

1. Pendahuluan

Pada bab ini menjelaskan latar belakang dari penelitian yang dilakukan, menjelaskan perumusan masalah, tujuan dan juga sistematika penulisan yang diterapkan dalam penelitian ini.

1.1 Latar Belakang

Al-Quran merupakan kitab suci Umat Islam, yang telah menjadi panduan hidup bagi kurang lebih 1,6 Miliar Muslim (Sebutan untuk pemeluk agama Islam) [1]. Dalam kepercayaan umat Islam, Al-Quran merupakan sumber hukum tertinggi yang merupakan perintah langsung dari Allah (Tuhan).

Namun, penelitian di bidang linguistik (ilmu tentang bahasa) [2] tentang Al-Quran masih sangat terbatas, padahal sebagai kumpulan teks yang mengandung informasi yang sangat penting, tentunya penelitian linguistik dengan menggunakan Al-Quran sebagai objek sangat penting dan bermanfaat terutama untuk umat Islam.

Terkait dengan penelitian di bidang linguistik khususnya pada *Natural Language Processing* (NLP) dalam beberapa tahun belakangan, perhitungan kesamaan semantik menjadi topik yang sangat menarik. Kesamaan semantik menggunakan dua unit bahasa untuk dibandingkan (bisa berupa paragraf, kalimat, frase ataupun kata) kemudian dinilai seberapa dua unit ini berbagi atribut atau properti yang sama. Semakin banyak dua unit ini memiliki karakteristik yang sama maka semakin tinggi pula nilai kesamaan semantik diantara keduanya [3]. Contohnya adalah menentukan seberapa besar nilai kesamaan semantik antara kata *Fire* dengan *Flame*.

Pendekatan *Knowledge-based* merupakan salah satu pendekatan yang sering digunakan dalam melakukan perhitungan semantik antar kata. Pendekatan ini menggunakan perhitungan untuk menentukan apakah dua kata saling berhubungan dengan menggunakan informasi yang didapatkan dari jaringan semantik [4]. Metode yang akan digunakan untuk menghitung nilai kesamaan semantik adalah metode

Resnik, Wu and Palmer, dan Lin. Ketiga metode ini menggunakan Jaringan semantik *Wordnet*.

Wordnet Merupakan sebuah *Database* kamus bahasa Inggris yang dikembangkan oleh Princeton University. Perbedaan yang terletak antara *Wordnet* dan kamus biasa adalah *Wordnet* berfokus pada makna kata sedangkan kamus biasa berfokus pada kata itu sendiri. Satu makna kata dalam *Wordnet* dinyatakan dalam *Synset (Synonym set)* [5]. Pada tugas akhir ini dibangun sebuah sistem yang menentukan nilai kesamaan semantik antara dua pasang kata di dalam Al-Quran dan dihitung korelasinya menggunakan *Pearson Correlation*.

Kedepannya diharapkan dengan penelitian ini dapat membawa permasalahan yang lebih menarik terkait dengan NLP pada teks Al-Quran, seperti *Text simplification* dalam teks Al-Quran agar lebih mudah dipahami, *Questioning and answering* dan juga *Text summarization* dengan menggunakan teks Al-Quran.

1.2 Perumusan Masalah

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

1. Bagaimana menganalisis kesamaan semantik diantara dua kata di dalam Al-Quran dengan menggunakan pendekatan *knowledge-based*.
2. Bagaimana menganalisis perbandingan antar tiap metode di dalam *knowledge-based*

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

1. Konsep term yang digunakan adalah kata atau frase.
2. Dataset yang digunakan adalah 200 pasang kata yang diambil dari penelitian berjudul “*Quranic Concept Similiarity Based on Lexical Database*”.

3. *Knowledge-Based* yang digunakan adalah *Wordnet*.
4. *Part of speech* yang digunakan adalah *Noun*.
5. *Input* dari sistem ini adalah dua pasang *Noun* dari Quran *Gold Standard*
6. *Output* dari sistem ini adalah skor kesamaan antar kedua noun yang dimasukkan.
7. Perhitungan korelasi menggunakan *Pearson Correlation*.

1.3 Tujuan

Adapun tujuan dari Tugas akhir ini adalah :

1. Menganalisis penggunaan berbagai metode dengan pendekatan *knowledge-based* pada *Wordnet* untuk penghitungan kesamaan semantik antar kata-kata di dalam Al-Quran.
2. Melakukan perhitungan nilai kesamaan semantik dengan berbagai metode dengan pendekatan *knowledge-based*
3. Melakukan perhitungan nilai korelasi antara nilai kesamaan yang dihasilkan dengan *Gold Standard*..

1.4 Sistematika Penulisan

Penulis menggunakan metodologi berikut untuk menyelesaikan tugas akhir ini:

1. Studi literatur

Pada tahap awal dari penyusunan tugas akhir ini dilakukan pengumpulan informasi dari berbagai sumber seperti paper, buku, jurnal, internet, dan lain-lain tentang semantik antar kata dan juga penelitian yang telah dilakukan sebelumnya yang berkaitan dengan masalah yang ingin diselesaikan.

2. *Data preprocessing*

Pada tahap ini, dikumpulkan data-data dari sumber yang tersedia. Pada tahap ini juga dilakukan proses untuk memproses data yang telah ada agar dapat diolah pada tahapan selanjutnya.

3. Implementasi

Implementasi dilakukan pada tahap ini. Arsitektur, metode maupun komponen lain pada sistem dibangun pada tahap ini untuk memecahkan masalah yang telah diidentifikasi sebelumnya.

4. Pengujian dan Analisis

Pada tahapan pengujian dilakukan pengujian terhadap sistem yang telah dibangun yang bertujuan untuk menentukan keakurasian sistem yang kemudian dianalisis sesuai dengan tujuan tugas akhir.