

BAB I

PENDAHULUAN

1.1. Latar Belakang

Ketersediaan perangkat beserta koneksi untuk mengakses jaringan internet pada zaman ini semakin meningkat. Dengan begitu kita dapat mencari informasi tanpa terikat waktu dan tempat. Banyak sekali orang mencari maupun bertukar informasi melalui forum-forum yang berada di internet, baik itu informasi untuk bisnis, tempat rekreasi, entertainment dan lain sebagainya. Salah satunya adalah Qatar Living Website Forum yang menyediakan informasi seputar tempat yang berada di Qatar. Setiap jawaban yang diberikan orang di forum tentunya berbeda-beda dan terdapat kemungkinan jawaban tersebut tidak mempunyai keterkaitan dengan pertanyaan yang diberikan. Oleh karena itu, dibutuhkan suatu system yang dapat membantu merangking istilah paling mendekati dengan pertanyaan yang diajukan di dalam Komunitas Tanya-Jawab. Salah satu metode yang dapat membantu memecahkan masalah ini adalah *Term Frequency* dan *Similarity Measure Features* dimana *Term Frequency* ini dapat menghitung *score* kalimat jawaban dari banyaknya term yang ada dari setiap jawaban yang diberikan oleh *responden* sedangkan *Similarity Measure Features* yang dibagi menjadi *Semantic Similarity* dan *Jaccard Similarity* dapat menghitung kesimilaritasan antar kalimat dari segi makna dan konten kalimat jawaban dengan pertanyaan tersebut.

Urutan proses yang dilakukan dalam penelitian ini adalah preprocessing data, pengklasifikasian jawaban, pembobotan jawaban dan yang terakhir adalah perankingan pada setiap jawaban. Pada pembobotan jawaban akan di terapkan *Term Frequency* dan *Similarity Measure Features*. Kegunaan dari *Term Frequency* ini untuk menghitung kemunculan term atau kata yang muncul pada suatu dokumen sehingga jumlah term jawaban telah dilihat berdasarkan pertanyaan yang berkaitan, sedangkan *Similarity Measure Features* untuk menghitung seberapa besar kemiripan atau keterkaitan antara pertanyaan dan jawaban yang telah diberikan.

1.2. Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana sistem mengetahui kriteria jawaban terbaik untuk pertanyaan yang terkait?
2. Bagaimana sistem dapat memasangkan jawaban dengan pertanyaan yang berhubungan?

1.3. Batasan Masalah

Batasan masalah yang diberikan dalam penelitian ini adalah sebagai berikut :

1. Dataset untuk penelitian ini diberikan oleh *Qatar Living Website Forum*.
2. Jawaban akan dikategorikan menjadi tiga kategori yaitu *good, potentially useful, dan bad*.
3. Jawaban akan dirangkingkan berdasarkan jawaban dari user pada *Qatar Living Website Forum*.
4. Proses *preprocessing* data dilakukan dengan *tools* yang sudah tersedia.

1.4. Tujuan

Tujuan dari penelitian tugas akhir ini adalah:

1. Mengimplementasikan metode *Term Frequency* dan *Similarity Measure Features* dalam menghitung bobot jawaban dan tingkat keakurasian similaritas data.
2. Merancang suatu sistem yang dapat merangking jawaban berdasarkan bobot dari setiap jawaban.

1.5. Metodologi Penyelesaian Masalah

Adapun tahapan metode yang digunakan pada penelitian ini adalah sebagai berikut :

1. Studi literatur

Mempelajari data set yang ada beserta berbagai referensi yang berhubungan dengan Answer Selection untuk *Questioning and Answering* terutama metode *Term Frequency* beserta *Similarity Measure Features*.

2. Pengumpulan dan Pengolahan Data

Proses ini merupakan tahapan pengumpulan studi kasus dan requirement yang akan digunakan untuk pengimplementasian aplikasi yang akan dibuat.

3. Desain Sistem

Proses ini merupakan perancangan dan pendefinisian masalah serta solusi yang akan diimplementasikan ke dalam system

4. Implementasi Sistem

Proses pengimplementasian suatu aplikasi yang dapat menghitung bobot setiap jawaban yang ada di dataset sehingga dapat memudahkan dalam menentukan jawaban yang terbaik dari semua jawaban yang ada.

5. Pengajuan aplikasi

Proses pengujian terhadap aplikasi yang telah dirancang juga kelayakan aplikasi tersebut.

6. Analisa hasil pengujian

Proses ini merupakan proses evaluasi terhadap hasil sampel jawaban yang diujikan.

7. Dokumentasi

Tahapan ini merupakan proses penyusunan laporan dari setiap kegiatan yang dilakukan.

1.6. Sistematika Penulisan

Penulisan Tugas Akhir ini dibagi menjadi beberapa bab berdasarkan poin-poin utama, yaitu :

1. Bab 1 : Pendahuluan
Bab Pendahuluan berisi landasan-landasan penelitian yang dilakukan, yaitu latar belakang, rumusan masalah, tujuan, batasan masalah, dan metodologi penyelesaian masalah.
2. Bab 2 : Kajian Pustaka
Bab Kajian Pustaka berisi kumpulan teori yang digunakan dalam penelitian ini, yaitu teori mengenai *Natural Language Processing*, *Text Preprocessing*, dan *Similarity Measure Features*.
3. Bab 3 : Perancangan Sistem
Bab Perancangan Sistem berisi langkah-langkah perancangan model, *Pairing Question and Answer*, *Preprocessing Data*, *Features Extraction*, dan Perankingan Jawaban.
4. Bab 4 : Pengujian dan Analisis
Bab Pengujian dan Analisis ini berisi hasil pengujian performansi *Similarity Measure Features* yang berupa tingkat ke akuratan jawaban terhadap pertanyaan yang di berikan.
5. Bab 5 : Kesimpulan dan Saran
Bab Kesimpulan dan Saran berisi kesimpulan hasil pengujian dan analisis, serta saran untuk pengembangan penelitian selanjutnya.