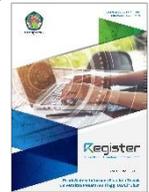


Tersedia online di www.journal.unipdu.ac.id
Unipdu

Terakreditasi S2 – SK No. 34/E/KPT/2018

Halaman jurnal di www.journal.unipdu.ac.id/index.php/register

Ambiguitas Machine Translation pada Cross Language Chatbot Bea Cukai

Muhammad Muharrom Al Haromainy^a, Dimas Ari Setyawan^b, Onny Kartika Waluya^c, Agus Zainal Arifin^d

^{a,b,c,d} Teknik Informatika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

email: ^aayommuharrom@gmail.com, ^bdimas.ariawan16@gmail.com, ^connykawe07@gmail.com, ^dagusza@cs.its.ac.id

INFO ARTIKEL

Sejarah artikel:

Menerima 18 Desember 2018
Revisi 10 April 2019
Diterima 10 April 2019
Online 29 Mei 2019

Kata kunci:

ambiguitas
chatbot
Cross Language
mesin translasi
POS Tagging

Keywords:

ambiguity
chatbot
Cross Language
machine translation
POS Tagging

Style APA dalam menyitasi artikel ini:

Al Haromainy, M. M. A., Setyawan, D. A., Waluya, O. K., & Arifin, A. Z. (2019). Ambiguitas Machine Translation pada Cross Language Chatbot Bea Cukai. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 5(1), 55-62.

ABSTRAK

Sistem *Information Retrieval* (IR) maupun *chatbot* semakin banyak dikembangkan. Salah satu bagian yang banyak diteliti adalah *cross language*. Masalah pada pengembangan *cross language* yaitu terjadinya kesalahan pada hasil terjemahan mesin translasi yang memberikan arti tidak sesuai dengan bahasa natural, sehingga pengguna tidak mendapatkan jawaban yang semestinya, bahkan tidak jarang pula pengguna tidak menemukan jawaban. Penelitian ini mengusulkan skema baru mesin translasi yang bertujuan meningkatkan performa dalam masalah ambiguitas. Mesin translasi bekerja dengan cek kebenaran kata kunci, kemudian melakukan *Part-of-Speech* (POS) *Tagging* pada kata benda (*noun*). Kemudian, setiap kata benda yang terdeteksi akan dicari sinonimnya. Lalu, sinonim yang didapatkan akan ditambahkan dan menjadi alternatif kueri baru. Kueri yang mempunyai nilai *confident* tertinggi diasumsikan sebagai kueri yang paling sesuai. Pada hasil yang didapatkan setelah dilakukan uji coba, melalui penambahan metode yang kami usulkan pada *machine translation*, dapat meningkatkan akurasi *chatbot* dibandingkan tanpa menggunakan skema yang diusulkan. Hasil akurasi bertambah 5%, dari yang semula 73% menjadi 77%.

ABSTRACT

Information retrieval and chatbot systems are increasingly being developed with its language part mostly studied. However, the problem associated with its development is the occurrence of errors in the translation machine resulting in inaccurate answers not in accordance with the natural language, thereby providing users with wrong answers. This study proposes a new translation machine scheme that aims to improve performance while translating ambiguous terms. Translation machines functions by checking the correctness of keywords, and carrying out Part-of-Speech (POS) Tagging on nouns (noun). The synonyms of any detected noun are searched for and obtained added to become alternative new queries. Those with the highest confident value are assumed to be the most appropriate. The results obtained after testing, through the addition of the method proposed in machine translation, can improve the accuracy of the chatbot compared to not using the proposed scheme. The results of the accuracy increased from the original 73% to 77%.

© 2019 Register: Jurnal Ilmiah Teknologi Sistem Informasi. Semua hak cipta dilindungi undang-undang.

1. Pendahuluan

Information Retrieval (IR) menjadi bahan penelitian yang masih populer beberapa tahun ini. Data yang semakin banyak dimanfaatkan untuk melakukan penelitian tentang IR (Asim, Wasim, Khan, Mahmood, & Mahmood, 2019). Penelitian yang dilakukan Madankar, Chandak, dan Chavhan (2016) menerapkan fungsi *cross language* pada sistem IR yang digunakan. Kemudian, IR yang diteliti Sharma dan Mittal (2016) menggunakan dual bahasa, yaitu bahasa Inggris dan India. Penelitian lainnya yang

dilakukan oleh Meade, Midgley, Dijkstra, dan Holcomb (2018) dan Zhang, Zhou, dan Zong (2016) menggunakan objek bahasa Mandarin pada IR yang dikembangkan. IR *cross language* juga telah dilakukan oleh Tran dan Christen (2015) yang diterapkan pada objek 5 bahasa, yaitu bahasa Inggris, Spanyol, Perancis, Jerman, dan Rusia.

