

INTISARI

Perbandingan Metode Evaluasi Otomatis Akurasi Mesin Penerjemah

Oleh

Gunawan Setiadi
11/317202/PA/14266

Bahasa Inggris yang digunakan masyarakat secara global saat ini merupakan *lingua franca*, yaitu bahasa penyambung bagi dua pihak yang memiliki bahasa ibu berbeda. Sebagai bahasa ibu bagi lebih dari 360 juta orang dan bahasa asing bagi lebih dari 600 juta orang, Bahasa Inggris membutuhkan Mesin Penerjemah (MP) dengan kualitas tinggi dalam penerjemahan publikasi dari bahasa lain, termasuk di antaranya dari Bahasa Indonesia. Terdapat beberapa MP Bahasa Indonesia — Bahasa Inggris yang telah dikembangkan dan dapat diakses secara daring, beberapa di antaranya adalah Google Translate dan Tradukka.

Dalam memastikan MP tersebut memiliki kualitas terjemahan yang tinggi, digunakan beberapa metode evaluasi secara otomatis oleh komputer. Tiga buah metode evaluasi otomatis yang digunakan pada penelitian ini adalah *Match Accuracy* (MAcc), *Bilingual Evaluation Understudy* (BLEU), dan *Metric for Evaluation of Translation with Explicit Ordering* (METEOR). Dari antara ketiga metode tersebut, dilakukan perbandingan untuk menentukan metode evaluasi otomatis terbaik dalam mengevaluasi MP. Perbandingan dilakukan menggunakan nilai Korelasi dan nilai Mean Absolut Persentase Error (MAPE) dengan didukung oleh data dari ahli bahasa.

Penelitian ini mengevaluasi hasil terjemahan kedua MP terhadap teks Referensi yang telah ditentukan sebelumnya. Lebih lanjut, nilai evaluasi yang didapatkan dari metode BLEU, MAcc, dan METEOR terhadap 150 kalimat dibandingkan menggunakan perhitungan Korelasi dan MAPE. Hasil yang didapatkan menunjukkan metode BLEU pada lingkup unigram dengan nilai mean MAPE 0.220563866 merupakan metode evaluasi otomatis terbaik dibandingkan dengan metode evaluasi otomatis MAcc dan METEOR. Hasil selanjutnya yang didapatkan menunjukkan MP Google dengan nilai MAPE lebih rendah lebih baik dibandingkan dengan MP Tradukka pada nilai mean BLEU yang sama.

Kata kunci : Bahasa, Bahasa Indonesia, Bahasa Inggris, Mesin Penerjemah, Evaluasi Otomatis

ABSTRACT

Automatic Evaluation Comparison of Machine Translation Accuracy

by

Gunawan Setiadi
11/317202/PA/14266

English used by the global community nowadays is a *lingua franca*, a connecting language for two parties with different mother tongues. As the mother tongue for more than 360 million people and foreign languages for over 600 million people, English requires a high quality Machine Translation (MT) in translating publications from other languages, including from Indonesian. There are several Indonesian — English MT that has been developed and could be accessed online, some of which are Google Translate and Tradukka.

In ensuring said MT has a high-quality translations, several methods of automatic evaluation are used by a computer. Three of the automatic evaluation methods used in this research are Match Accuracy (MAcc), Bilingual Evaluation Understudy (BLEU), and Metric for Evaluation of Translation with Explicit Ordering (METEOR). Among those three methods, a comparison is performed to determine the best automatic evaluation method in evaluating MT. Comparisons are performed using Correlation value and Mean Absolute Percentage Error (MAPE) value supported with data from the linguists.

This research evaluates the results of both MT's translation against a predetermined reference text. Furthermore, the evaluation values obtained from BLEU, MAcc, and METEOR methods towards 150 sentences are compared using Correlation and MAPE calculations. The results gained indicated that BLEU method on unigram scope with MAPE Mean value of 0.220563866 is the best automatic evaluation method compared to the MAcc and METEOR automatic evaluation method. Next results obtained shows that Google MT with a lower MAPE value are better than Tradukka MT with the same value of BLEU Mean.

Keywords : Language, Indonesian, English, Machine Translation, Automatic Evaluation