

BAB I

Pendahuluan

1.1. Latar Belakang

Di era digital seperti saat ini, informasi merupakan salah satu hal yang tidak dapat dipisahkan dari masyarakat global. Salah satu bentuk informasi tersebut adalah berita. Setiap orang dari segala umur bahkan dapat memperoleh berita tanpa harus beranjak keluar dari tempat duduknya. Namun, tidak semua orang dapat langsung mengolah dan mencerna yang telah didapat. Bagi anak-anak, penggunaan bahasa yang tidak umum atau tidak sering didengar oleh anak yang digunakan dalam suatu teks berita dapat menjadi penghambat mereka untuk mengerti dan memahami makna sebenarnya yang ingin disampaikan. Menurut Nugroho (2015), penguasaan kosa kata dan struktur kalimat memiliki pengaruh yang signifikan terhadap pemahaman membaca. Penguasaan kosa kata juga merupakan hal yang fundamental karena apabila suatu kosa kata tidak dapat dipahami, kosa kata tersebut dapat menjadi penghalang utama dalam komunikasi antara pemberi informasi dan penerima informasi [1]. Disisi lain, penyediaan kanal berita khusus anak yang membutuhkan waktu yang tidak singkat untuk tetap berjalan [2] mengakibatkan sangat sedikitnya kanal berita yang diperuntukkan khusus untuk anak-anak.

Penyederhanaan teks (*Text Simplification*) dapat digunakan untuk membantu menyelesaikan masalah tersebut. Penyederhanaan teks sendiri merupakan bagian dari pengembangan sistem pemrosesan bahasa natural (*Natural Language Processing*) yang ditujukan untuk melakukan penyederhanaan kalimat dengan mengubah kompleksitas sintaksis atau leksikal dari sebuah kalimat dengan mempertahankan arti semula dari kalimat tersebut [3].

Penyederhanaan teks yang difokuskan dalam tugas akhir ini adalah penyederhanaan teks pada tingkatan leksikal dan tingkatan sintaksis dari sebuah teks. Penyederhanaan leksikal adalah proses untuk mengganti kata-kata kompleks dalam sebuah kalimat menjadi kata alternatif yang memiliki arti yang sama [4] seperti contoh yang terdapat dalam tabel 1. Sedangkan penyederhanaan sintaksis adalah proses untuk menurunkan kompleksitas susunan kalimat dari sebuah teks dengan tetap mempertahankan informasi yang dikandungnya [5] seperti contoh yang terdapat dalam tabel 2.

Tabel 1 : Simplifikasi Leksikal

Masukkan / <i>Input</i>	<i>Akhir bulan Oktober, Doni akan melakukan lawatan bersama keluarganya.</i>
Keluaran / <i>Output</i>	<i>Akhir bulan Oktober, Doni akan melakukan perjalanan bersama keluarganya.</i>

Tabel 2 : Simplifikasi Sintaksis

Masukkan / <i>Input</i>	<i>Bukan saja dia tidak menyapa, menoleh pun dia tidak.</i>
Keluaran / <i>Output</i>	<i>Dia tidak menyapa. Dia tidak menoleh.</i>

Dalam proses penyederhanaan teks tingkatan leksikal, terdapat kemungkinan munculnya kata pengganti yang tidak sesuai dengan konteks kalimat yang dapat mengakibatkan perubahan makna kalimat dari makna aslinya. *Word sense disambiguation* merupakan suatu metode yang dikembangkan untuk menentukan makna sebenarnya dari suatu kata ambigu dalam suatu kalimat / konteks tertentu [6]. Pada penelitian sebelumnya, Natadiarta, et al. (2017) menggunakan metode *word sense disambiguation* dengan pendekatan berbasis pengetahuan untuk melakukan penyederhanaan teks berbahasa Inggris. Namun, pendekatan tersebut sulit diterapkan untuk penyederhanaan teks berbahasa Indonesia dikarenakan masih terbatasnya basis pengetahuan yang tersedia untuk bahasa Indonesia. Oleh karena itu, metode *word sense disambiguation* yang akan digunakan dalam penelitian ini adalah metode yang dikembangkan oleh Pelevina, et al. (2016) yaitu *word sense disambiguation* dengan pendekatan berbasis *word embeddings*.

Untuk proses penyederhanaan teks tingkatan sintaksis, salah satu metode yang dapat diterapkan adalah metode berbasis aturan (*Rule-based Reasoning*). Metode berbasis aturan bekerja dengan cara melakukan perintah “*if-then-else*” berdasarkan set perintah yang sudah didefinisikan oleh manusia. Metode penyederhanaan teks berbasis aturan untuk teks Berbahasa Indonesia sebelumnya sudah pernah diterapkan oleh Widyastuti, et al. (2014).

