

1 Pendahuluan

1.1 Latar Belakang

Representasi sebuah argumen semantik dalam sebuah susunan kalimat merupakan hal yang penting dalam *text mining*, dalam penentuan siapa yang melakukan apa kepada siapa, kapan, dimana, mengapa, bagaimana dan lain sebagainya dalam sebuah text [5].

Proses identifikasi suatu argumen yang terdapat pada sebuah kalimat lewat predikat yang ada hingga mendapatkan role atau peran semantiknya dikenal sebagai *semantic role labelling* (SRL). *Semantic role labelling* atau yang lebih sering dikenal juga sebutan *shallow semantic parsing* telah dikembangkan dan diterapkan untuk mengekstraksi informasi pada *open domain* dan *biomedical subdomain*. Namun terdapat suatu masalah dimana terkadang hanya terdapat sedikit data yang teranotasi yang menyebabkan sistem SRL tidak dapat melakukan proses ekstraksi informasi dengan semestinya. Untuk mengatasi masalah tersebut diperlukan pengembangan secara lanjut sistem SRL yang sesuai dengan data tersebut. Dan dalam pembangunan sistem SRL dibutuhkan waktu yang lama serta biaya yang tidak sedikit.

Tugas akhir ini akan dilakukan sebuah metode untuk menyelesaikan masalah-masalah yang dipaparkan di atas yang disebut dengan adaptasi domain. Tujuannya adalah untuk mengadaptasi pengetahuan sebuah domain yang telah dibangun sebelumnya ke dalam domain lain yang masih memiliki sedikit data yang teranotasi. Dengan dilakukannya *transfer* pengetahuan yang dimiliki domain yang lebih kaya ditujukan untuk meningkatkan performa dari sistem SRL yang telah ada. Dalam tugas akhir ini data yang digunakan sebagai data target adaptasi terjemahan Al Qur'an dalam bahasa Inggris yang akan dilabelkan secara manual. Dan korpus PropBank yang akan digunakan sebagai sumber data yang akan ditransferkan pengetahuannya. Serta algoritma adaptasi domain yang akan digunakan dalam tugas akhir adalah *transfer self-training*.

1.2 Perumusan Masalah

Adapun rumusan permasalahan yang dilakukan dalam perumusan tugas akhir ini adalah sebagai berikut :

1. Bagaimana hasil dari klasifikasi adaptasi domain menggunakan algoritma learning *Transfer Self-Training* ?
2. Bagaimana pengaruh akurasi *BaselineFeature* dengan penambahan *Noun Head of PP*, *Syntactic Frame*, *Argument Order*, *Constituent*, *First* dan *Last Word/POS in Constituent* ?

1.3 Tujuan

Adapun tujuan yang ingin dicapai dalam pengerjaan tugas akhir ini adalah sebagai berikut :

1. Menganalisis hasil dari klasifikasi adaptasi domain dengan algoritma learning *Transfer Self-Training*.

2. Menganalisis hasil kinerja klasifikasi yang dihasilkan berdasarkan *Baseline Feature* dengan penambahan *Noun Head of PP*, *Syntactic Frame*, *Argument Order*, *Constituent*, *First* dan *Last Word/POS in Constituent*.

1.4 Batasan Masalah

Tugas akhir ini memiliki beberapa batasan masalah untuk membatasi ruang penelitian yang ada. Batasan masalah dari tugas akhir ini adalah :

1. Fitur yang akan digunakan dalam tugas akhir ini adalah fitur yang telah dikembangkan dalam penelitian sebelumnya yang dilakukan oleh Daniel *et al* dan tidak termasuk dalam *Baseline Feature*.
2. Pemilihan algoritma yang akan digunakan dalam tugas akhir ini adalah algoritma yang sebelumnya pernah digunakan dalam pengujian *Semantic Role Labelling* yaitu *Transfer Self-Training*.
3. *Dataset* yang akan digunakan dalam tugas akhir ini adalah berupa korpus yang berasal dari PropBank dan terjemahan Al-Qur'an dalam bahasa Inggris.
4. *Classifier* yang digunakan dalam tugas akhir ini adalah SVM lib yang diambil dari api weka.
5. Kernel SVM lib yang digunakan adalah kernel linear $u \cdot v$ yang menunjukkan hasil paling tinggi diantara kernel lainnya.

