

PENINGKATAN AKURASI MESIN PENERJEMAH BAHASA INGGRIS- INDONESIA DENGAN MEMAKSIMALKAN KUALITAS DAN KUANTITAS KORPUS PARALEL

Herry Sujaini*¹

¹Program Studi Informatika, Universitas Tanjungpura

Email: ¹hs@untan.ac.id

*Penulis Korespondensi

(Naskah masuk: 10 Juni 2019, diterima untuk diterbitkan 22 April 2020)

Abstrak

Korpus paralel memiliki peran yang sangat penting dalam mesin penerjemah statistik (MPS). Korpus paralel yang diperoleh berbagai sumber biasanya memiliki kualitas yang kurang baik, sedangkan kuantitas korpus paralel merupakan tuntutan utama bagi hasil penerjemahan yang baik. Penelitian ini bertujuan untuk mengetahui efek ukuran dan kualitas korpus paralel di MPS. Penelitian ini menggunakan metode *bilingual evaluation understudy* (BLEU) untuk mengklasifikasikan pasangan kalimat paralel sebagai kalimat berkualitas tinggi atau buruk. Metode ini diterapkan ke korpus paralel yang berisi 1,5 M pasangan kalimat Inggris-Indonesia paralel dan memperoleh 900K pasangan kalimat paralel berkualitas tinggi. Beberapa sistem MPS dengan berbagai ukuran korpus paralel mentah dan korpus berkualitas tinggi yang difilter dilatih dengan MOSES dan dievaluasi kinerjanya. Hasil percobaan yang dilakukan menunjukkan bahwa ukuran korpus paralel merupakan faktor utama dalam kinerja terjemahan. Selain itu, kinerja terjemahan yang lebih baik dapat dicapai dengan korpus berkualitas tinggi yang lebih kecil menggunakan metode filter berkualitas. Hasil eksperimen pada MPS bahasa Inggris-Indonesia menunjukkan bahwa dengan menggunakan 60% kalimat yang kualitas terjemahannya baik, kualitas terjemahan dapat meningkat sebesar 7,31%.

Kata kunci: mesin penerjemah statistik, korpus paralel, MOSES, BLEU

IMPROVE THE ACCURACY OF ENGLISH-INDONESIAN TRANSLATOR MACHINES BY MAXIMIZING THE QUALITY AND QUANTITY OF PARALLEL CORPUS

Abstract

The parallel corpus has a very important role in the statistical machine translator (SMT) system. The parallel corpus obtained by various sources usually has poor quality, while the quantity of parallel corpus is the main demand for good translation results. This study aims to determine the effect of the size and quality of parallel corpus at SMT. This study uses the *bilingual evaluation understudy* (BLEU) method to classify pairs of parallel sentences as high-quality or bad sentences. This method is applied to a parallel corpus containing 1.5 M parallel English-Indonesian sentence pairs and obtaining 900K pairs of high-quality parallel sentences. Some SMT systems with various sizes of raw parallel bodies and high-quality corpus filtered are trained with MOSES and evaluated for performance. The experimental results show that the size of the parallel corpus is a major factor in translation performance. In addition, better translation performance can be achieved with a smaller high-quality corpus using a quality filter method. The experimental results in the English-Indonesian SMT show that by using 60% of sentences whose translation quality is good, the quality of the translation can increase by 7.31%.

Keywords: statistical machine translator, parallel corpus, MOSES, BLEU

1. PENDAHULUAN

Korpus paralel dapat didefinisikan sebagai koleksi teks yang disejajarkan atau diselaraskan dengan dua atau lebih bahasa lainnya. Korpus paralel dapat diterapkan pada pertanyaan teoretis termasuk studi tentang proses penerjemahan dan bagaimana mengekspresikan ide dalam dua (atau

lebih) bahasa yang berbeda, atau untuk membandingkan karakteristik dari teks asli (sumber) dan teks terjemahan. Pada saat yang sama, penggunaan praktis korpora paralel dalam studi terjemahan mencakup berbagai penggunaan korpora dalam pengembangan mesin penerjemah (Alotaibi, 2017).

