

# 1. Pendahuluan

## 1.1. Latar Belakang

Informasi – informasi tersebut dapat berupa teks, gambar, video, dan lainnya. Sehingga dengan internet semua orang dapat membagikan ide mereka dengan bebas. Banyak informasi yang dibagikan di internet menggunakan bahasa Inggris, sehingga sebagian orang yang baru mempelajari bahasa tersebut sulit untuk memahami informasi tersebut. Dapat juga menyebabkan perbedaan pemahaman terhadap informasi tersebut, karena pembaca hanya memahami sebagian dari informasi yang diberikan. Hal tersebut juga berpengaruh pada *second language learner*. *Second language learner* adalah orang yang mempelajari bahasa selain bahasa aslinya [1]. Penyederhanaan teks dapat membantu mereka untuk mengerti dan sambil mempelajari suatu bahasa yang baru.

Oleh karena itu, dibutuhkan bantuan yang membuat bahasa yang lebih mudah dimengerti tetapi tidak merubah makna asli dari isi teks tersebut. Sehingga dikembangkan sebuah sistem yang mampu menyederhanakan teks yaitu, penyederhanaan teks otomatis (*Automatic Text Simplification*) [2]. *Text Simplification* di implementasikan dengan menganalisis kata atau kalimat yang sulit kemudian menyederhanakannya tanpa mengurangi makna dari kalimat tersebut. *Text Simplification* adalah proses untuk mengurangi kompleksitas sehingga teks dapat lebih mudah dibaca dan dimengerti [1]. Pendekatan yang pertama kali digunakan adalah dengan mengganti kata yang sulit dengan sinonimnya yang lebih mudah dipahami [3]. Tingkat kesulitan suatu kata dapat diukur dengan frekuensi kata tersebut, semakin kecil nilai frekuensi suatu kata maka kata tersebut dianggap semakin sulit. Selain itu dapat juga dengan mengukur panjang dari setiap kata, semakin banyak jumlah huruf pada kata maka kata tersebut dianggap semakin sulit.

Tugas akhir ini menggunakan fokus pada metode *Lexical Simplification* dan *Substitution Generation* yang menangani teks dalam bahasa Inggris. *Lexical Simplification* bertujuan untuk mengganti kata yang sulit dengan kata yang lebih mudah dimengerti. Selain itu *Substitution Generation* membantu memberikan penjelasan dari kata yang dianggap sulit. Penelitian ini menganalisis pengaruh penggunaan metode pada *Complex Word Identification* untuk menyeleksi kata yang dianggap sulit, *Substitution Generation* untuk mencari kata pengganti, *Substitution Selection* untuk menyeleksi kata yang tidak berhungan dengan kalimat asalnya, dan *Substitution Ranking* untuk mengambil kata yang dianggap paling sederhana. Agar tidak mengubah makna aslinya maka pada tahap *Substitution Selection*, penelitian ini menggunakan *Word Sense disambiguation* dengan pendekatan *Knowledge-based*.

## 1.2. Perumusan Masalah

Berdasarkan pada latar belakang diatas, rumusan masalah yang akan diuraikan dan diteliti adalah sebagai berikut:

1. Bagaimana pengaruh penggunaan frekuensi kata dari Wikipedia dan Simple Wikipedia terhadap *Complex Word Identification* dan *Subtitution Ranking* ?
2. Bagaimana pengaruh NER (*Named Entity Recognizer*) terhadap *Complex Word Identification* dan *Subtitution Ranking* ?
3. Bagaimana pengaruh *Word Sense Disambiguation* dengan pendekatan *Knowledge-Based* terhadap *Subtitution Selection* ?
4. Bagaimana mengimplementasikan *Text Simplification* yang didapatkan dari penggabungan metode *Complex Word Identification*, *Subtitution Generation*, *Subtitution Selection*, dan *Subtitution Ranking* ?

Dalam penelitian Tugas Akhir ini, terdapat beberapa batasan masalah dalam pembahasan dan pengerjaannya, diantaranya :

1. Sistem yang akan dibangun menggunakan bahasa pemograman Python.
2. Masukkan program berupa teks yang menggunakan bahasa Inggris.
3. Pada pengujian digunakan dataset NNSeval yang terdiri dari 239 kalimat.
4. Sistem yang akan dibangun dibatasi hanya hanya *Lexical Simplification*
5. *Knowledge* yang digunakan dalam penelitian ini adalah Wordnet.
6. Korpus yang digunakan dalam penelitian ini adalah Wikipedia dan Simple Wikipedia

## 1.3. Tujuan

Tujuan yang ingin dicapai dalam pengerjaan Tugas Akhir ini sebagai berikut :

1. Menganalisis dan membandingkan pengaruh penggunaan frekuensi kata dari Wikipedia dan Simple Wikipedia terhadap *Complex Word Identification* dan *Subtitution Ranking*.
2. Menganalisis dan membandingkan pengaruh *preprocessing* menggunakan NER (*Named Entity Recognizer*) pada *Complex Word Identification* dan *Subtitution Ranking*.
3. Menganalisis dan membandingkan metode *Word Sense Disambiguation* dengan pendekatan *Knowledge-Based* terhadap *Subtitution Selection*.
4. Menganalisis dan mengimplementasikan *Text Simplification* dengan menggabungkan *Complex Word Identification*, *Subtitution Generation*, *Subtitution Selection*, dan *Subtitution Ranking*.

## 1.4. Sistematika Penulisan

Untuk mendapatkan hasil akhir yang sesuai dengan tujuan, disusunlah metodologi penelitian sebagai berikut :

1. Studi Literatur  
Dilakukan untuk memahami literatur yang berhubungan dengan tugas akhir ini yaitu *Text Simplification*, *Lexical Simplification*, *Complex Word Identification*, *Subtitution Generation* dan *Selection*, *Subtitution Ranking*, algoritma Leks, *F-Score evaluation*, dan *TRank evaluation*.
2. Pengumpulan Data  
Dilakukan pencarian dan pengumpulan data yang dibutuhkan. Data yang digunakan adalah *dataset* dalam bahasa Inggris dengan hasil penyederhanaannya yang di buat secara manual. *Dataset* ini akan digunakan dalam tugas akhir.
3. Analisis dan Perancangan Sistem  
Dilakukan analisis metode dan perancangan sistem dimulai dari input hingga menghasilkan keluaran berupa teks yang sudah disederhanakan berdasarkan referensi yang ada.
4. Implementasi Sistem  
Implementasi dari rancangan sistem yang telah dibuat.
5. Pengujian dan Analisis  
Penyelesaian masalah dan pengujiannya dan melakukan analisis dari hasil output sistem.
6. Kesimpulan  
Membuat kesimpulan hasil penelitian berdasarkan tujuan yang telah ditentukan sebelumnya.
7. Penyusunan Laporan  
Tahap pembuatan laporan tugas akhir berdasarkan pengerjaan penelitian yang sudah dilakukan.