

# 1 PENDAHULUAN

## 1.1 Latar Belakang

Pada era digital ini setiap orang dimudahkan untuk mencari dan mengambil informasi. Informasi – informasi tersebut dapat berbentuk berita, paper/journal ataupun artikel. Kesamaan yang ada pada contoh – contoh tersebut adalah semua bentuk penyampaian pendapat tersebut berbentuk data teks. Data teks mudah dibuat oleh siapapun. Hal ini yang membuat data teks memiliki jumlah yang banyak.

Meskipun dengan banyaknya jumlah data teks, pemanfaatan dari data ini belum maksimal terutama pada data teks yang panjang. Salah satu penyebab dari hal ini adalah data teks memerlukan waktu untuk dibaca dan juga dipahami. Padahal terkadang tujuan utama membaca data teks adalah mencari inti topik. Pada dokumen yang tidak memiliki ringkasan ataupun abstrak kita perlu membaca dokumen tersebut untuk mengetahui topik dari dokumen dan proses ini membutuhkan waktu yang lama. Karena itu kita perlu membuat peringkasan teks otomatis sehingga kita dapat membuat ringkasan dari dokumen yang tidak memiliki ringkasan.

Peringkasan dokumen teks bertujuan untuk membuat ringkasan dari suatu dokumen, menjadikan suatu teks lebih pendek tetapi tetap berisi informasi ini yang sama dengan dokumen aslinya. Tujuan peringkasan dokumen adalah meng ekspresikan dokumen dalam kata yang sedikit [1]. Peringkasan berguna untuk mempermudah seseorang memahami isi dari suatu dokumen yang cukup panjang dengan hanya mengambil informasi yang paling penting dari dokumen tersebut.

Terdapat dua jenis text summarization yaitu extraction dan abstraction [2]. Abstraction merupakan jenis text summarization yang masih sulit karena perlu membentuk kalimat sendiri. Meskipun sudah banyak metode alternatif yang diriset, peringkasan menggunakan metode ekstraksi masih menjadi mayoritas [3]. Pada tugas akhir yang di ajukan, jenis teks summarization yang akan di lakukan adalah extraction, dengan menggunakan metode document index graph.

Document index graph merupakan metode peringkasan dokumen dengan cara merepresentasikan dokumen sebagai graf. Dengan representasi ini kita dapat menilai kalimat tidak hanya berdasarkan kata per kata. Penilaian kalimat dalam suatu dokumen akan dihitung berdasarkan kesamaan frasa [4], karena suatu frasa akan membentuk hubungan diantara node yang dapat di kenali oleh komputer. Kemampuan tersebut merupakan kelebihan dari metode document index graph.

Masalah yang ditemukan pada segala jenis pemrosesan bahasa natural adalah ambiguitas kata. Ambiguitas kata berefek pada evaluasi hasil peringkasan dokumen. Ringkasan yang menjadi gold standard dapat menggunakan kata berbeda dengan makna yang sama dengan kata pada ringkasan yang dibuat oleh sistem. Metode document index graph juga mendapatkan efek dari ambiguitas kata, karena itu dalam tugas akhir ini akan diterapkan word sense ambiguity dengan menggunakan metode Lesk.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang diatas berikut rumusan masalah yang akan dibahas dari tugas akhir ini:

1. Bagaimana penerapan metode document index graf pada peringkasan dokumen?
2. Bagaimana cara penggunaan word sense disambiguation pada peringkasan dokumen otomatis?
3. Bagaimana pengaruh penggunaan word sense disambiguation pada peringkasan dengan metode document index graph?

Berikut batasan – batasan masalah pada tugas akhir ini:

1. Peringkasan di lakukan hanya pada dokumen tunggal.
2. Dokumen yang diringkas menggunakan bahasa inggris.
3. Hasil peringkasan berjenis ringkasan ekstraktive.
4. Evaluasi hasil ringkasan menggunakan toolkit ROUGE.

## 1.3 Tujuan

Berikut tujuan dari tugas akhir ini:

1. Menerapkan metode document index graph untuk melakukan peringkasan dokumen otomatis.
2. Melakukan penerapan *word sense disambiguation* pada peringkasan dokumen otomatis.
3. Membandingkan hasil peringkasan dengan metode document index graph dengan mengimplementasikan word sense disambiguation dan hasil peringkasan dengan hanya menggunakan document index graph.

## 1.4 Metodologi Penelitian

Tahapan yang akan dilakukan untuk menyelesaikan tugas akhir ini adalah sebagai berikut:

1. Studi literature, mempelajari literature yang relevan dengan tugas akhir ini yaitu text summarization, word sense disambiguation, document index graph, lesk algorithm, ROUGE analisis.
2. Analisis permasalahan, mencakup kekurangan metode document index graph, representasi dokumen yang akan diproses, penggunaan word sense disambiguation pada peringkasan teks.
3. Mengumpulkan kebutuhan tugas akhir, yaitu toolkit NLTK dan ROUGE, dataset teks berbahasa inggris beserta ringkasan yang dibuat manual, tool/software yang akan digunakan dalam tugas akhir.
4. Mengimplementasikan sistem yang telah dirancang pada proposal.
5. Menguji hasil implementasi sistem, menggunakan jumlah data yang kecil
6. Melakukan evaluasi dan analisis pada hasil ringkasan sistem.
7. Mengambil kesimpulan dari hasil penelitian yang telah di lakukan.

## **1.5 Sistematika Penulisan**

Berikut sistematika penulisan pada buku tugas akhir ini:

### **1 Pendahuluan**

Bagian pendahuluan dimulai dengan latar belakang dari penilitan, dilanjutkan dengan perumusan masalah, batasan masalah, tujuan, dan metodologi penelitian.

### **2 Studi Literatur**

Bagian studi literatur berisi ringkasan tinjauan pustaka dari rujukan yang terkait dengan tujuan dari penelitian yang dilakukan. Ringkasan tinjauan pustaka berisi teori atau metode yang akan digunakan dalam penelitian.

### **3 Perancangan Sistem**

Bagian perancangan sistem berisi rancangan sistem yang dibangun, rancangan sistem dijelaskan dengan diagram flowchart, dan deskripsi tiap proses. Bagian ini juga berisi kebutuhan data dan kebutuhan sistem pada penelitian ini. Scenario pengujian juga perlu dituliskan pada bagian ini.

### **4 Pengujian dan Analisis**

Bagian pengujian dan analisis berisi hasil pengujian yang dilakukan menggunakan scenario pada bagian perancangan sistem. Bagian ini juga berisi analisis terhadap hasil pengujian.

### **5 Kesimpulan dan Saran**

Bagian kesimpulan dan saran berisi kesimpulan dari hasil penelitian yang dilakukan dan saran terhadap penelitian yang dapat dilakukan lebih lanjut.