1.5 Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah dalam tugas akhir ini adalah sebagai berikut :

1. Studi Literatur
Tahap ini merupakan tahapan dilakukannya pengumpulan teori dan data yang berkaitan dengan metode yang akan digunakan dalam pengerjaan tugas akhir ini sehingga mendapatkan gambaran yang jelas tentang implementasi nantinya. Tahap studi literatur dilakukan dari awal hingga akhir pengerjaan Tugas Akhir.
2. Pengumpulan Data yang Dibutuhkan
Tahap ini dilakukan pengumpulan *dataset* berupa data teks, yang nantinya akan digunakan dalam pengujian implementasi klasifikasi argumen semantik. *Dataset* diperoleh dari PropBank, sebuah korpus semantik yang telah tersedia untuk bahasa inggris dan terjemahan Al-Qur'an dalam bahasa inggris.
3. Perancangan dan Implementasi Sistem
Pada tahap ini dilakukan perancangan serta pengimplementasian *tools* yang digunakan untuk melakukan *semantic role labelling* dan klasifikasi argumen semantik.
4. Pengujian dan Analisis
Dalam tahap ini nantinya akan dilakukan pengujian dan analisis hasil dari algoritma dan set fitur yang telah ditentukan, dalam hal ini analisis akan dititikberatkan pada hasil kecepatan proses learning dan keakuratan hasil pelabelan
5. Penyusunan Laporan Tugas Akhir
Penyusunan laporan merupakan tahap pendokumentasian dari apa yang telah dikerjakan pada tugas kahir ini, lengkap dengan hasil pengujian serta

analisis. Tidak lupa juga dengan penyajian hasil kesimpulan dan saran yang diperoleh dari hasil tugas akhir ini.

1.6 Sistematika Penulisan

Adapun sistematika penulisan Tugas Akhir ini adalah sebagai berikut:

1. Bab 1 Pendahuluan
Pada Bab 1 diuraikan isi dan gambaran umum Tugas Akhir secara menyeluruh yang meliputi latar belakang, perumusan masalah, batasan masalah dan tujuan dari penyelesaian masalah yang diterapkan serta sistematika penulisan.
2. Bab 2 Landasan Teori
Bab 2 memaparkan dasar-dasar teori yang berkaitan dengan semua aspek yang digunakan dalam penelitian Tugas Akhir ini seperti dasar *text mining*, korpus PropBank, terjemahan Al Qur'an dalam bahasa Inggris, penjelasan mengenai fitur-fitur yang digunakan, adaptasi domain, hingga algoritma *classifier* yang diimplementasikan.
3. Bab 3 Perancangan Sistem
Dipaparkan gambaran umum sistem yang akan dibangun, mulai dari diagram alur pengerjaan Tugas Akhir, rancangan sistem berupa perancangan data, *preprocessing*, *feature extraction*, adaptasi domain, hingga klasifikasi, juga dilengkapi dengan deskripsi kebutuhan sistem.
4. Bab 4 Pengujian dan Analisis
Berisikan pembahasan tentang hasil pengujian yang diperoleh berdasarkan implementasi dari bab sebelumnya. Hasil pengujian yang telah dilakukan selanjutnya dianalisis, kemudian dari hasil analisis yang diperoleh tersebut selanjutnya digunakan sebagai dasar untuk pengambilan kesimpulan.
5. Bab 5 Kesimpulan dan Saran
Berisikan hasil kesimpulan yang diperoleh dari hasil analisis yang telah dijelaskan pada bab sebelumnya, juga memaparkan saran berkaitan dengan pengembangan penelitian selanjutnya berdasarkan penelitian yang sudah dilakukan pada tugas akhir ini.