

1. Pendahuluan

Latar Belakang

Media sosial merupakan media komunikasi yang memungkinkan penggunanya untuk saling berinteraksi secara daring untuk berbagai tujuan yakni seperti bekerja, belajar, bermain dan juga menghabiskan waktu bersama teman dan juga keluarga secara publik[1]. Twitter merupakan salah satu media sosial yang paling populer untuk beberapa tahun terakhir. Hal ini ditandai dengan jumlah pengguna aktifnya yang mencapai 313 juta pengguna dan lebih dari 500 juta postingan per harinya[2]. Terdapat banyak topik informasi yang tersebar melalui postingan di Twitter, salah satu topik yang bisa dicari adalah informasi mengenai beasiswa[3]. Namun pada penerapannya, terdapat beberapa kendala pada pencarian informasi yang dimana salah satu penyebabnya adalah karena tweet dari pengguna yang memiliki format yang beragam sehingga menyulitkan pengguna lainnya untuk berhasil mendapatkan informasi beasiswa yang ingin dicari[4]. Akan tetapi, jika tweet dari pengguna dapat diolah dengan baik sehingga menghasilkan format tertentu, maka informasi mengenai beasiswa akan bisa didapatkan. Hal ini merupakan salah satu permasalahan pada bidang Natural Language Processing (NLP). Terdapat beberapa task yang dapat dikerjakan dalam bidang NLP, dan untuk task berupa mengolah tweet agar dapat menjadi informasi yang berguna merupakan salah satu task NLP yakni Information Extraction (IE). Information Extraction merupakan proses untuk mengekstrak informasi yang terstruktur dari teks yang tidak terstruktur[5].

Pada penelitian ini penulis menggunakan dataset berupa tweet berbahasa Inggris dari pengguna Twitter yang berisi informasi mengenai beasiswa S3. Penulis melakukan task IE dengan menggunakan metode Maximum Entropy Markov Models (MEMM). Metode ini digunakan karena dapat menggabungkan konteks teks terhadap karakteristik informasi dan karakteristik informasi yang terkandung dalam sebuah teks ke dalam model markov sehingga dapat meningkatkan performansi beberapa task NLP yang mana salah satunya adalah IE[6]. Pada penelitian di tahun 2009, Rong Li menyimpulkan bahwa dengan memasukkan metode Maximum Entropy ke dalam model HMM, permasalahan representasi pengetahuan dapat diselesaikan dan menambah pengetahuan bahasa yang baru kedalam model[7].

Topik dan Batasannya

Topik dan batasan masalah yang digunakan dalam penelitian tugas akhir ini mencakup bagaimana pengimplementasian metode MEMM untuk ekstraksi informasi beasiswa pada Twitter dan performansi metode MEMM terhadap ekstraksi informasi pada Twitter. Dataset yang digunakan berupa kumpulan tweet berbahasa Inggris dan berisikan informasi mengenai beasiswa S3. Dataset menggunakan format skema BIO yang memiliki 7 label, yakni:

- Field : Bidang studi dari beasiswa
- Link : Tautan dari situs web mengenai informasi beasiswa
- Lab : Lembaga yang menawarkan beasiswa
- Country : Negara dari tempat universitas beasiswa
- Topic : Topik yang berkaitan dengan beasiswa
- Deadline : Tenggat waktu dari beasiswa
- Univ : Nama universitas dari beasiswa

Tujuan

Tujuan yang ingin dicapai dalam penelitian tugas akhir ini adalah untuk mengetahui hasil dari pengimplementasian metode MEMM untuk ekstraksi informasi beasiswa pada Twitter serta performansi dari metode tersebut dalam menyelesaikan task ekstraksi informasi.

Organisasi Tulisan

Pada bab 2 dibahas studi terkait penelitian yang dilakukan, bab 3 dibahas sistem yang dibangun, bab 4 dibahas evaluasi dari model, dan bab 5 dibahas kesimpulan dari penelitian yang dilakukan.