

BAB I Pendahuluan

1.1 Latar Belakang

Setiap orang pasti memiliki penilaian yang berbeda mengenai seberapa besar kesamaan makna dari sepasang kata bahasa Indonesia. Pencarian informasi mengenai kesamaan makna atau semantik pada kata bahasa Indonesia merupakan hal yang baru. Keterbatasan sumber data menjadi tantangan tersendiri dalam penelitian ini, namun kendala tersebut masih dapat dilalui dengan memanfaatkan kamus bahasa Indonesia yang ada.

Kesamaan semantik adalah proses yang digunakan untuk memperkirakan kekuatan hubungan semantik antara unit bahasa, konsep atau contoh seperti mengetahui kemiripan antara ayah dan bapak, melalui deskripsi numerik yang diperoleh sesuai dengan perbandingan informasi pendukung makna atau menggambarkan sifat [1]. Kesamaan antara dua kata atau kalimat berupa angka menggambarkan kedekatan makna antara kedua kata atau kalimat tersebut. Perhitungan kesamaan digunakan dalam berbagai keperluan, misalnya untuk melakukan pencarian informasi di Internet, klasifikasi dokumen dalam arsip, dan kegiatan menganalisis informasi di dunia maya (*data analysis*) [12]. Salah satu metode yang digunakan untuk mengukur nilai kesamaan semantik adalah dengan menggunakan metode berbasis vektor.

Metode berbasis vektor merupakan salah satu metode yang dapat digunakan untuk mengukur kesamaan semantik dari sepasang dokumen. Berdasarkan definisi kata yang didapat dari Kamus Besar Bahasa Indonesia (KBBI), dibentuk vektor yang merepresentasikan kata tersebut. Dengan menggunakan fungsi kosinus yang terdapat pada metode berbasis vektor, dapat diukur seberapa besar kesamaan semantik dari sepasang kata berdasarkan sudut vektornya.

Fungsi kesamaan kosinus adalah kesamaan antara dua teks yang berasal dari nilai kosinus antara dua vektor teks kata [2]. Fungsi ini merupakan fungsi yang umum digunakan pada metode berbasis vektor. Untuk mencerminkan betapa pentingnya sebuah kata dalam sebuah kumpulan dokumen atau korpus [13], maka digunakanlah pembobotan *term frequency-inverse document frequency* (tf-idf) sebagai faktor bobot.

Pada studi kasus kali ini akan dilakukan implementasi dalam bentuk aplikasi untuk mengukur kesamaan dari sepasang kata bahasa Indonesia, dengan metode berbasis vektor, dihitung menggunakan fungsi kosinus, melalui pembobotan tf-idf, menggunakan basis pengetahuan KBBI.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, berikut rumusan masalah dalam Tugas Akhir ini yaitu:

1. Bagaimana menghitung nilai kesamaan semantik pada pasangan kata bahasa Indonesia?
2. Bagaimana mengimplementasikan kesamaan semantik pasangan kata bahasa Indonesia dengan menggunakan metode berbasis vektor?
3. Bagaimana korelasi antara nilai kesamaan semantik dengan nilai dari *gold standard*?
4. Apa parameter terbaik yang mempengaruhi nilai kesamaan semantik pada pasangan kata bahasa Indonesia?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah disampaikan, berikut tujuan dari penelitian yang dibangun, antara lain:

1. Melakukan perhitungan nilai kesamaan semantik pada pasangan kata bahasa Indonesia.
2. Mengetahui apakah metode berbasis vektor dapat diimplementasikan dalam menghitung kesamaan semantik pasangan kata bahasa Indonesia.
3. Menganalisis nilai korelasi dengan membandingkan nilai kesamaan semantik dengan nilai dari *gold standard*.
4. Menganalisis parameter terbaik yang mempengaruhi nilai kesamaan semantik pasangan kata bahasa Indonesia.

1.4 Batasan Masalah

Batasan masalah dari penelitian yang akan dibangun, antara lain:

1. Basis pengetahuan yang digunakan adalah KBBI daring tiga.
2. Bahasa yang digunakan hanya bahasa Indonesia.
3. Hanya menggunakan definisi yang pertama kali muncul dari API kateglo.
4. Menggunakan unit linguistik definisi dan sinonim sebagai parameter pengujian.

1.5 Metodologi Penyelesaian Masalah

Metodologi berikut digunakan untuk menyelesaikan Tugas Akhir:

1. Studi Literatur
Melakukan studi literatur dari berbagai sumber seperti paper, jurnal dan internet untuk mendapatkan informasi mengenai kesamaan semantik dengan metode berbasis vektor dan hal lain yang dapat membantu untuk penyusunan Tugas Akhir ini.
2. Pengumpulan Data
Data yang digunakan didapat dengan menggunakan dari KBBI daring 3 melalui API kateglo.com. Kemudian *gold standard* didapatkan dari hasil *crowdsourcing* kuesioner.
3. Implementasi
Implementasi dilakukan setelah dataset yang digunakan telah siap, kemudian sistem yang akan dibangun dirancang sesuai dengan metode

yang digunakan, lalu membangun sistem berdasarkan pada desain dan analisis yang telah dibuat sebelumnya. Sistem akan dibangun dengan menggunakan bahasa pemrograman java.

4. Pengujian

Pengujian sistem dilakukan untuk meminimalisir *error* serta untuk mengetahui sistem dapat bekerja dengan semestinya.

5. Pembuatan Laporan

Membuat laporan Tugas Akhir dari tahap metodologi yang telah dilakukan.

1.6 Sistematika Penulisan

Tugas akhir ini disusun dengan sistematika penulisan sebagai berikut:

1. Pendahuluan

Dalam bab ini, dijelaskan latar belakang masalah yang diangkat, dijelaskan pula rumusan masalah, tujuan, batasan masalah dan metodologi penyelesaian masalah dari Tugas Akhir ini.

2. Dasar Teori

Dalam bab ini, dijelaskan teori-teori yang mendukung dan terkait dalam Tugas Akhir ini.

3. Perancangan Sistem

Bab ini, berisi gambaran umum sistem yang dibangun, seperti *preprocessing*, perhitungan kosinus dan korelasi, dan alur proses berupa *flowchart*.

4. Pengujian dan Analisis

Bab ini, berisi skenario pengujian, hasil pengujian dan analisis dari pengujian.

5. Kesimpulan dan Saran

Dalam bab ini, dijelaskan hasil dan kesimpulan dari penelitian Tugas Akhir yang telah dilakukan dan saran yang membangun untuk ke depannya.