Pada penerapan aplikasi yang lain, yaitu *chatbots*, membutuhkan pencarian sebuah informasi seperti sistem IR dan mengembalikan jawaban kepada penggunanya. Penelitian yang menggunakan *chatbots* dilakukan Fryer, dkk. (2017) untuk membandingkan efek dari pemakaian *chatbots* pada siswa kursus bahasa. Literatur lain yaitu Oard dan Wang (1999) dan Jovita, dkk. (2015) yang meneliti *chatbots* menghasilkan kesimpulan bahwa ketika seseorang melakukan percakapan terhadap mesin dapat berinteraksi lebih lama atau berbalas pesan lebih banyak, tetapi pesan tersebut pendek.

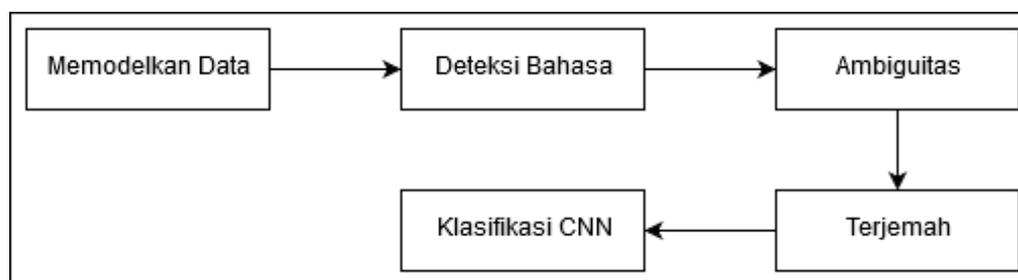
Menurut Qu, Nguyen, dan Shimazu (2016) ada tiga pendekatan untuk melakukan translasi: *machine translation*, *bilingual*, dan *parallel method*. Tujuan utama tranlasi adalah mendapatkan arti suatu kata atau kalimat dari suatu bahasa ke bahasa yang lain. Ada beberapa kesulitan mesin translasi, seperti kata-kata yang mempunyai makna ganda dan kalimat dengan struktur bahasa ganda. Pada metode berbasis korpus, pertanyaan yang diterjemahkan berdasarkan istilah multibahasa diekstraksi dari koleksi dokumen paralel (Feng, et al., 2017). Penelitian yang menggunakan pendekatan berbasis kamus mengeksplorasi kamus yang dapat dibaca oleh mesin dan strategi pemilihan seperti memilih semua, pilih n terbaik, dan pilih secara acak.

Penelitian seputar *cross language* menggunakan *machine translation* ditambah dengan deteksi *disambiguation* pada preposisi kalimat telah dilakukan (Kumar, Rajendran, & Soman, 2015). Begitu juga Zhou, dkk. (2016) yang telah melakukan penelitian tentang *ambiguity machine translation*, tetapi berfokus pada ambiguitas frasa dan kata majemuk yang diterapkan pada IR bahasa Melayu ke bahasa Inggris. Namun, pada beberapa literatur, belum ada penelitian yang melakukan antisipasi terhadap masalah yang signifikan, yaitu ambiguitas dari hasil terjemahan *machine translation*. Masalah yang terjadi seperti pada kata *custom*, jika dilakukan metode translasi biasa akan didapatkan hasil "adat istiadat", padahal hasil yang diinginkan adalah "bea cukai", sehingga hal tersebut mengakibatkan sistem tidak mendapatkan hasil yang sesuai dan bahkan tidak memunculkan hasil pencarian.

Penelitian ini mengusulkan skema baru untuk mengurangi masalah ambiguitas mesin translasi. Kemudian, sistem *chatbot* yang dibangun mendapatkan hasil yang sesuai keinginan pengguna, sehingga permasalahan ambiguitas mesin translasi dapat ditangani dengan metode yang kami usulkan. Pertama, *chatbots* mendeteksi bahasa yang dimasukkan pengguna. Kedua, terjemahkan menggunakan *machine translation*. Ketiga, ambil semua kemungkinan hasil terjemahan, biasanya pada mesin translasi hanya muncul satu terjemahan. Keempat, pilih kueri yang berhubungan dengan data sistem dan yang mempunyai nilai *confident* terbaik.