Mesin Penerjemah Statistik (MPS) merupakan sebuah mesin penerjemah berdasarkan pendekatan model statistik, dimana parameter-parameternya bersumber dari hasil analisis korpus paralel bertipe teks. Korpus paralel juga berguna untuk aplikasi pemrosesan bahasa alami lainnya seperti pencarian informasi lintas bahasa, disambiguasi kata, dan proyeksi anotasi. Membangun korpus yang mencakup sejumlah besar kalimat paralel adalah salah satu pekerjaan yang paling memakan waktu dan penting untuk sistem MPS berkinerja tinggi (Hoang dkk., 2012).

Pelatihan komponen model translasi dalam MPS membutuhkan korpus paralel yang besar untuk parameter yang akan diprediksi (Taghipour dkk., 2010). Oleh karena itu, akurasi terjemahan yang lebih tinggi dapat dicapai ketika sistem terjemahan mesin dilatih untuk meningkatkan jumlah korpus paralel (Apriani dkk., 2016). Kualitas output sistem MPS sangat tergantung pada kualitas pasangan kalimat. Taghipour dkk., (2010) menyatakan bahwa sebagian kecil dari koleksi teks paralel yang tersedia dihasilkan secara alami di dalam organisasi dan pemerintah yang menggunakan multi-bahasa seperti PBB, Parlemen Eropa, dan Kanada untuk serangkaian bahasa yang terbatas. Kompilasi manual korpus paralel terlalu mahal, sehingga sebagian besar korpus paralel yang tersedia dihasilkan secara otomatis. Metode otomatis untuk menyusun pasangan kalimat paralel sebagian besar tidak tepat dan menggunakan korpus pelatihan berkualitas rendah yang memiliki banyak pasangan kalimat non-paralel akan menyebabkan terjemahan berkualitas rendah. Kesalahan dalam korpus yang dihasilkan secara otomatis mungkin disebabkan oleh perbedaan antara konten sumber dan dokumen target, terjemahan non-literal atau kesalahan perataan kalimat. Sangat tidak mungkin untuk menghilangkan kesalahan penyelarasan secara manual dalam sebuah korpus paralel yang besar. Oleh karena itu, metode evaluasi otomatis diperlukan untuk menentukan apakah pasangan kalimat paralel akurat atau tidak (Liu dan Ming, 2010). Metode evaluasi otomatis seperti itu dapat bergantung pada beberapa faktor seperti kata yang tumpang tindih, panjang kalimat, kebenaran tata bahasa, kelancaran, dan kebenaran penggunaan kata.

Proses penyaringan pasangan kalimat non-paralel dianggap sebagai langkah pasca-proses dari proses data mining dwibahasa. Santos (2011) berupaya untuk membandingkan kalimat untuk korpora paralel didasarkan pada panjang kalimat, bersama dengan penyelarasan kosa kata secara otomatis. Pendekatan berbasis panjang bekerja sangat baik pada pasangan bahasa dengan korelasi tinggi dalam panjang kalimat, seperti bahasa Perancis dan Inggris.

Yasuda dkk. (2008) mengembangkan metode pemilihan set pelatihan untuk sistem MPS Inggris-Cina menggunakan skor *perplexity*. Skor *perplexity*

dihitung untuk setiap kalimat dengan menggunakan probabilitas model bahasa dan jumlah kata dari kalimat. *Perplexity* pasangan kalimat paralel dianggap sebagai cara geometris dari *perplexity* kalimat sumber dan target. Mereka melatih sistem MPS dengan inialisasi korpus in-domain awal dan pasangan terjemahan terpilih yang memiliki *perplexity* lebih kecil dari ambang. Metode mereka meningkatkan skor BLEU sebesar 1,76% dibandingkan dengan penggunaan korpus in-domain awal.

Liu dan Zhou (2010) mengusulkan metode berbasis pembelajaran mesin untuk mengevaluasi kualitas penyelarasan. Fitur dan konstruksi linguistik yang diekstraksi dari kalimat disatukan bersama-sama ke dalam vektor fitur untuk mengkarakterisasi dimensi yang berbeda dari kualitas penyelarasan. Mereka menggunakan *Support Vector Machines* (SVM) untuk membedakan pasangan kalimat paralel yang berkualitas tinggi dan berkualitas rendah. Fitur dalam penelitian ini adalah jumlah kata yang salah eja, skor model bahasa untuk setiap kalimat bahasa Inggris, jumlah kata yang tidak terhubung yang disediakan oleh pengurai tata bahasa tautan dan ukuran ekivalensi terjemahan. Penjajaran kata diperoleh dengan menggunakan kamus dwibahasa yang komprehensif dan jumlah penjajaran kata dinormalisasi oleh panjang kalimat untuk mendapatkan pengukuran kesetaraan. Mereka melatih dengan pengklasifikasi SVM mereka pada 40K set pelatihan yang diperiksa oleh anotator dwibahasa. Klasifikasi mereka dilaporkan memiliki tingkat ketelitian dan penarikan 0,88.

