

BAB 1 PENDAHULUAN

Bab ini berisi tentang latar belakang (subbab 1.1), tujuan penelitian (subbab 1.2), perumusan dan pembatasan masalah (subbab 1.3), metodologi penelitian (subbab 1.4), serta penjelasan mengenai sistematika penulisan laporan (subbab 1.5).

1.1 Latar Belakang

Bahasa merupakan alat komunikasi yang digunakan oleh seseorang untuk menyampaikan ide, gagasan, pengalaman, dan sebagainya kepada orang lain. Bahasa memiliki peranan yang besar dalam kehidupan manusia. Terdapat berbagai ragam macam bahasa yang ada di dunia, salah satunya adalah Bahasa Inggris dan Bahasa Indonesia.

Saat ini dengan sarana salah satu teknologi informasi (TI) yang disebut internet, informasi dan ilmu pengetahuan bisa dengan mudah didapatkan oleh semua orang di seluruh dunia. Akan tetapi, bagi sebagian orang, keragaman dan perbedaan bahasa yang ada seringkali menjadi kendala dalam berinteraksi dan berkomunikasi. Masyarakat Indonesia sekarang ini masih tergolong lemah dalam penguasaan Bahasa Inggris, sehingga faktor ini menjadi faktor yang sangat penting dalam terhambatnya penyebaran informasi. Banyak orang seperti: ilmuwan, insinyur, pengusaha harus membaca banyak dokumen dan berkomunikasi dalam Bahasa Inggris. Dengan adanya kemajuan teknologi yang sudah merambah semua bidang, saat ini sedang dikembangkan mesin penerjemah untuk mengatasi kendala penerjemahan bahasa tersebut.

Mesin penerjemah merupakan alat penerjemah otomatis pada sebuah teks dari satu bahasa ke bahasa lainnya. Mesin penerjemah juga merupakan salah satu aplikasi yang penting dalam bidang *Natural Language Processing* (NLP). Tujuan dari mesin penerjemah adalah dapat membuat orang – orang yang berasal dari berbagai budaya, yang memiliki bahasa yang berbeda, dapat berkomunikasi satu sama lain dengan mudah. Agar bisa melakukan proses penerjemahan, mesin

penerjemah harus mengetahui sinonim kata - kata dan frase – frase dari dua buah bahasa dan mempunyai pengetahuan semantik tentang kedua bahasa tersebut.

Salah satu caranya adalah dengan menggunakan dua ahli bahasa untuk memasukkan informasi bahasa yang diperlukan ke dalam mesin penerjemah secara manual. Akan tetapi cara ini sangatlah tidak efektif dan sangat mahal. Cara lainnya adalah dengan membiarkan mesin penerjemah mempelajari informasi bahasa secara otomatis dari pasangan dokumen yang paralel (dalam penelitian ini pasangan dokumen yang digunakan adalah dokumen Bahasa Inggris dan Bahasa Indonesia) dan menggunakan konsep penerjemahan secara statistik untuk menghasilkan terjemahan yang optimal. Mesin penerjemah yang menggunakan cara tersebut disebut sebagai mesin penerjemah statistik atau *statistical machine translation*. Akan tetapi, kualitas hasil terjemahan tersebut masih jauh dari sempurna dan memiliki nilai akurasi yang rendah. Diawali dari permasalahan ini, muncullah sebuah ide untuk melakukan percobaan agar dapat meningkatkan kualitas dan nilai akurasi hasil terjemahan dari mesin penerjemah statistik dengan menggunakan teknik restrukturisasi teks Bahasa Indonesia.

1.2 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan penulis pada tugas akhir ini adalah:

- Menyelidiki pengaruh restrukturisasi teks pada korpus paralel Bahasa Inggris dari mesin penerjemah statistik Inggris - Indonesia.
- Membandingkan nilai akurasi hasil terjemahan yang menggunakan strategi *word reordering* berdasarkan *Part Of Speech tagging* dengan hasil terjemahan tanpa *word reordering* dari proses penerjemahan Bahasa Inggris ke Bahasa Indonesia.
- Membandingkan nilai akurasi hasil terjemahan yang menggunakan strategi *phrase reordering* berdasarkan *parse tree* dengan hasil terjemahan tanpa *phrase reordering* dari proses penerjemahan Bahasa Inggris ke Bahasa Indonesia.
- Membandingkan nilai akurasi hasil terjemahan yang menggunakan strategi *word reordering* dan *phrase reordering* berdasarkan *Part Of Speech*

tagging dan *parse tree* dengan hasil terjemahan tanpa *word reordering* dan *phrase reordering* dari proses penerjemahan Bahasa Inggris ke Bahasa Indonesia.