Penelitian ini dibagi menjadi beberapa bagian. Pada bagian kedua akan dijelaskan tentang *reseacrh methodology* dari penelitian ini. Bagian ketiga akan disajikan hasil dan pembahasan dari percobaan yang telah dilakukan. Bagian keempat menampilkan kesimpulan dari hasil percobaan.

2. Metode Penelitian



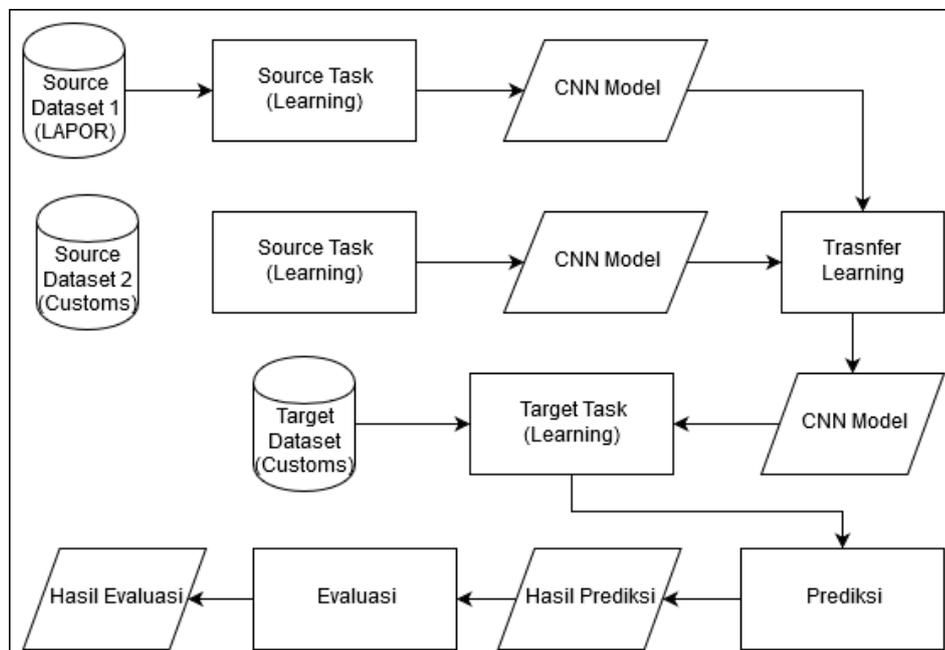
Gambar 1. Arsitektur metode penelitian

Pada Gambar 1 dijelaskan langkah penelitian yang dilakukan yaitu membentuk *Global vectors for word representations* (Glove), kemudian dilakukan pembentukan model *Convolutional Neural Network* (CNN). *Testing* dilakukan dengan cara memasukkan beberapa pertanyaan lalu dideteksi bahasanya, jika berupa bahasa Inggris akan dilakukan deteksi ambiguitas dan diterjemahkan ke dalam bahasa Indonesia. Selanjutnya, diolah oleh CNN, sehingga menghasilkan jawaban. Hasil tersebut akan dievaluasi menggunakan akurasi, presisi, *recall*, dan *f-measure*.

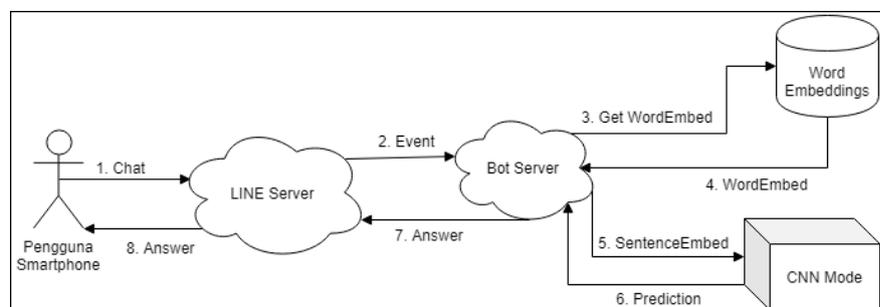
2.1. Global vector for word representation (Glove) dan Convolutional Neural Networks (CNN)

Penelitian ini mengusulkan skema arsitektur baru untuk mencegah ambiguitas hasil *machine translation* yang diterapkan pada aplikasi *chatbot* bea cukai. Pertama kali, kami membuat data Glove (*Global vector for word representation*). *Output* dari Glove adalah vektor kata yang merepresentasikan makna dari setiap kata (*word embeddings*). Untuk mendapatkan *word embeddings*, kami melakukan *parsing* korpus dalam jumlah yang besar (bisa puluhan juta atau bahkan miliar). Korpus yang digunakan pada penelitian ini adalah Wikipedia Bahasa Indonesia dan LAPOR! (portal pemerintah yang menyimpan data <https://data.go.id>).