Taghipour dkk. (2010) juga mengusulkan metode klasifikasi untuk membersihkan data paralel. Banyak fitur yang telah diuji dan digunakan untuk membangun model seperti probabilitas terjemahan berdasarkan model terjemahan IBM, jumlah kata yang selaras nol, fitur berdasarkan panjang dan fitur berdasarkan model bahasa. Mereka memilih model entropi maksimum untuk proses klasifikasi dan mereka mencapai akurasi 98,3% pada paralel paralel 48K Farsi-Inggris yang diperluas dengan noise buatan.

Dalam studi Cui dkk. (2013) metode tanpa pengawasan yang dinamakan algoritma random walk, dilakukan untuk menghitung skor kepentingan setiap pasangan kalimat secara iteratif yang menunjukkan kualitasnya. Metode mereka menggunakan probabilitas terjemahan frasa berdasarkan *Maximum Likelihood Estimation*. Mereka menguji metode mereka pada berbagai korpus paralel Cina-Inggris dan akhirnya, metode mereka meningkatkan kinerja sistem sekitar 0,5 ~ 1,0 nilai BLEU.

Munteanu dan Marcu (2005) melatih pengklasifikasi maximum entropy untuk mengekstraksi data paralel dari korpus surat kabar non-paralel besar Cina, Arab dan Inggris. Mereka mulai dari kumpulan paralel 5K kalimat pasangan

untuk menghitung keberpihakan kata dan mengekstraksi nilai fitur. Pengklasifikasi mereka menggunakan fitur-fitur berikut: panjang kalimat serta perbedaan panjang dan rasio panjang, persentase kata di setiap sisi yang memiliki terjemahan di sisi lain dan fitur perataan yang diperoleh dari model perataan kata. Mereka menggunakan kamus yang dipelajari dari korpus paralel awal yang ukurannya 100K, 1M, 10M, 50M, dan 95M token. Nilai *precision* dan *recall* pengklasifikasinya masing-masing adalah 0,97 dan 0,45 dengan data dari korpus domain-bahasa Arab-Inggris sedangkan nilai *precision* 0,94 dan *recall* 0,67 dengan korpus *out-of-domain* Arab-Inggris.

Hoang dkk. (2012) menyajikan model untuk mengekstraksi pasangan kalimat paralel dari sumber korpus non-paralel di domain yang berbeda. Mereka menggabungkan penyaringan berbasis panjang, kondisi kognitif dan pengukuran kesamaan konten. Kesamaan berdasarkan konten mereka adalah kesamaan antara kalimat target dan terjemahannya dari kalimat sumber yang diperoleh oleh sistem MPS. Awalnya mereka melatih sistem MPS dengan korpus *out-of-domain* awal (50K) dan menggunakan sistem ini untuk kesamaan berdasarkan konten. Setelah itu, mereka menyaring domain yang sebanding dalam domain (958K). Jika pasangan kalimat kandidat melewati kondisi pengukuran kesamaan, mereka memperluas data latih mereka dengan pasangan kalimat paralel baru ini. Mereka melatih kembali sistem MPS dan mengulangi proses memperluas paralel paralel serta meningkatkan sistem MPS. Mereka memperluas 50K korpus paralel awal menjadi 95K setelah 5 iterasi dan mencapai untuk meningkatkan skor terjemahan BLEU dari 8,92 menjadi 24,07.

Suatu pendekatan yang terkait dengan metode *supervised training* oleh Schwenk (2008) yang digunakan untuk menerjemahkan sejumlah besar teks monolingual, untuk menyaringnya dan menambahkannya ke data latih model terjemahan. Pendekatan lain yang memberikan bobot untuk setiap kalimat dalam bitext pelatihan diusulkan dalam Matsoukas dkk. (2009). Foster dkk. (2010) mendeskripsikan pendekatan baru untuk adaptasi MPS yang menentukan bobot pasangan frase domain sesuai dengan relevansinya dengan domain target. Axelrod dkk. (2011) mengusulkan metode untuk mengekstraksi kalimat dari korpus paralel umum berdomain besar dengan metode pemilihan data yang relevan dengan domain.

