

Abstrak

Penelitian ini berfokus dalam pembuatan mesin penerjemah statistik bahasa Minang – Indonesia serta melihat seberapa baik hasil terjemahan mesin. Sumber data *training* dan *test* berupa *corpus parallel* dan *corpus monolingual* berasal dari *Wikipedia* bahasa Minang dan *website* berita bahasa Indonesia. Semua eksperimen konfigurasi dilakukan dengan menggunakan data uji sebanyak 600 baris kalimat. Dua skenario pengujian dilakukan yaitu dengan menggunakan skenario yang berdasarkan *monolingual corpus* dan *parallel corpus*. Untuk melihat seberapa baiknya terjemahan akan dilihat dengan menggunakan pengujian otomatis *Bilingual Evaluation Understudy* (BLEU). Hasil pengujian pada 6 konfigurasi menunjukkan adanya peningkatan nilai akurasi mesin penerjemah setelah jumlah *corpus monolingual* dan *parallel* ditambahkan. Pada skenario pertama konfigurasi 3 dan 2 meningkat sebesar 3,6% konfigurasi 2 dan 1 meningkat sebesar 2,59%. Pada skenario dua konfigurasi 5 dan 4 meningkatkan sebesar 0,44% konfigurasi 4 dan 1 meningkat sebesar 0,06%. Hasil penelitian menunjukkan bahwa pada pengujian skenario pertama memiliki dampak yang signifikan dibandingkan dengan pengujian skenario kedua dalam segi terjemahan. Kurangnya sumber *corpus* menjadi masalah dalam membuat mesin penerjemah statistik.

Kata kunci: bahasa minang, bahasa Indonesia, mesin penerjemah statistik, bilingual evaluation understudy