Glove *word embedding* menjadi input data, kemudian dimodelkan dalam bentuk CNN (*Convolutional Neural Networks*). Dalam pembentukan model CNN, dilakukan 6 tahap, yaitu *text preprocessing*, *sentence mapping*, pemisahan *dataset*, *CNN learning*, prediksi, dan evaluasi. Penelitian ini menggunakan arsitektur seperti Gambar 2 dengan 2 *source task* yang terdiri atas *dataset* LAPOR! dan bea cukai. Dari 2 data tersebut dilakukan proses *transfer learning* kembali untuk membentuk model CNN dan dilatih menggunakan data FAQ bea cukai. Pembentukan model CNN menggunakan *layer* konvolusi dari *source task* LAPOR!, sedangkan *layer hidden* dan *layer output* menggunakan *source task* dari bea cukai.



Gambar 2. Glove dan CNN



Gambar 3. Alur komunikasi *end-point* dengan *chatbot*

2.2. Chatbot

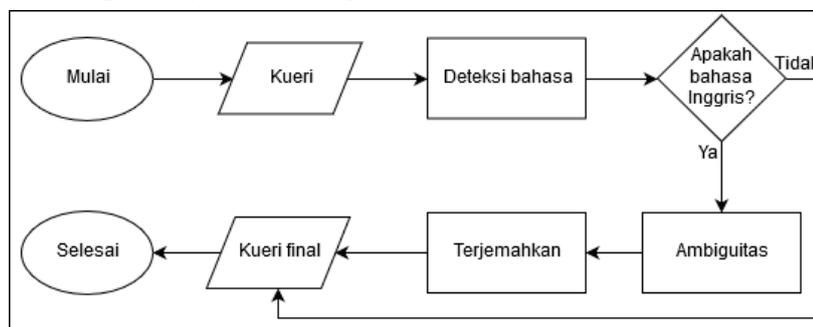
Model yang didapat dari *transfer learning* akan digunakan untuk memprediksi kelas dari kalimat-kalimat yang masuk pada sistem *chatbot*. Pada penelitian ini, digunakan *LINE Messenger API* (*Application Program Interface*), sehingga interaksi *chat* melalui aplikasi LINE. *Chat* masuk akan dikenali terlebih dahulu, termasuk *direct question* atau tidak. Jika tidak, *chat* kemudian dijadikan input model CNN dan

setelah itu dicek kelas dengan nilai *probability* tertinggi. Apabila nilai *probability* kurang dari ambang (*threshold*), *chat* tidak termasuk dalam kelas dan *bot* akan memberikan jawaban kelas tidak ditemukan. *Bot* akan memberikan jawaban seputar Bea Cukai Negara Indonesia. Gambar 3 menjelaskan arsitektur aplikasi mulai pengguna mengirim pertanyaan sampai dijawab oleh *bot*.

2.3. Cross Language Machine Translation

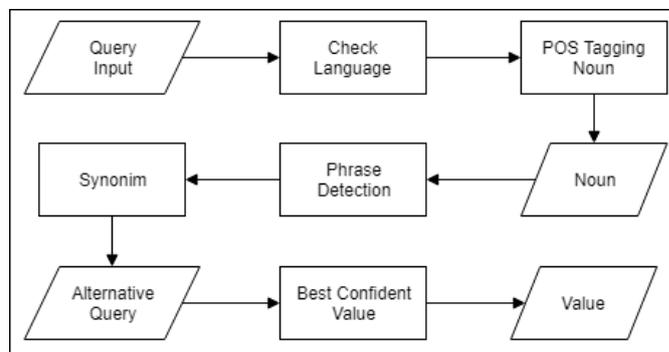
Bahasa yang digunakan pada sistem *chatbot* pada penelitian ini berfokus pada dua bahasa, yaitu bahasa Inggris dan Indonesia. Ketika kueri dikenali berupa bahasa Indonesia, akan diproses biasa tanpa translasi. Namun, ketika sistem mendeteksi kueri masukan adalah bahasa Inggris, akan dilakukan translasi dan pengenalan ambigu. Bahasa kueri memengaruhi hasil *chat* yang akan disampaikan kepada pengguna.