Penelitian ini berfokus pada efek menggunakan teks paralel yang besar versus menggunakan korpus berkualitas tinggi yang difilter. Eksperimen dilakukan dengan mengambil objek bahasa Inggris ke Indonesia.

2. METODE PENELITIAN

Eksperimen ini menggunakan korpus paralel OPUS yang diambil dari *subtitles* film dan TV (Tiedemann, 2016). OPUS merupakan korpus paralel *open source* dimana semua data dikumpulkan secara otomatis, dan sebagai hasilnya, teks asli berkualitas rendah. Jumlah kalimat dari korpus paralel OPUS yang dipergunakan pada penelitian ini adalah sebesar 1,5 M pasang kalimat, korpus ini selanjutnya disebut Korpus A. Meskipun kalimat-kalimat dalam sumber-sumber ini seharusnya paralel, pemeriksaan cepat dari kumpulan data ini memperlihatkan adanya masalah penyelarasan kalimat yang serius seperti terlihat pada Gambar 1.

Pasangan kalimat paralel dari setiap sumber telah diproses terlebih dahulu untuk mengatasi masalah keberpihakan dan mempertahankan tingkat kualitas penyejajaran dasar. Untuk mengeliminasi pasangan kalimat yang tampaknya tidak selaras dilakukan pemfilteran dengan menggunakan metode BLEU (*bilingual evaluation understudy*) (Sujaini, 2018). Nilai *n* yang digunakan adalah sebesar 10% seperti yang dipergunakan pada penelitian Sujaini dan bijaksana (2014). Setelah melakukan pemfilteran, didapatkan 900K pasang kalimat yang selaras, korpus ini selanjutnya disebut Korpus B. Contoh korpus B seperti pada Gambar 2.

You see this stuff every day , you know , people getting hurt and killed.	Kau melihat hal seperti ini setiap hari...
They're not sure.	Kau tahu , orang-orang yang terluka dan terbunuh... dan...
It freaks me out.	Mereka tidak yakin.
Just the thought of that happening to you , you know...	Ini membuatku takut
By the computer , that 's where I saw your keys.	Hanya berpikir jika itu terjadi padamu , kau tahu...
Hmm ?	Sebelah komputer...
They were under the newspaper.	Aku melihat kuncimu disana.
What would I do without you ?	Ada dibawah koran.
I gotta run.	Apa yang bisa kulakukan tanpamu ?

Gambar 1. Potongan Korpus A

Korpus A dan Korpus B digunakan untuk melatih sistem MPS terpisah menggunakan MOSES *toolkit* (Koehn dkk, 2007). Pada eksperimen ini, 90% data kalimat digunakan untuk data latih yang dijadikan sistem MPS, sedangkan 10% sisanya dijadikan data uji.

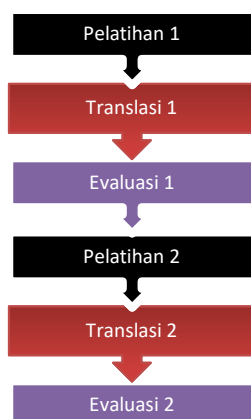
Penelitian ini menggunakan beberapa instrumen sebagai alat penelitian, yaitu:

1. Moses: sebagai *decoder*,
2. Giza++: sebagai alat latih model terjemahan,
3. SRILM: sebagai alat latih model bahasa, dan
4. BLEU: digunakan untuk penilaian hasil translasi (Papineni dkk., 2002).