Penelitian pada Tugas Akhir ini akan menitikberatkan pada kombinasi penggunaan *Word Sense Disambiguation* berbasis *Word Embeddings* dengan model *word2vec* yang dilatih berdasarkan korpus yang dihimpun dari situs Wikipedia Bahasa Indonesia untuk penyederhanaan teks tingkat leksikal dan penggunaan metode berbasis aturan untuk penyederhanaan teks tingkat sintaksis.

1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan di atas, maka hal tersebut dapat dirumuskan menjadi beberapa masalah sebagai berikut:

1. Bagaimana pengaruh penggunaan *Word Sense Disambiguation* dengan pendekatan berbasis pengetahuan pada proses penyederhanaan leksikal pada hasil penyederhanaan teks?
2. Bagaimana pengaruh penggunaan metode berbasis aturan pada proses penyederhanaan sintaksis teks pada hasil penyederhanaan teks?
3. Bagaimana pengaruh penggabungan metode *Word Sense Disambiguation* dengan pendekatan berbasis pengetahuan dan metode berbasis aturan pada proses penyederhanaan teks berita?
4. Apakah proses penyederhanaan teks yang dihasilkan dapat membantu anak dalam lebih memahami isi berita dengan lebih baik?

Dalam penyusunan Tugas Akhir ini, terdapat beberapa batasan dalam proses pembahasan dan pengerjaannya, yaitu :

- a. Bahasa yang digunakan dalam penelitian ini menggunakan bahasa pemrograman *Python*.
- b. Penyederhanaan teks yang dilakukan hanya pada tingkatan leksikal dan tingkatan sintaksis.
- c. Korpus yang akan digunakan dalam penelitian ini dihimpun dari situs Wikipedia Bahasa Indonesia.

1.3. Tujuan

Penelitian ini dilaksanakan dengan tujuan untuk :

- a. Mengetahui pengaruh penggunaan metode *Word Sense Disambiguation* dengan pendekatan berbasis *Word Embeddings* dan metode berbasis aturan pada proses penyederhanaan teks berita.
- b. Mengetahui apakah proses penyederhanaan teks berita dapat membantu anak dalam lebih memahami isi suatu berita dengan lebih baik.
- c. Membangun korpus penyederhanaan teks untuk Bahasa Indonesia.

1.4. Rencana Kegiatan

Agar proses perancangan Tugas Akhir ini sesuai dengan tujuan, ada beberapa tahapan penyusunan yang akan dilakukan, yaitu :

- a. Studi Pustaka

Studi pustaka dilakukan untuk memahami metode-metode yang akan digunakan dalam Tugas Akhir ini.

b. Pengumpulan Data

Data yang dikumpulkan dalam Tugas Akhir ini adalah teks berita dalam Bahasa Indonesia dengan hasil penyederhanaannya yang dibuat secara manual yang akan digunakan sebagai korpus dalam proses penyederhanaan tingkat leksikal dan proses penyusunan aturan dalam metode berbasis aturan.

c. Analisa dan Perancangan Sistem

Analisis metode dan perancangan sistem dalam perancangan Tugas Akhir ini akan dilakukan mulai dari proses masukkan / *input* data hingga menghasilkan keluaran / *output* berupa teks yang telah dilakukan penyederhanaan.

d. Implementasi Sistem

Implementasi sistem dari hasil analisa dan perancangan sistem menjadi sistem yang dapat dijalankan.

e. Pengujian dan Analisis Hasil Pengujian

Keluaran (*Output*) dari sistem yang telah dibangun akan diujikan dengan cara dilakukan wawancara kepada anak-anak dengan membandingkan kedua teks sebelum dan setelah dilakukan penyederhanaan. Kemudian, hasil pengujian akan dianalisis dan dievaluasi apakah algoritma dan basis aturan (dalam metode berbasis aturan) yang digunakan sudah tepat dan cocok untuk menyelesaikan masalah penyederhanaan teks atau tidak.

f. Kesimpulan

Penarikan kesimpulan dari hasil penelitian akan ditarik berdasarkan tujuan yang telah ditentukan sebelumnya dengan memperhatikan hasil pengujian dan analisis hasil pengujian sistem.

1.5. Jadwal Kegiatan

Nama Kegiatan	Mei				Jun				Jul				Agu				Sep			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pembangunan <i>Dataset</i>																				
Pembangunan Sistem																				
Evaluasi Hasil Penyederhanaan (Wawancara)																				
Analisis Hasil Wawancara																				
Penarikan Kesimpulan																				