

BAB 5

PENUTUP

Bab ini merupakan penutup dari laporan tugas akhir yang berisi kesimpulan dari percobaan yang dilakukan. Selain itu, pada subbab terakhir juga diberikan saran-saran untuk pengembangan lebih lanjut dalam penelitian penerjemahan dokumen bahasa Indonesia ke bahasa Inggris dengan menggunakan Mesin Penerjemah Statistik (MPS) berdasarkan frase.

5.1 Kesimpulan

Eksperimen penerjemahan teks bahasa Indonesia ke bahasa Inggris dilakukan dengan menggunakan 4 jenis korpus dwibahasa, yaitu korpus dwibahasa biasa (tanpa menggunakan perangkat bahasa apapun), korpus dwibahasa dengan POS Tag, korpus dwibahasa dengan lema dan korpus dwibahasa dengan POS Tag serta lema. Selain itu, eksperimen juga dilakukan dengan menggunakan 3 jenis model bahasa, yaitu model bahasa 1-gram, model bahasa 3-gram dan model bahasa 5-gram. Dari hasil eksperimen dan analisis yang telah dilakukan, penulis menarik sejumlah kesimpulan, antara lain:

1. Sebuah sistem MPS berdasarkan frase untuk bahasa Indonesia-Inggris telah berhasil dibuat.
2. Terdapat faktor-faktor yang dapat mempengaruhi hasil terjemahan Indonesia-Inggris, antara lain adalah penggunaan anotasi tambahan POS Tag dan identifikasi lema.
3. Dari 4 jenis korpus yang digunakan, nilai akurasi yang paling baik untuk semua kategori dicapai oleh korpus dwibahasa biasa dan dari ke-5 nilai akurasi pada masing-masing kategori tersebut, nilai akurasi yang paling baik dicapai oleh kategori novel pada jumlah korpus pelatihan 16000 dengan menggunakan model bahasa 5-gram yaitu 0,2696.
4. Pada korpus dwibahasa biasa, semakin besar jumlah korpus pelatihan yang digunakan, maka nilai akurasi yang dihasilkan cenderung semakin naik. Kemudian, semakin besar nilai n pada model bahasa n -gram yang

digunakan, maka nilai akurasi yang dihasilkan cenderung semakin naik juga.

5. Pada korpus dwibahasa yang menggunakan POS Tag, lema dan POS Tag serta lema, semakin besar jumlah korpus pelatihan yang digunakan, nilai akurasi yang dihasilkan ada yang naik dan turun. Kemudian, semakin besar nilai n pada model bahasa n -gram yang digunakan, nilai akurasi yang dihasilkan juga ada yang naik dan turun.
6. Penurunan nilai akurasi pada eksperimen dengan menggunakan POS Tag dan lema disebabkan karena beberapa hal, antara lain:
 - a. POS Tag untuk bahasa Indonesia yang digunakan masih belum sempurna serta adanya perbedaan jenis *tag-set* yang cukup banyak antara *tag-set* untuk bahasa Indonesia dan bahasa Inggris.
 - b. Adanya kekurangan di dalam melakukan identifikasi lema untuk bahasa Indonesia yaitu KBBI acuan yang digunakan untuk melakukan pengecekan lema belum mencakup semua kemungkinan kata berimbuhan yang ada dalam bahasa Indonesia. Sementara itu kekurangan yang terjadi di dalam melakukan identifikasi lema untuk bahasa Inggris adalah belum digunakannya acuan berupa kamus bahasa Inggris untuk melakukan pengecekan lema yang dihasilkan.

5.2 Saran

Untuk pengembangan sistem MPS berdasarkan frase bahasa Indonesia-bahasa Inggris yang lebih baik di masa mendatang, penulis memiliki beberapa saran sebagai berikut:

1. Jenis kategori korpus dwibahasa yang digunakan diperbanyak, seperti kategori ilmu pengetahuan alam, ilmu sosial, politik, hukum, budaya, hiburan dan lain sebagainya sehingga dapat lebih memperkaya isi korpus yang digunakan, terutama perbendaharaan kata yang digunakan.

2. Jumlah korpus dwibahasa yang digunakan diperbanyak, terutama untuk kategori artikel berita dan percakapan agar nilai akurasi yang dihasilkan bisa lebih tinggi.
3. POS Tag untuk bahasa Indonesia yang digunakan dikembangkan lebih lanjut dan disempurnakan lagi agar ketika digunakan kembali dalam penelitian, hasil terjemahan yang dihasilkan dapat memberikan nilai akurasi yang lebih tinggi. Selain itu, jenis *tag-set* untuk bahasa Indonesia yang digunakan bisa lebih disamakan dengan jenis *tag-set* bahasa Inggris. Penyamaan ini memang tidak akan bisa dilakukan untuk semua jenis *tag-set*, seperti halnya jenis kata kerja (*verb*) yang ada di bahasa Indonesia tidak bisa disamakan semuanya dengan jenis kata kerja yang ada di bahasa Inggris dimana jenis kata kerja yang ada di bahasa Inggris sangat kompleks dan beragam sementara jenis kata kerja yang ada di bahasa Indonesia sangatlah sederhana.
4. Identifikasi lema yang dilakukan bisa lebih dikembangkan dan disempurnakan. KBBI acuan yang digunakan dalam melakukan identifikasi lema untuk bahasa Indonesia perlu disempurnakan lagi, seperti menambahkan berbagai macam kata berimbuhan dan kata dasar yang belum tercakup. Dalam melakukan identifikasi lema bahasa Inggris, perlu ditambahkan penggunaan kamus acuan bahasa Inggris sehingga kata-kata yang telah dilakukan *stemming* dengan menggunakan *stemmer* (seperti Porter Stemmer) bisa dicek kebenarannya di kamus tersebut.
5. Perangkat bahasa tambahan yang digunakan bisa diperbanyak variasinya, seperti faktor-faktor morfologi dan sebagainya. Perangkat-perangkat bahasa yang digunakan tersebut harus mendukung untuk kedua bahasa, yaitu bahasa sumber dan bahasa target sehingga nilai akurasi yang dihasilkan bisa bagus.
6. Selain menggunakan model bahasa yang bervariasi, ada parameter-parameter lain yang dapat divariasikan untuk membandingkan nilai akurasi, diantaranya adalah panjang frase maksimum pada tabel frase yang

dihasilkan oleh sistem MPS berdasarkan frase ketika melakukan pelatihan. Dengan menggunakan berbagai parameter yang bisa divariasikan, maka perbandingan nilai akurasi bisa lebih akurat dan mantap.