What happened ?	Apa yang terjadi ?
What is wrong ?	Apa yang salah ?
Get away from her !	Menyingkir dari dia !
Call an ambulance.	Panggil ambulans.
She can 't leave here.	Dia tak bisa pergi dari sini.
Her heart stopped beating.	Jantungnya berhenti berdetak.
We are going to the hospital.	Kita harus ke Rumah Sakit.
Forty-five seconds	45 detik
Hurry up and open the door !	Cepat buka pintunya !
You 'll be in America by morning.	Kau akan berada di Amerika besok pagi.
We 're taking you to Washington , D.C.	Kami akan membawamu ke Washington , D.C.
Since when did Jolene Parker 's case go from missing person to murder investigation ?	Sejak kapan kasus Jolene Parker berubah dari orang hilang menjadi penyelidikan pembunuhan ?
The police say they have a person of interest.	Polisi bilang mereka punya orang yang berkepentingan.

Gambar 2. Potongan Korpus B

Metode yang digunakan untuk menghasilkan MPS dengan korpus yang lebih berkualitas sampai pada evaluasi hasil translasi MPS yang telah dioptimasi adalah seperti ditunjukkan pada Gambar 3.



Gambar 3. Tahapan eksperimen

Model bahasa dan model terjemahan dihasilkan pada proses pelatihan 1 dengan melakukan pelatihan dengan menggunakan Korpus A. Selanjutnya, model-model tersebut digunakan pada *decoder* MOSES (Translasi 1). Seluruh kalimat yang ada pada bahasa sumber pada korpus A dijadikan masukan pada Translasi 1 dengan bahasa target sebagai referensi. Selanjutnya, pada proses Evaluasi 1, masing-masing nilai BLEU kalimat terjemahannya dihitung terhadap kalimat referensi. Semua kalimat yang memiliki nilai BLEU yang kurang dari 0,1 dieliminasi, sisanya kalimat digunakan sebagai korpus paralel untuk mesin yang baru. Proses pelatihan 2 melakukan pelatihan untuk memperoleh model bahasa dan model terjemahan dengan menggunakan Korpus B hasil optimasi. Model-model tersebut digunakan pada *decoder* (Translasi 2). Selanjutnya, dilakukan proses Evaluasi 2 untuk mengukur nilai BLEU dari mesin yang baru.

Nilai BLEU didapat dari perhitungan dengan dimulai dengan menghitung *Brevity Penalty* (BP). BP dihitung dengan memilih referensi efektif dengan panjang token n-gram, dimana panjang token ditandai dengan r. Kemudian menghitung panjang total n-gram token terjemahan yang dilambangkan dengan c. Nilai BP kemudian dihitung berdasarkan persamaan (1) (Hadla dkk, 2015):

$$BP = \begin{cases} 1, & c > r \\ e^{(1-\frac{r}{c})}, & c \leq r \end{cases} \quad (1)$$

Kemudian nilai BLEU dicari dengan persamaan (2).

$$BLEU = BP \times \exp(\sum_{n=1}^N w_n \log p_n) \quad (2)$$

dimana :

$w_n = 1/N$ (pada eksperimen ini menggunakan nilai standar = 4), p_m = Output terjemahan berupa jumlah n-gram yang sesuai dengan rujukan terjemahan dari input dibagi jumlah n-gram hasil terjemahan.

3. HASIL DAN PEMBAHASAN

Pada Tabel 1, disajikan kinerja terjemahan dari masing-masing sistem. Ada perbedaan yang signifikan dalam kualitas terjemahan, kualitas mesin dari Korpus A meningkat sebesar $(33,49-31,21)/31,21 = 0,0731$ atau 7,31%.

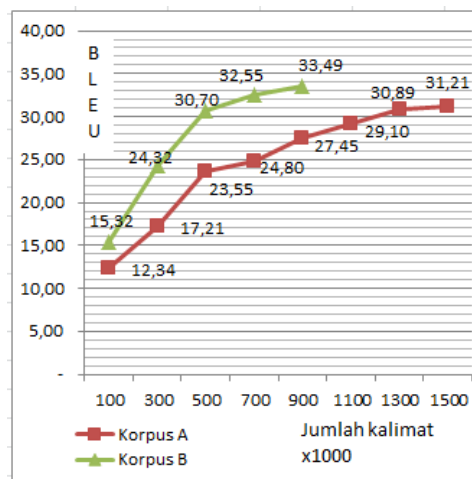
Tabel 1. Evaluasi Sistem MPS untuk Korpus A dan B

Korpus	Data Training	Data Uji	Nilai BLEU (%)
Korpus A	1.350 K	150 K	31.21
Korpus B	810 K	90 K	33.49