Penerjemahan bahasa dari kueri yang dimasukkan pengguna menggunakan metode mesin translasi. Karena hasil terjemahan mesin translasi dengan metode biasa kurang maksimal, diusulkan skema seperti Gambar 4 untuk memperbaiki hasil terjemahan yang memiliki arti ambigu agar hasil translasi tersebut mendapatkan jawaban yang sesuai.



Gambar 4. Alur *cross language* mesin translasi

2.4. Ambiguitas mesin translasi



Gambar 5. *Cross language* ambiguitas

Kata ambigu didapatkan karena hasil terjemahan dari bahasa Inggris ke bahasa Indonesia dapat berupa banyak kata. Misalnya, satu kata dalam bahasa Inggris dapat menjadi berbagai macam kata jika diartikan ke dalam bahasa Indonesia, seperti kata *custom* yang artinya adat, bea cukai, kebiasaan, dan lain sebagainya. Teknik yang diusulkan untuk menangani ambiguitas terjemahan di *cross language information retrieval dictionary-based* melibatkan POS (*Part-of-Speech*). Seperti pada penelitian Madankar, dkk. (2016) yang menggunakan bahasa Inggris ke bahasa India, ditemukan bahwa terjemahan *dictionary-based* yang dilengkapi dengan disambiguasi POS atau disambiguasi berbasis POS meningkatkan efektivitas. Langkah yang kami usulkan setelah pengguna menuliskan pertanyaan adalah sebagai berikut. Pertama, dilakukan *spelling checker query* untuk memastikan apakah pertanyaan yang diberikan pengguna benar atau ada kesalahan. Jika benar, dilanjutkan POS (*Part-of-Speech*) *Tagging*, maksudnya adalah menyeleksi atau memilih bagian kata. Penelitian ini mengambil kata benda (*noun*) saja untuk diproses ke bagian selanjutnya. Dari kata benda yang telah terpilih, dicari frasa dan sinonim dari proses sebelumnya. Pada tahap ini, sistem sudah mendapatkan berbagai macam terjemahan yang kemungkinan relevan dengan pertanyaan awal. Kemudian, berbagai macam terjemahan tersebut dibentuk ulang menjadi macam-macam pertanyaan baru sehingga terbentuklah

alternatif kueri input. Untuk menentukan final kueri yang akan dipakai, dipilih alternatif kueri yang mempunyai nilai *confident* tertinggi. Gambar 5 menunjukkan proses ambiguitas pada mesin translasi.

Nilai *confident* merupakan nilai *probability* dari proses klasifikasi CNN yang digunakan untuk memprediksi kelas yang terdapat pada Tabel 1. Ketika sistem sudah ditambahkan dengan metode yang kami usulkan, akan mendapatkan banyak alternatif pertanyaan dan menghasilkan banyak nilai *confident* yang dapat dibandingkan. Nilai *confident* menampilkan angka nol sampai satu. Hasil perhitungan diurutkan dari nilai terbesar ke terkecil. Hasil nilai terbesar atau yang mendekati angka satu, maka itulah hasil jawaban yang terbaik.

Tabel 1. Kelas FAQ Bea Cukai

Bahasa Indonesia	Kelas
Pengertian Bea masuk	1
Pengertian Impor	2
Pengertian ekspor	3
Pengertian daerah Pabean	4
Pengertian kawasan Pabean	5
Pengertian cukai	6
Pengertian Barang Kena Cukai (BKC)	7
Aturan atau Undang-undang BKC	8
Pengertian Penerimaan Negara Bukan Pajak (PNBP)	9
Membawa binatang dan tumbuhan ke Indonesia	10
Kunjungan pertama ke Indonesia	11
Bea masuk untuk barang impor	12
Uang tunai	13
Rokok dan minuman beralkohol	14
Waktu importasi	15
Kelas tidak ditemukan	16

2.5. Data

Data yang digunakan adalah data dari bea cukai berupa pertanyaan yang paling sering ditanyakan/FAQ (*Frequently Asked Question*) yang diambil dari situs resmi <http://www.beacukai.go.id/>. Di sana sudah disediakan *dataset*, baik bahasa Indonesia maupun bahasa Inggris, serta terdapat sekitar 15 kelas untuk masing-masing bahasa yang akan dipakai. Dengan demikian, 15 kelas tersebut akan dijadikan sebagai data latih dari prediksi jawaban dan 1 kelas lain untuk kueri yang tidak sesuai, yaitu kelas 16.

