

1 PENDAHULUAN

1.1 Latar Belakang

Text simplification dapat didefinisikan sebagai proses memahami aspek yang membuat teks menjadi sulit dan menemukan solusi untuk mengatasinya dengan memodifikasi struktur teks sehingga menjadi lebih sederhana untuk dipahami oleh *user*. Berdasarkan sifat kelinguistikannya, *text simplification* dapat diklasifikasikan menjadi tiga kategori, yaitu *lexical simplification*, *syntactic simplification*, *discourse simplification*. Adapun yang akan dibahas pada penelitian ini adalah *text simplification* dengan pendekatan leksikal (*lexical simplification*) [1].

Lexical simplification pada dasarnya adalah suatu proses penyederhanaan teks (*Text Simplification*) dengan mengidentifikasi dan mengganti kata sulit dengan kata alternatif (sinonim) yang lebih sederhana atau mudah dimengerti [1]. *Lexical simplification* merupakan pendekatan paling simpel daripada pendekatan lainnya [2]. Namun, Paetzold dan Specia berpendapat bahwa *lexical simplification* sangat menantang karena hasil substitusi harus mempertahankan keaslian makna dan tata bahasa kalimat yang disederhanakan [3]. Pada teks jurnal ilmiah, catatan medis atau bisnis seringkali ditemui istilah yang tidak memiliki sinonim. *Lexical simplification* bisa menjabarkan definisi istilah tersebut secara langsung.

Terdapat empat langkah untuk mengimplementasikan *lexical simplification*. Pertama, kata-kata kompleks pada suatu teks harus diidentifikasi. Kedua, bangkitkan list kata pengganti/substitusi setiap kata kompleks yang sudah diidentifikasi. Ketiga, seleksi kata pengganti pada list sehingga kata-kata tersebut tidak memberikan arti yang ambigu. Terakhir, kata pengganti yang tersisa diberi peringkat kemudian tentukan mana yang paling sederhana dan umum digunakan [2].

Shardlow menjelaskan bahwa salah satu masalah utama penerapan *lexical simplification* adalah kehilangan makna asli kalimat akibat keambiguan. Hal tersebut terjadi ketika sebuah kata mengandung beberapa arti atau sinonim sehingga sistem mengambil kata pengganti yang tidak relevan dengan makna kalimat. Oleh karena itu, diperlukan metode menyeleksi kata-kata substitusi untuk menghindari keambiguan [2].

Penelitian terdahulu berhasil melahirkan beberapa metode penyeleksian kata substitusi antara lain dengan menggunakan pendekatan *word sense disambiguation* atau dengan membuang kata substitusi yang memiliki *Part of Speech Tags (POS tags)* tidak sama dengan kata kompleks [2, 4, 5].

Nunes et. al. melakukan penelitian membuat sistem *lexical simplification* dengan menggunakan pendekatan *word sense disambiguation*. Fungsi *word sense disambiguation* digunakan pada saat memfilter kata sinonim yang memiliki *part-of-speech tag* yang sama. Penelitian tersebut berhasil memperoleh presisi yang cukup baik. Dengan menggunakan *wordnet* sebagai sumber data kata sinonim, rata-rata skor presisi (persamaan makna) yang diperoleh sebesar 81% dan rata-rata skor presisi (kebenaran gramatikal) sebesar 61,25% [5].

Adapun yang dilakukan pada penelitian tugas akhir ini adalah menganalisa pengaruh *word sense disambiguation* dalam penyeleksian kata substitusi pada sistem *lexical simplification*. Pada penelitian ini dibuatkan simulasi sistem *lexical simplification* sederhana. Pada tahap penyeleksian kata substitusi akan diterapkan

word sense disambiguation. Hasil penyeleksian akan dievaluasi untuk mengetahui seberapa besar pengaruh *word sense disambiguation* terhadap sistem dalam membangkitkan kata substitusi yang relevan dengan makna asli kalimat yang ingin disederhanakan.

1.2 Perumusan Masalah

Rumusan masalah dalam tugas akhir ini adalah bagaimana pengaruh *word sense disambiguation* pada *lexical simplification* ?

1.3 Tujuan

Tujuan dari pembuatan tugas akhir ini adalah membangun simulasi sistem *lexical simplification* sederhana dan menganalisis pengaruh *word sense disambiguation* dalam menyeleksi kata substitusi sehingga dapat membangkitkan kata substitusi yang relevan dengan makna kalimat yang ingin disederhanakan.

1.4 Metodologi Penelitian

Metode penelitian yang digunakan pada perancangan sistem ini adalah :

1. Studi literatur, dengan mempelajari referensi dan dokumentasi yang bersumber dari *Internet*, buku, jurnal, artikel, dan lain-lain yang meliputi :
2. Studi lapangan, dengan berkonsultasi kepada orang yang lebih berpengalaman di bidang *software* dan *data engineering* serta linguistik, khususnya *lexical simplification* dan literatur bahasa Inggris seperti dosen, senior, dan forum komunitas yang berkonsentrasi di bidang yang sama.
3. Desain dan analisis kebutuhan, dengan memperkirakan apa saja yang dibutuhkan pada saat pembuatan dan pada saat waktu pengerjaan sistem, baik *hardware* maupun *software*.
4. Pengujian dan analisis perbandingan performa sistem yang telah dibuat.
5. Penyusunan laporan tugas akhir yang berfungsi sebagai dokumentasi terhadap proses dan hasil penelitian yang telah dilakukan mengikuti kaidah institusi yang ditetapkan.