Pengamatan terperinci di dalam kumpulan data mengungkapkan fakta bahwa, bahkan setelah penyaringan, kumpulan data pada korpus B masih mengandung pasangan yang tidak selaras atau pasangan kalimat yang memiliki kualitas terjemahan yang buruk. Meskipun kinerja sistem MPS dapat ditingkatkan dengan memasukkan data monolingual yang lebih besar untuk komponen model bahasa (*language model*), itu juga merupakan ide yang baik untuk menyaring pasangan kalimat yang berkualitas rendah untuk meningkatkan model terjemahan (*translation model*). Selain itu, jelas bahwa penyaringan yang efektif secara substansial mengurangi fase pelatihan MPS yang memakan waktu bahkan jika itu tidak berkontribusi dalam akurasi terjemahan. Jadi, sebuah pengklasifikasi yang dapat membedakan pasangan kalimat berkualitas tinggi dari yang berkualitas rendah di korpus paralel dianggap sebagai langkah pra-pemrosesan yang menguntungkan.

Beberapa percobaan dilakukan dengan menggunakan Korpus A yang berupa data mentah dan Korpus B yang telah difilter. Untuk melihat dampak dari ukuran korpus paralel, Eksperimen dilakukan dengan melakukan tes menggunakan ukuran korpus berbeda hingga 1,5M kalimat. Teks di

sisi bahasa Indonesia sebagai bahasa target digunakan untuk pelatihan model bahasa dan 10% dari masing-masing korpus digunakan sebagai set data uji. Skor BLEU dari output MPS diplot pada Gambar 4.



Gambar 4. Skor BLEU berbagai variasi korpus

Seperti yang diharapkan, argumen yang menyatakan semakin banyak data, semakin tinggi akurasi MPS terbukti valid. Ketika jumlah kalimat paralel ditingkatkan dari 100K menjadi 1500K, skor BLEU meningkat dari 12,34 menjadi 31,21 untuk seluruh korpus A. Kecenderungan dalam plot menunjukkan bahwa skor BLEU yang lebih tinggi dapat dicapai dengan pengenalan lebih banyak data paralel dari 1500K. Begitu juga pada korpus B yang sudah difilter, skor BLEU meningkat dari 15,32 menjadi 33,49.

Di sisi lain, Gambar 4 mengungkapkan fakta penting lainnya bahwa kinerja terjemahan yang lebih baik dapat diperoleh dengan data pelatihan yang sedikit tetapi berkualitas tinggi. Dengan hanya menggunakan $900/1500 = 60\%$ dari data pelatihan mentah, skor BLEU meningkat sebesar $(33,49 - 31,21)/31,21 = 7,31\%$ yang dapat dicapai dengan penyaringan berbasis kualitas. Pengurangan dalam ukuran data pelatihan juga mengarah pada pengurangan kompleksitas waktu fase pelatihan sistem MPS yang memerlukan proses yang kompleks dan memakan waktu selama beberapa hari atau minggu.

4. KESIMPULAN

Makalah ini membahas masalah kualitas korpus bilingual di MPS. Percobaan menunjukkan bahwa metode penyaringan dapat digunakan untuk mengekstraksi pasangan kalimat yang tidak cocok dari korpus paralel yang berkualitas buruk dan pasangan yang tersisa masih dapat efektif dalam mencapai hasil setinggi hasil dari seluruh korpus mentah. Hasil penelitian ini juga menunjukkan bahwa masih ada kebutuhan untuk data paralel yang lebih banyak.

Ketersediaan sumber daya paralel dalam pasangan bahasa Inggris-Indonesia jauh lebih terbatas daripada ketersediaan pasangan bahasa yang memiliki sumber daya yang andal dan besar. Oleh karena itu, diperlukan lebih banyak data paralel, namun sumber daya yang tidak dapat diandalkan seperti Wikipedia dan subtitle film tidak boleh ditambahkan secara langsung sebagai data pelatihan untuk sistem MPS.

Metode penyaringan berguna dalam penggabungan sumber daya ini secara efektif dalam proses MPS. Walaupun hanya menggunakan 60% kalimat yang kualitas terjemahannya baik, MPS yang dibangun dapat meningkatkan kualitas sebesar 7,31%.

