

INTISARI

Seiring meningkatnya persebaran informasi lintas dunia seperti berita pada media sosial, peran fitur *auto-translate* yang disediakan menjadi krusial bagi pembaca. Untuk menentukan seberapa efektif sistem *auto-translate* yang ada, evaluasi penerjemahan mesin perlu dilakukan.

Penelitian ini bertujuan untuk menjabarkan bagaimana hasil evaluasi terjemahan luaran sistem *auto-translate* Facebook dan Twitter pada tajuk berita, dan keselarasan hasil evaluasi dengan minat baca. Terjemahan diambil dari kanal berita Reuters pada periode 1-7 September 2023, dan didapatkan 158 pasang tajuk berita pada 16 topik. Evaluasi linguistik dilakukan menurut taksonomi Farrus et al., yaitu Ortografi, Morfologi, Leksikal, Semantik dan Sintaksis. Dijabarkan juga kemungkinan alasan mengapa ketidaksepadanan linguistik tersebut terjadi. Kemudian, evaluasi persepsi dilakukan oleh 10 evaluator dengan kemampuan dwibahasa Indonesia-Inggris untuk menilai luaran terjemahan dengan skor 1-5. Lalu, metrik otomatis seperti BLEU, TER, BERTScore, BLEURT dan COMET dilakukan untuk mendapatkan penilaian luaran terjemahan dengan standar emas dari kantor penerjemahan Telitikata. Terakhir, survey dilakukan untuk mendapat preferensi dari responden minat baca terhadap luaran sistem yang diberikan.

Secara keseluruhan, ditemukan sekitar 530 ketidaksepadanan linguistik pada luaran terjemahan Facebook, dan 260 ketidaksepadanan linguistik pada luaran terjemahan Twitter. Kemudian, dari nilai rata-rata dari 10 evaluator, didapatkan penilaian 2,52 pada luaran terjemahan Facebook dan 3,76 pada luaran terjemahan Twitter. Hasil metrik otomatis juga menilai luaran Twitter lebih baik ketimbang Facebook. Lalu, dari 220 responden minat baca, secara umum ditemukan bahwa pembaca lebih meminati sistem dengan luaran terjemahan yang memiliki nilai lebih baik dengan korelasi lemah-sedang, dan tingkat keselarasan 59,49%-81.65%

Kata Kunci: Evaluasi Penerjemahan, Penerjemahan Mesin, Penerjemahan Berita, Ketertarikan Membaca

ABSTRACT

As news spreads across the world through social media, the role of the auto-translate feature provided becomes crucial for readers. To determine how effective existing auto-translate systems are, a machine translation evaluation needs to be performed.

This research aims to describe the evaluation results of auto-translated news headlines from Facebook and Twitter, and how the results aligned with reading interest. The translation was taken from the Reuters news channel on 1-7 September 2023, and 158 pairs of news headlines were obtained. Linguistic evaluation was carried out according to Farrus et al.'s taxonomy. The possible reasons why these linguistic errors occur are also explained. Then, a perception evaluation was carried out by 10 evaluators with bilingual Indonesian-English to assess the translation output with a score of 1-5. Then, automatic metrics such as BLEU, TER, BERTScore, BLEURT and COMET are carried out using a gold standard translation output assessment from Telitikata translation office. Finally, a survey was conducted to obtain preferences from respondents.

Overall, around 532 linguistic errors were found in Facebook's translation output, and 262 linguistic errors in Twitter's translation output. Then, from the average score of 10 evaluators, an assessment of 2.52 was obtained for the Facebook translation output and 3.76 for the Twitter translation output. Automatic metric results also assess Twitter's output as better than Facebook. Then, from 220 respondents interested in reading, it was generally found that readers were more interested in systems with translation output that had better scores with weak-moderate correlation, and an alignment of 59.49%-81.65%

Keywords: Translation Evaluation, Machine Translation, News Translation, Reading Interest