		Predicted	
		Positive(+)	Negative(-)
Actual	Positive(+)	True Positive (TP)	False Positive (FP)
	Negative(-)	False Negative (FN)	True Negative (TN)

Gambar 6. *Confusion matrix*

Gambar 6 menunjukkan *confusion matrix*, yang pada penelitian ini, *True Positive* (TP) adalah jumlah kueri yang kelasnya benar dan hasil klasifikasi benar. *True Negative* (TN) menyatakan jumlah kueri yang tidak ada kelas dan klasifikasinya tidak ada kelas. *False Positive* (FP) merupakan jumlah kueri yang benar dan hasil klasifikasinya salah tidak sesuai dengan kelas. Kemudian, *False Negative* (FN) adalah jumlah kueri yang benar dan hasil klasifikasinya tidak ada kelas. *Confusion matrix* tersebut digunakan untuk mengevaluasi akurasi, presisi, *recall*, dan *f-measure* ambiguitas mesin translasi. Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai asli. Presisi adalah tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan sistem. *Recall* adalah nilai ketepatan informasi yang diprediksi relevan pada suatu kelas terhadap data asli pada kelas tersebut.

Sementara, *f-measure* memanfaatkan nilai presisi dan *recall* yang dijadikan alasan penggunaan beberapa evaluasi untuk menyeimbangkan nilai evaluasi yang seimbang, karena presisi dan *recall* cenderung berbanding terbalik hasilnya.

$$\text{Akurasi} = \frac{TP + \sum TN}{TP + \sum TN + \sum FP + \sum FN} \quad (1)$$

$$\text{Presisi} = \frac{TP}{\sum TP + \sum FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{\sum TP + \sum FN} \quad (3)$$

$$F - \text{Measure} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

3. Hasil dan Pembahasan

Eksperimen yang dilakukan menggunakan data *testing* berupa 75 pertanyaan berbahasa Inggris yang terdiri atas 15 kelas jawaban dan 1 kelas jawaban untuk kueri yang tidak mempunyai kelas, yaitu kelas 16. Pada Tabel 2, kami menunjukkan sampel hasil percobaan sejumlah 15 pertanyaan. Kami membanding dua skema percobaan, pertama, saat ditambahkan metode yang kami usulkan, kedua, tanpa menggunakan metode yang kami usulkan agar mudah untuk dianalisis dampaknya.

Tabel 2. Sampel data *testing*

No	Pertanyaan	Kelas	Kelas tanpa metode	Kelas dengan metode
1	<i>what is the import duty</i>	1	1	1
2	<i>what is the purpose of import?</i>	2	6	2
3	<i>what kind of export?</i>	3	12	3
4	<i>I want to know about customs territory</i>	4	16	16
5	<i>what is tax customs?</i>	6	6	16
6	<i>What is excisable goods?</i>	7	6	6
7	<i>will go to Indonesia with my dog</i>	10	16	10
8	<i>what kind of wine?</i>	14	16	7
9	<i>type of customs territory</i>	4	5	4
10	<i>do you agree with export?</i>	16	16	16
11	<i>I will go to Bali for the first time</i>	11	11	11
12	<i>I want to bring an apple tree to indonesia</i>	10	16	10
13	<i>what is customs area</i>	5	5	5
14	<i>time to import goods</i>	15	15	15
15	<i>what is customs</i>	6	7	6

Pertanyaan yang digunakan pada Tabel 2 menggunakan bahasa Inggris. Pertanyaan yang dilakukan untuk data *testing* terdiri atas pertanyaan dasar biasa dan pertanyaan yang menyebabkan ambigu seperti pada pertanyaan nomor 12 dan nomor 15. Kata *apple* bisa berarti 2 dalam bahasa Indonesia, yaitu apel sebagai buah dan *apple* sebagai produk. Sementara, kata *customs* bisa berarti adat istiadat dan bea cukai dalam bahasa Indonesia.