DAFTAR PUSTAKA

- APRIANI, T., SUJAINI. H., dan SAFRIADI, N., 2016. Pengaruh Kuantitas Korpus terhadap Akurasi Mesin Penerjemah Statistik Bahasa Bugis Wajo ke Bahasa Indonesia. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 4(1), pp. 168-173.
- ALOTAIBI, H.M., 2017. Arabic-English Parallel Corpus: A New Resource for Translation Training and Language Teaching, *Arab World English Journal (AWEJ)*, 8(3), pp. 319-337.
- AXELROD, A., HE, X., dan GAO., J., 2011. Domain adaptation via pseudo in-domain data selection. *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics*.
- CUI, L., ZHANG, D., LIU, S., LI, M., dan ZHOU, M., 2013. Bilingual data cleaning for SMT using graph-based random walk. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 2, pp. 340-345.
- FOSTER, G., GOUTTE, C.R., dan KUHN. 2010. Discriminative Instance Weighting for Domain Adaptation in Statistical Machine Translation. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Cambridge, US-MA, pp 451-459.
- HADLA, L.S., HAILAT, T.M., dan AL-KABI, M.N. 2015, Comparative Study Between METEOR and BLEU Methods of MT: Arabic into English Translation as a Case Study. *International Journal of Advanced Computer Science and Applications*, 6(11).
- HOANG, C., THAI, M.P., dan BAO, H.T., 2012. Exploiting non-parallel corpora for statistical machine translation, In *Proceedings of The 9th IEEE-RIVF International Conference on Computing*

- dan Communication Technologies, IEEE Computer Society, pp. 97 – 102.
- KOEHN, P., HOANG H., BIRCH, A., BURCH, C.C., FEDERICO, M., BERTOLDI, N., dan HERBST, E., 2007. Moses: Open source toolkit for statistical machine translation, In Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster dan Demonstration Sessions , pp. 177-180.
- LIU, X., dan ZHOU, M., 2010. Evaluating the quality of web-mined bilingual sentences using multiple linguistic features, Asian Language Processing (IALP), 2010 International Conference on. IEEE, 2010.
- MATSOUKAS, S., ROSTI, A.I., dan ZHANG, B., 2009. Discriminative Corpus Weight Estimation for Machine Translation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Singapore, pp. 708-717
- MUNTEANU, D.S., dan MARCU, D., 2005. Improving machine translation performance by exploiting non-parallel corpora, Computational Linguistics, 31(4), pp. 477-504.
- PAPINENI, K., ROUKOS, S., WARD T., dan ZHU, W.J., 2002. BLEU: A Method for Automatic Evaluation of Machine Translation, Proceedings of the 40th annual meeting on association for computational linguistics, pp. 311-318
- SANTOS, A., 2011. A survey on parallel corpora alignment, MI-STAR 2011, pp. 117–128.
- SCHWENK, H., 2008. Investigations on Large-Scale Lightly-Supervised Training for Statistical Machine Translation, In Proc. of the International Workshop on Spoken Language Translation Evaluation
- SUJAINI, H., 2018. Peningkatan Akurasi Penerjemah Bahasa Daerah dengan Optimasi Korpus Paralel. Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), 7(1).
- SUJAINI, H., dan BIJAKSANA, A., 2014. Strategi Memperbaiki Kualitas Korpus untuk Meningkatkan Kualitas Mesin Penerjemah Statistik, Seminar Nasional Teknologi Informasi XI Tahun 2014, Jakarta.
- TAGHIPOUR, K., AFHAMI, N., KHADIVI, S., dan SHIRY, S., 2010. A discriminative approach to filter out noisy sentence pairs from bilingual corpora, Telecommunications (IST), 5th International Symposium on 2010, pp. 537-541.
- TIEDEMANN, J., 2016. Finding Alternative Translations in a Large Corpus of Movie Subtitles. In Proceedings of the 10th International Conference on Language Resources dan Evaluation (LREC 2016).
- YASUDA, K., ZHANG, R., YAMAMOTO, H., dan SUMITA, E., 2008. Method of Selecting Training Data to Build a Compact dan Efficient Translation Model. In: Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), Hyderabad, India
- YILDIZ, E., dan CÜNEYD, T.A., 2012. Evaluation of Sentence Alignment Methods for EnglishTurkish Par-allel Texts, LREC 2012: The International Con-ference on Language Resources dan Evaluation. Istanbul.