

## Abstrak

Evaluasi secara otomatis untuk mengukur kualitas dari terjemahan mesin translasi dilakukan dengan menggunakan *automatic metric*. *Metric* didefinisikan sebagai *measurement* (ukuran). *Metric* untuk evaluasi keluaran terjemahan mesin merupakan ukuran kualitas keluaran terjemahan mesin tersebut. Oleh karena kualitas terjemahan bersifat subjektif maka tugas dari berbagai *metric* adalah untuk menetapkan *score* terhadap kualitas dalam berbagai cara sehingga *score* memiliki hubungan dengan *human judgement*. Ukuran untuk evaluasi *metric* adalah nilai korelasi antara *score metric* dan *human judgement*.

Pengevaluasian sistem mesin translasi menggunakan *automatic metric* menjadi lebih cepat, mudah dan murah dibandingkan *human evaluation*. Pada tugas akhir ini, pengimplementasian dilakukan dengan membuat modul *stemming* untuk teks berbahasa Indonesia yang diintegrasikan pada *tool METEOR metric* yang akan digunakan untuk mengevaluasi hasil terjemahan mesin translasi. Kemudian melakukan analisis terhadap pengaruh jumlah referensi yang digunakan serta pengaruh *stemming* yang diimplementasikan terhadap *score* dan nilai korelasi yang dihasilkan dari pengujian. Selain itu dilakukan analisis terhadap performa *METEOR metric* dengan *stemming* Indonesia dalam mengevaluasi beberapa mesin translasi serta membandingkan nilai korelasi antara *METEOR* dan *BLEU metric*.

Hasil pengujian menunjukkan jumlah referensi yang digunakan dalam pengevaluasian dapat meningkatkan *score metric* dan nilai korelasi yang dihasilkan. Modul *stemming* Indonesia yang diimplementasikan dapat memberikan kontribusi untuk meningkatkan *score metric* tetapi pengaruhnya terhadap nilai korelasi tidak signifikan. Hasil komparasi dengan *BLEU metric* menunjukkan bahwa *METEOR metric* mampu menghasilkan nilai korelasi yang lebih tinggi daripada *BLEU metric*. Sedangkan untuk evaluasi mesin translasi, Google Translator menghasilkan nilai korelasi paling tinggi sebesar 0.736, sedangkan Transtool menghasilkan nilai korelasi dengan urutan tertinggi kedua yaitu sebesar 0.478 lebih tinggi daripada Rekso Translator yang menghasilkan nilai korelasi sebesar 0.469.

**Kata kunci:** *METEOR metric, stemming, score, correlation value.*