1.3 Perumusan dan Pembatasan Masalah

Permasalahan pada tugas akhir ini adalah rendahnya kualitas dan nilai akurasi terjemahan dari mesin penerjemah statistik Bahasa Inggris ke Bahasa Indonesia. Permasalahan lain dari tugas akhir ini adalah bagaimana mendefinisikan aturan restrukturisasi teks pada kalimat Bahasa Inggris agar sesuai dengan sintaks Bahasa Indonesia dan menyelidiki apakah aturan restrukturisasi teks tersebut dapat meningkatkan kualitas penerjemahan dari mesin penerjemah statistik yang menggunakan model frase. Pembatasan masalah pada tugas akhir ini adalah penerjemahan hanya dilakukan satu arah dari Bahasa Inggris ke Bahasa Indonesia

1.4 Metode dan *Tools* Pengembangan

Penelitian diawali dengan melakukan studi pustaka mengenai *statistical machine translation*. Studi pustaka dilakukan dengan membaca referensi *paper* mengenai perkembangan dari *machine translation* dan mempelajari penelitian – penelitian yang telah dilakukan dalam *machine translation*, *word reordering*, dan *phrase reordering*. Setelah melakukan studi pustaka, penelitian dilanjutkan dengan mempelajari cara penggunaan *tools* yang akan digunakan untuk penelitian, yaitu: Moses, Stanford POS Tagger, Stanford Parser dan program Automatic Evaluation. Kemudian penelitian dilanjutkan dengan pembuatan aturan restrukturisasi teks untuk Bahasa Inggris dengan pola struktur Bahasa Indonesia. Setelah itu dirancang beberapa arsitektur eksperimen. Lalu dilakukan analisis pada hasil eksperimen tersebut.

1.5 Sistematika Penulisan

Laporan tugas akhir ini terdiri atas enam bab, dengan rincian sebagai berikut:

- BAB 1 Pendahuluan. Bab ini berisi tentang latar belakang (subbab 1.1), tujuan penelitian (subbab 1.2), perumusan dan pembatasan masalah (subbab 1.3), metodologi penelitian (subbab 1.4), serta penjelasan mengenai sistematika penulisan laporan (subbab 1.5).
- BAB 2 Landasan Teori. Bab ini menjelaskan tentang landasan teori yang digunakan dalam penelitian ini. Pembahasan dimulai dari sejarah *machine translation* (subbab 2.1), *statistical machine translation* (subbab 2.2), *automatic evaluation* (subbab 2.3), *word reordering* dan *phrase reordering* (subbab 2.4), POS *tagging* (subbab 2.5), dan *parsing* (subbab 2.6).
- Bab 3 Arsitektur dan Perancangan Sistem. Bab ini menjelaskan tentang arsitektur dan perancangan sistem pada saat pengerjaan tugas akhir. Pembahasan dimulai dari arsitektur sistem (subbab 3.1), korpus yang digunakan (subbab 3.2), *K-fold cross validation* (subbab 3.3), rancangan *word reordering* (subbab 3.4), dan rancangan *phrase reordering* (subbab 3.5). Subbab pada bab ini penulis susun dalam urutan sistematis, sesuai dengan alur yang penulis lalui selama masa tugas akhir.
- BAB 4 Implementasi. Bab ini menjelaskan tentang proses yang dilalui pada saat implementasi penelitian. Pembahasan dimulai dari persiapan awal dokumen (subbab 4.1), Stanford POS Tagger (subbab 4.2), *word reordering* (subbab 4.3), *cleaning* dokumen (subbab 4.4), Stanford Parser (subbab 4.5), *phrase reordering* (subbab 4.6), MOSES (subbab 4.7), dan *evaluation tools* (subbab 4.8). Subbab pada bab ini penulis susun dalam urutan sistematis, sesuai dengan alur yang penulis lalui selama masa tugas akhir.
- BAB 5 Hasil Implementasi dan Uji Coba. Bab ini menjelaskan tentang hasil implementasi dan uji coba yang dilakukan dalam penelitian. Pembahasan dimulai dari hasil dan analisis eksperimen arsitektur *baseline* (subbab 5.1), hasil dan analisis eksperimen arsitektur sistem POS *tagging* (subbab 5.2), hasil dan analisis eksperimen arsitektur sistem *parsing* (subbab 5.3), hasil dan analisis eksperimen arsitektur sistem *mixed* (subbab 5.4), rangkuman analisis eksperimen (subbab 5.5). Subbab pada

bab ini penulis susun dalam urutan sistematis, sesuai dengan alur yang penulis lalui selama masa tugas akhir.

- **BAB 6 Penutup.** Bab ini berisikan tentang kesimpulan yang didapatkan dari penelitian yang telah dilakukan (subbab 6.1), keterbatasan dan hambatan (subbab 6.2), serta saran penulis untuk pengembangan selanjutnya dari penelitian ini (subbab 6.3).

