

1. Pendahuluan

1.1. Latar Belakang

Alquran merupakan kitab suci umat Islam yang dijadikan sebagai sumber utama serta pedoman hidup. Alquran merupakan teks yang sangat populer dan dipercaya sebagai perkataan yang sakral dari Tuhan oleh lebih dari 1,6 milyar umat Islam di dunia [1]. Alquran terdiri dari 30 Juz, 114 Surat dan 6236 ayat dengan mayoritas ayat berupa kalimat yang pendek. Pada 6236 ayat ini terdapat ayat-ayat yang memiliki makna yang sama dan saling berkaitan, namun tidak berada di dalam satu surat atau juz (*chapter*) yang sama, sehingga sulit untuk mengaitkan informasi antara ayat-ayat di dalam Alquran bagi orang awam. Untuk dapat memahami Alquran, salah satunya dengan cara mengetahui ayat-ayat Alquran yang memiliki kesamaan dan keterkaitan makna, sehingga tidak memahami ayat-ayat tersebut secara parsial dan dapat memperoleh informasi yang lengkap mengenai makna yang terkandung dalam ayat-ayat tersebut. Metode seperti ini juga diterapkan dalam penafsiran ayat Alquran, yaitu menafsirkan suatu ayat dengan ayat yang lain, akan tetapi, untuk mengukur kesamaan semantik atau kesamaan makna antara dua buah dokumen teks masih sangat sulit dilakukan oleh komputer/sistem. Karena komputer/sistem tidak memiliki kemampuan intuisi seperti manusia yang dapat dengan tepat menentukan kesamaan antara dua buah dokumen teks. Oleh karena itu, diperlukan suatu sistem yang dapat memberikan informasi kesamaan antar ayat di dalam Alquran dengan ketepatan yang mirip dengan hasil intuisi manusia. Sehingga, dapat membantu umat Islam dalam mempelajari dan memahami Alquran.

Salah satu konsep yang dapat melakukan pengukuran kesamaan makna antara dua buah data teks adalah *Semantics Textual Similarity* (STS). STS dalam konteks teks pendek, merupakan salah satu penerapan teknologi penggalian data teks dan *Natural Language Processing* (NLP). Identifikasi terhadap teks pendek merupakan permasalahan penelitian yang penting untuk diaplikasikan dalam bidang NLP (*Machine Translation, Text Summarization, Question Answering, Short Answer Scoring, Information Retrieval*) [2]. Konsep ini dapat diterapkan untuk mengukur kesamaan antar dua ayat Alquran. Adapun untuk penelitian tentang STS ini terdapat berbagai metode yang digunakan, salah satunya adalah *word alignment*. *Word alignment* melakukan identifikasi pada kata-kata atau frasa yang terdapat di dalam duah buah kalimat untuk mengetahui kesamaan makna dua kalimat tersebut.

Pada tugas akhir ini akan diimplementasikan STS dengan menggunakan pendekatan *word alignment* pada terjemahan ayat Alquran dalam bahasa Indonesia untuk mengukur kesamaan antara dua ayat di dalam Alquran. Pendekatan *word alignment* dipilih karena metode tersebut merupakan salah satu metode yang cukup banyak digunakan pada *unsupervised system* dan pada salah satu kompetisi di bidang STS yaitu SemEval, merupakan metode yang sederhana dan menunjukkan performansi yang paling baik dalam dalam salah satu *running di task* STS pada SemEval 2014 dan SemEval 2015 .

1.2. Perumusan Masalah

Berdasarkan latar belakang masalah yang diuraikan, maka permasalahan yang dapat dirumuskan adalah sebagai berikut :

1. Bagaimana implementasi dari *Semantics Textual Similarity (STS)* dengan menggunakan pendekatan *word alignment* pada potongan ayat Alquran terjemahan bahasa Indonesia?
2. Bagaimana membuat *gold standard* untuk set data potongan ayat Alquran terjemahan bahasa Indonesia?
3. Bagaimana performansi sistem pada implementasi pendekatan *word alignment* pada set data pasangan potongan ayat Alquran terjemahan bahasa Indonesia?
4. Bagaimana pengaruh fitur yang digunakan dalam implementasi pendekatan *word alignment* pada set data pasangan potongan ayat Alquran terjemahan bahasa Indonesia?

1.3. Tujuan

Berdasarkan perumusan masalah yang ditujukan, tujuan tugas akhir ini adalah sebagai berikut:

1. Mengimplementasikan *Semantics Textual Similarity (STS)* pada potongan ayat Alquran terjemahan bahasa Indonesia dengan menggunakan pendekatan *word alignment*.
2. Membuat *gold standard* untuk set data pasangan potongan ayat Alquran terjemahan bahasa Indonesia yang digunakan dalam penelitian ini.
3. Menganalisis performansi sistem pada implementasi pendekatan *word alignment* pada set data pasangan potongan ayat Alquran terjemahan bahasa Indonesia .
4. Menganalisis pengaruh penggunaan fitur yang digunakan dalam pendekatan *word alignment* pada set data pasangan potongan ayat Alquran terjemahan bahasa Indonesia.

