

Bab I

Pendahuluan

1.1 Latar Belakang

Keanekaragaman pertanyaan yang muncul dalam sebuah forum diskusi akan menghasilkan berbagai jawaban untuk setiap pertanyaan yang diajukan. Beberapa forum diskusi memungkinkan adanya jawaban yang tidak relevan dengan pertanyaan yang diunggah sehingga tidak sesuai dengan keinginan pengunggah dan memenuhi halaman pertanyaan tersebut dengan kata lain jawaban tersebut dikategorikan sebagai *spam* [5][4]. Ketika jawaban yang tidak relevan terlebih dahulu diunggah sebelum jawaban relevan maka pengunggah pertanyaan akan membutuhkan waktu yang tidak singkat untuk mencari jawaban relevan dari pertanyaannya. Contoh pertanyaan “Apakah ada ikan didekat Pantai Seraph?” akan menghasilkan jawaban yang banyak. Beberapa jawaban akan bersifat relevan seperti “Ada, dibagian barat mendekati ujung pantai” ataupun “Ditengah pantai sekitar 20 meter arah barat” dan beberapa jawaban akan bersifat tidak relevan seperti “Tidak mungkin ada ikan di dekat pantai”.

Masalah yang diuraikan diatas banyak dijumpai pada forum diskusi dan dapat menambah waktu pengunggah untuk melakukan pencarian jawaban yang relevan dengan pertanyaannya. Community Question Answering merupakan cabang bidang dari Question Answering yang sekarang ini digeluti pada SemEval 2016. Bidang ini membantu pengguna untuk mendapatkan jawaban yang diharapkan oleh pengguna. Forum diskusi membutuhkan sebuah sistem yang dapat *me-ranking* jawaban pada setiap halaman forum tersebut secara otomatis agar waktu yang digunakan untuk pencarian jawaban berkurang dan Community Question Answering menyediakan solusi dari permasalahan tersebut. Pemberian *ranking* pada setiap jawaban tersebut dapat dilakukan dengan bantuan Community Question Answering : *Question - Comment Similarity*(QCS), *Question - Question Similarity*(QQS), *Question - External Comment Similarity*(QECS) [9].

Pada tugas akhir ini, akan dibuat sistem yang akan memberikan *ranking*

pada setiap jawaban yang diunggah pada halaman pertanyaan. Sistem ini akan mengimplementasikan QCS pada Community Question Answering yang secara umum, algoritma sistem akan melakukan *Part of Speech Tagging*(PoS-T) untuk setiap kata pada pertanyaan. Kemudian dengan bantuan *Regular Expression*(RegEx), setiap jawaban akan dicocokkan dengan setiap kata pada pertanyaan. PoS-T juga dilakukan pada jawaban-jawaban yang cocok dan akan dicocokkan dengan hasil PoS-T kalimat pertanyaan. Jawaban yang tersisa dan relevan akan di-*ranking* sesuai nilai yang diberikan melalui proses yang diuraikan di atas. Sistem ini menggunakan QCS dan selanjutnya akan disebut S-QCS.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, berikut perumusan masalah :

1. Apa saja PoS Tag yang digunakan dalam prosedur PoS-T ?
2. Bagaimana cara mengimplementasikan PoS-T pada pertanyaan ?
3. Bagaimana mengimplementasikan RegEx yang digunakan pada setiap jawaban?
4. Bagaimana cara memberi nilai pada setiap jawaban yang diunggah pada halaman pertanyaan ?

1.3 Tujuan

Tujuan utama dari tugas akhir ini yaitu pembuatan sistem yang memiliki algoritma untuk melakukan PoS-T, RegEx dan pemberian *ranking* untuk setiap jawaban pada halaman pertanyaan. Pemberian *ranking* ini akan digunakan untuk mensortir jawaban sehingga pengunggah pertanyaan dapat mendapatkan jawaban yang paling relevan tanpa harus melihat jawaban-jawaban yang tidak relevan. Tujuan lain tugas akhir ini yakni menentukan penggunaan PoS-T yang baik dan aturan RegEx yang baik pula.

1.4 Batasan Masalah

Adapun batasan masalah pada tugas akhir ini, yaitu :

1. Data diambil dari forum *Community Question Answering*(CQA) SemEval-2016 Task 3.
2. Pembuatan sistem dengan bantuan *NetBeans* dan bahasa pemrograman Java.
3. Beberapa algoritma dalam sistem merupakan algoritma yang sudah tersedia.