikannya yang mengacu pada nilai kesamaan (*similarity*) rangking pertama dari hasil pencarian. Nilai keunikan yang tinggi membuat semakin besar peluang Tugas Akhir yang diajukan berbeda dengan Tugas Akhir sebelumnya dan pencegahan terhadap tindakan plagiarisme. Salah satu mekanisme sistem tersebut adalah dengan mengimplementasikan sistem pencarian untuk mendeteksi nilai keunikan proposal Tugas Akhir berdasarkan judul dan abstrak menggunakan *Latent Semantic Analysis* (LSA) dengan metode *Semi Discrete Decomposition* (SDD) dan algoritma *O'Leary-Peleg*. LSA memiliki kemampuan untuk merepresentasikan konteks yang digunakan sebagai sebuah arti kata dengan memanfaatkan komputasi statik untuk sejumlah corpus yang besar. Kemudian LSA akan merepresentasikan seluruh kata dalam semua dokumen sebagai baris dalam matriks dan kumpulan dokumen sebagai kolom dalam matriks. Setelah pembuatan matriks yang merepresentasikan kumpulan kata dan kumpulan dokumen yang ada, salah satu tahap penting dari LSA adalah melakukan proses dekomposisi matriks dengan tujuan mereduksi dimensi matriks yang sangat besar menjadi dimensi yang lebih kecil sesuai dengan nilai reduksi yang diinginkan (nilai  $k$ ). Oleh karena itu, untuk melakukan dekomposisi tersebut penulis akan menggunakan metode SDD dan algoritma *O'Leary Peleg*. Dekomposisi SDD dengan menggunakan algoritma *O'Leary Peleg* melakukan penguraian sebuah matriks menjadi tiga bagian, yaitu matriks  $X$ , matriks  $D$ , dan matriks  $Y$ . Ketiga bagian hasil dekomposisi matriks tersebut akan digunakan untuk proses perhitungan nilai kesamaan antara sebuah *query* (judul dan abstrak TA) dengan seluruh dokumen yang ada dalam dataset. Nilai kesamaan (*similarity*) tersebut akan menjadi acuan untuk nilai keunikan.

## 1.2 Perumusan masalah

Berdasarkan latar belakang, dapat dirumuskan masalah sebagai berikut :

1. Bagaimana membangun sistem *Latent Semantic Analysis* (LSA) dengan metode *Semi Discrete Decomposition* (SDD) dan algoritma *O'Leary Peleg* dalam menghasilkan nilai keunikan sebuah judul dan abstrak proposal Tugas Akhir?
2. Bagaimana perbandingan performansi R-Precision dan waktu dari tiga jenis tipe inisial vektor  $y$  (CYC, ONE, PER) pada algoritma *O'Leary-Peleg* dalam menemukan dokumen yang relevan dengan *query* ?

## 1.3 Tujuan

Tujuan yang ingin dicapai dalam Tugas Akhir ini adalah :

1. Merancang dan menganalisis nilai keunikan dari sistem *Latent Semantic Analysis* dengan metode *Semi Discrete Decomposition* (SDD) & algoritma *O'Leary-Peleg* pada dataset yang berasal dari Tugas Akhir Teknik Informatika IT Telkom.
2. Menganalisis perbandingan performansi R-Precision dan waktu dari tiga jenis tipe inisial vektor  $y$  (CYC, ONE, PER) pada algoritma *O'Leary-Peleg* dalam menemukan dokumen relevan dengan *query*.

## 1.4 Batasan Masalah

Batasan permasalahan yang terdapat pada Tugas Akhir ini adalah :

1. Koleksi dataset yang digunakan untuk menguji perbandingan R-Precision dan waktu dari CYC, ONE, dan PER dalam algoritma *O'Leary Peleg* adalah Tugas Akhir (TA) IT Telkom Teknik Informatika kelompok keahlian SIDE ( *Software Engineering, Information System, and Data Engineering* ) dari tahun 2007 - 2008 (140 dokumen). Sedangkan koleksi dataset untuk menguji keunikan adalah Tugas Akhir IT Telkom Teknik Informatika kelompok keahlian SIDE tahun 2007 - 2010 (280 dokumen).
2. Melakukan proses *stop list* pada tahap *indexing* untuk menghilangkan kata yang tidak memiliki makna dan frekuensi kemunculannya sangat sering terjadi pada setiap dokumen serta tidak melakukan proses *stemming* sehingga kata dasar dan kata berimbuhan diasumsikan sebagai dua kata yang berbeda.
3. Deteksi keunikan proposal Tugas Akhir hanya dibatasi pada nilai ketidaksamaan saja yang mengacu pada nilai *cosine similarity* tertinggi pertama dari hasil pencarian.
4. Acuan sistem untuk mendeteksi nilai keunikan Tugas Akhir hanya dibatasi pada judul dan abstrak dalam format *text*.
5. Keluaran (*output*) sistem LSA ini adalah hasil pencarian dokumen terhadap *query* yang dimasukkan oleh *pengguna* dan nilai keunikan dengan jumlah rangking yang telah ditentukan terhadap *query* tersebut.
6. Hasil pencarian dokumen yang relevan dibatasi pada nilai *threshold* (0.50).
7. Nilai keunikan berubah sesuai dengan dokumen yang terpilih.

## 1.5 Metodologi Penyelesaian Masalah

Pada Tugas Akhir ini penulis menerapkan beberapa metode untuk menyelesaikan masalah sebagai berikut:

1. Studi literatur  
Mempelajari konsep-konsep *Latent Semantic Analysis*, *Semi Discrete Decomposition*, dan algoritma *O'Leary-Peleg* yang terdapat pada beberapa sumber literatur. Sumber literatur dapat berupa buku teks, *paper*, *website*, blog, jurnal, dan dokumen teknis sistem terdahulu.
2. Koleksi data  
Penulis meminta secara langsung database Tugas Akhir Teknik Informatika IT Telkom lulusan tahun 2007 - 2010 pada waktu tertentu terhadap institusi melalui izin dari pihak Fakultas Informatika IT Telkom terlebih dahulu untuk dijadikan penelitian bahan Tugas Akhir.
3. Perancangan Sistem  
Merancang sistem menggunakan konsep *Latent Semantic Analysis* dengan metode *Semi Discrete Decomposition* (SDD) dan Algoritma *O'Leary Peleg* dengan menentukan terlebih dahulu deskripsi sistem, kebutuhan *hardware* dan *software*, dan tujuan sistemnya.
4. Implementasi  
Mengimplementasikan *Latent Semantic Analysis* dengan metode *Semi Discrete Decomposition* (SDD) dan Algoritma *O'Leary Peleg* dalam menemukan dokumen yang relevan dan menghasilkan nilai keunikan dari *query* (judul dan abstrak TA) yang diinputkan oleh pengguna.
5. Pengujian dan Analisis  
Penulis akan menguji dan menganalisis perbandingan tiga tipe vektor  $y$  sebagai inisialisasi proses penguraian matriks dalam algoritma *O'Leary Peleg* untuk menentukan solusi nilai optimal vektor  $x$  dan vektor  $y$  di proses *inner iteration*, yaitu CYC, ONE, dan PER terhadap parameter R-Precision dalam menemukan dokumen yang relevan dan menguji perbandingan waktunya. Kemudian menguji dan menganalisis pengaruh tipe CYC, ONE, dan PER dalam menghasilkan nilai keunikan.