1.4. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data pasangan ayat yang digunakan adalah pasangan potongan ayat alquran terjemahan bahasa Indonesia versi Kementerian Agama Republik Indonesia.
2. Skala nilai gold standard yang digunakan adalah 0 sampai dengan 5.
3. Fitur pendekatan *word alignment* yang digunakan adalah fitur yang dapat diimplementasikan pada data berbahasa Indonesia.

1.5. Metodologi Penyelesaian Masalah

Adapun metode penyelesaian masalah dalam melakukan tugas akhir ini adalah sebagai berikut :

- a. Identifikasi masalah
Metode yang digunakan dalam melakukan identifikasi masalah adalah dengan melakukan studi literatur. Studi literatur dilakukan untuk mencari informasi seputar topik yang diteliti melalui hasil penelitian yang sudah dilakukan sebelumnya baik berupa jurnal penelitian, buku dan referensi ilmiah lainnya. Bagian-bagian penting yang dipelajari pada tahap ini yaitu, mulai dari konsep, data yang digunakan, permasalahan pada penelitian, metode penyelesaian, analisis dan hasil penelitian.
- b. Pengumpulan Data
Pada tahapan ini dilakukan pengumpulan data yang akan diperlukan untuk melakukan penelitian. Data yang diperlukan dalam penelitian ini berupa data pasangan potongan ayat Alquran, data *gold standard* dari pasangan potongan ayat tersebut dan juga data *knowledge base* yang akan diperlukan dalam implementasi metode *word alignment*. Pengumpulan data dilakukan dalam tiga bagian, yaitu pengumpulan data pasangan potongan ayat Alquran, pembuatan set data *gold standard* dan set data *knowledge base* berupa data pasangan parfrase teks Alquran.
- c. Pemodelan dan Perancangan
Pemodelan dan Perancangan sistem merupakan tahapan untuk memodelkan sistem yang akan dibuat, dan melakukan perancangan sistem yang akan dibangun.
- d. Implementasi
Pada tahapan implementasi dilakukan anotasi data dan *coding* program menggunakan algoritma dengan pendekatan *word alignment*, berdasarkan pemodelan dan perancangan sistem yang telah dibuat.
- e. Pengujian dan Analisis
Tahapan pengujian dilakukan setelah implementasi selesai dilaksanakan. Acuan yang digunakan dalam pengujian ini adalah beberapa parameter yang memiliki keterkaitan dalam menjawab tujuan penelitian. Kemudian, dilakukan analisis untuk mencari tahu penyebab dari hasil yang telah didapatkan pada proses pengujian.
- f. Pembuatan Laporan
Tahapan ini merupakan tahapan akhir dalam penelitian ini, dimana hasil dari penelitian didokumentasikan ke dalam sebuah laporan.

1.6. Sistematika Penulisan

Sistematika penulisan dari Tugas Akhir “ **Pengukuran Kesamaan Semantik pada Potongan Ayat Alquran dengan Pendekatan *Word Alignment*** ” adalah sebagai berikut:

1. Pendahuluan

Bab 1 menjelaskan tentang latar belakang masalah yang menjadi dasar penelitian, identifikasi terhadap rumusan masalah dan tujuan penelitian serta memberikan batasan masalah pada penelitian yang dilakukan. Selain itu, metodologi penyelesaian masalah yang menjelaskan tentang mekanisme penelitian yang dilakukan dideskripsikan secara rinci bersamaan dengan sistematika penulisan pada bab ini.

2. Dasar Teori

Bab 2 menjelaskan tentang teori-teori pendukung dan penelitian-penelitian terkait yang telah dilakukan sebelumnya yang akan digunakan sebagai landasan dalam pengembangan sistem. Istilah-istilah umum, metode yang terkait dengan penelitian, dan juga proses perhitungan kesamaan semantik tercantum didalam bab ini.

3. Perancangan Sistem

Bab 3 menjelaskan tentang perancangan sistem yang akan dibangun dan alur proses sistem bekerja termasuk proses pembuatan dan evaluasi set data yang akan dipakai pada sistem. Perancangan sistem disusun berdasarkan dasar teori yang telah diuraikan pada Bab 2. Perancangan sistem mendefinisikan setiap proses dari penelitian ini dalam melakukan pembuatan data dan perhitungan kesamaan semantik secara mendetail.

4. Pengujian dan Analisis

Bab 4 menjelaskan tentang set data yang digunakan dalam sistem, pengujian dan analisis terhadap sistem yang telah diimplementasikan. Selain itu, pada bab ini dilakukan analisis terhadap faktor -faktor yang mempengaruhi pengukuran kesamaan semantik.

5. Kesimpulan dan Saran

Bab 5 memuat kesimpulan dari hasil penelitian yang dilakukan, saran terkait dengan hasil penelitian yang telah dianalisis terlebih dahulu, dan rencana kedepan dari penelitian tersebut.