BAB I
Pendahuluan

1.1 Latar Belakang

Perkembangan teknologi informasi yang begitu cepat semakin memudahkan setiap orang dalam memperoleh informasi terhadap suatu hal yang baru, salah satunya ketika memasuki lingkungan yang baru, dan lain-lain. Hal tersebut membuat setiap orang dapat memahami secara umum terkait lingkungan yang akan dimasuki salah satunya dengan cara membaca *Frequently Asked Question* (FAQ) yang disediakan pihak penyedia tentang pertanyaan yang sudah pernah ditanyakan sebelumnya dan telah dijawab. Seiring berjalannya waktu, FAQ dari suatu produk akan semakin banyak sehingga akan menghabiskan waktu lebih banyak dalam membaca halaman FAQ dari produk tersebut, dan tidak jarang harus membaca setiap pertanyaan yang ada walaupun pertanyaan tersebut bukan pertanyaan yang kita butuhkan saat itu.

Oleh karena itu, dibutuhkan adanya suatu sistem *question answering* yang mampu secara otomatis menjawab pertanyaan yang dibutuhkan user [1] berdasarkan FAQ yang telah ada agar dapat memangkas waktu pencarian, salah satunya teknik *question answering* dengan menggunakan o*verall* s*imilarity*. Pada penelitian sebelumnya [2], dijelaskan bahwa *overall similarity* yakni menghitung kecocokan pertanyaan yang diajukan user dengan pertanyaan yang telah ada di FAQ sebelumnya. SemEval 2016 Task 3 merupakan salah satu kegiatan yang mengadakan kompetisi *community question answering* terutama untuk kasus *question-question similarity* dengan *mean average precision* (MAP) sebagai satuan evaluasi resminya [3].

Overall similarity merupakan hasil kombinasi nilai kecocokan dari dua pendekatan, yakni *semantic similarity* [4] dan *statistic similarity*. *Question answering* ini memiliki 3 tahapan pemrosesan utama, yaitu: (1) *Preprocessing,* (2) *Similarity measurement*, dan (3) *Answer Generation*. Adapun proses *preprocessing* dilakukan untuk mendapatkan data yang bebas dari *noisy text* lengkap dengan kelas katanya sehingga memudahkan penggunaan data dalam proses pembangunan sistem selanjutnya dan dapat digunakan untuk menambah nilai akurasi. Proses *similarity measurement* bertujuan untuk mengukur tingkat kecocokan dari pertanyaan yang diajukan *user* dengan pertanyaan yang ada pada *dataset* menggunakan kombinasi dari *semantic similarity* dengan menggunakan dan *statistic similarity* dengan istilah *overall similarity*. Langkah terakhirnya yaitu meng-*generate* jawaban dari pertanyaan yang berada diperingkat pertama dengan asumsi jawaban tersebut dapat menjawab pertanyaan yang diajukan *user*.

1.2 Perumusan Masalah

Berdasarkan masalah yang telah diutarakan diatas, maka dapat ditarik beberapa rumusan masalah, yaitu :

1. Bagaimana impelementasi *question similarity* pada k*asus FAQ Answering* dengan s*imilarity measurement* menggunakan s*emantic similarity* dan s*tatistic similarity*?
2. Bagaimana *analisis* performa *question similarity* pada kasus *FAQ answering* dengan *similarity measurement* menggunakan *semantic similarity* dan *statistic similarity*?

Adapun batasan masalah pada Tugas Akhir ini adalah sebagai berikut:

1. *Dataset* yang digunakan diperoleh dari *Qatar Living Forum* 2016 Task 3 dengan format .xml

1.3 Tujuan

Tujuan dari penelitian Tugas Akhir ini adalah sebagai berikut :

1. Mampu memahami impelementasi *question similarity* pada k*asus FAQ answering* dengan s*imilarity measurement* menggunakan s*emantic similarity* dan s*tatistic similarity*
2. Mampu menganalisis *question similarity* pada kasus *FAQ answering* dengan *similarity measurement* menggunakan *semantic similarity* dan *statistic similarity*

1.4 Metodologi Penyelesaisan Masalah

Metodologi berupa tahapan untuk menyelesaikan tugas akhir:

1. Kajian Pustaka

   Tahap ini digunakan untuk pencarian dan pengumpulan berbagai informasi berdasarkan buku maupun jurnal ilmiah untuk pembangunan sistem, baik itu tentang *semantic* dan *statistic similarity* serta pemodelannya maupun pendalaman materi tentang tahapan-tahapan dan metode yang ada di masing-masing pendekatan serta materi tentang text mining.
2. Pengumpulan dan analisis data

   Dataset yang digunakan berupa daftar pertanyaan dan jawaban terhadap suatu informasi yabg diperoleh dari *Qatar Living Forum* dengan format .xml.
3. Analisis dan Perancangan Sistem

   Pada tahap ini dapat digambarkan menggunakan *flowchart* untuk menggambarkan setiap proses yang hendak dijalankan dalam pembangunan sistem sehingga memudahkan untuk mengetahui alur dalam sistem..
4. Implementasi Model

Tahap ini dilakukan untuk merealisasikan tahap sebelumnya yaitu pemodelan. Model akan dikembangkan menjadi suatu perangkat lunak lengkap dengan fungsionalitas yang telah direncanakan.

5. Analisis dan Pengujian

   Melakukan analisis terhadap penggunaan *Semantic Similarity* dan *Statistic Similarity* untuk *Similarity Measurement*, analisis linear kombinasi dalam penentuan *similar question.*

6. Pembuatan Laporan Tugas Akhir

   Pembuatan laporan tugas akhir bertujuan untuk mendokumentasikan hasil dari tahapan penelitian yang telah dilakukan, disertai dengan lampiran yang mendukung penelitian Tugas Akhir ini.

1.5 Sistematika Penulisan

Dalam penulisan Tugas Akhir ini akan dibagi menjadi beberapa bab yang meliputi hal-hal sebagai berikut:

a. BAB 1 : PENDAHULUAN

   Berisi latar belakang, perumusan dan batasan masalah, tujuan, metodologi penyeleseaian yang digunakan dalam penelitian tugas akhir, dan sistematika laporan tugas akhir ini.

b. BAB 2 : LANDASAN TEORI

   Bab ini memuat tentang teori-teori yang mendukung dalam perancangan sistem pengenalan angka tulisan tangan yang dibuat. Teori dasar yang dibahas antara lain tentang *Natural Languange Processing, Question Answering, Penelitian Terkait, Text Preprocessing, Statistic Similarity, Semantic Similarity,dan metode evaluasi.*

c. BAB 3 : PERANCANGAN SISTEM

   Bab ini membahas tentang perencanaan sistem yang akan dilakukan untuk membuat aplikasi ” Question Similarity pada Kasus FAQ Answering dengan Similarity Measurement Menggunakan Semantic Similarity dan Statistic Similarity.”

d. BAB 4 : PENGUJIAN DAN ANALISIS SISTEM

   Bab ini menjabarkan secara singkat mengenai dataset yang digunakan, baseline, evaluation metric yang digunakan dan hasil pengujian. Analisis diterapkan berdasarkan hasil yang diperoleh.

e. BAB 5 : KESIMPULAN DAN SARAN

   Bab terakhir ini berisi kesimpulan terhadap pengaruh kombinasi linear antara nilai statistic dan semantic similarity terhadap nilai-nilai evaluasi yang dihasilkan. Saran sebagai masukan penelitian lebih lanjut juga disertakan pada bab ini.