Kueri nomor 2, nomor 3, nomor 7, dan nomor 9 dapat diklasifikasikan sesuai dengan kelasnya ketika menggunakan skema terbaru. Skema yang lama tanpa deteksi ambiguitas tidak bisa mengklasifikasikan dengan benar, karena jumlah terjemahan kueri hanya satu, sedangkan untuk skema baru terjemahan kueri bisa banyak karena diambil per *noun* dan dipilih nilai probabilitas yang tertinggi.

Tabel 3. Perbandingan nilai *confident*

Skenario	Terjemahan (<i>i want to bring an apple tree to indonesia</i>)	Kelas	Nilai <i>confident</i>
Dengan skema ambiguitas	saya ingin membawa pohon apel ke indonesia	10	0,971
	saya ingin membawa pohon malus pumila ke indonesia	10	0,951
	saya ingin membawa pohon pohon apel ke indonesia	10	0,938
Tanpa skema	saya ingin membawa pohon apel ke indonesia	16	0,673

Kemudian, pada Tabel 3 menunjukkan nilai *confident* kueri dengan menggunakan skema ambiguitas lebih tinggi dibandingkan dengan yang tanpa skema. Meskipun terjemahan tanpa skema sama dengan yang menggunakan skema ambiguitas. Hal ini dipengaruhi oleh banyaknya kueri masukan per *noun* pada skema ambiguitas sehingga nilai *confident* lebih tinggi.

Tabel 4. Hasil percobaan

Skenario	Accuracy	Precision	Recall	F1
Skema	0,77	0,83	0,90	0,86
Ambiguitas				
Tanpa skema	0,73	0,82	0,85	0,83

Dari hasil percobaan yang tampak pada Tabel 4, terlihat penggunaan metode untuk mengatasi ambigu. Uji coba yang dilakukan meliputi 4 evaluasi, seperti Persamaan 1 akurasi, Persamaan 2 presisi, Persamaan 3 *recall*, dan Persamaan 4 *f-measure*, sedangkan Gambar 6 berupa *confusion matrix* sebagai langkah awal sebelum menghitung 4 evaluasi tersebut.

Hasil pengujian meningkat, seperti akurasi semula 73% menjadi 77% dan *recall* 85% menjadi 90%. Peningkatan hasil pengujian paling banyak dipengaruhi oleh percobaan kueri yang mempunyai makna ambigu. Seperti pada contoh pertanyaan nomor 12, "*I want to bring an apple tree to Indonesia*", tanpa metode yang kami usulkan, tidak dapat memunculkan hasil jawaban yang sesuai. Kemudian, pada pertanyaan nomor dua, "*what is purpose of import*", ketika menggunakan metode ini berhasil diklasifikasikan dengan benar.

4. Kesimpulan

Dari hasil pengujian yang telah dilakukan bahwa pemakaian metode untuk mengatasi hasil translasi yang ambigu berhasil dilakukan dan meningkatkan hasil evaluasi percobaan, seperti akurasi, presisi, *recall*, dan *f-measure*. Ketika metode yang kami usulkan tidak diterapkan, beberapa pertanyaan tidak dapat menghasilkan jawaban yang benar karena terjadi kesalahan terjemahan dari *machine translation*. Secara detail, peningkatan evaluasi yang terjadi yaitu akurasi semula 73% menjadi 77%, presisi 82% menjadi 83%, *recall* 85% menjadi 90%, dan *f-measure* 83% menjadi 86%. Pada penelitian selanjutnya, kami mengusulkan penggunaan metode optimasi lain.

5. Ucapan Terima Kasih

Kami sangat berterima kasih kepada dosen mata kuliah topik Sistem Temu Kembali Informasi, yaitu Bapak Agus Zainal yang telah membimbing dan memberi saran pada penelitian yang telah kami lakukan, seperti pengembangan metode yang harus dilakukan. Kami juga mengucapkan terima kasih kepada asisten dosen yang menjaga semangat dan membantu kami di luar jam pelajaran, serta teman sekelas yang kadang-kadang kami jadikan tempat berkonsultasi mengenai aplikasi dan dukungan ketika ada masalah.

7. Referensi

- Asim, M. N., Wasim, M., Khan, M. U., Mahmood, N., & Mahmood, W. (2019). The Use of Ontology in Retrieval: A Study on Textual, Multilingual, and Multimedia Retrieval. *IEEE Access*, 7, 21662-21686.
- Feng, X., Huang, L., Qin, B., Lin, Y., Ji, H., & Liu, T. (2017). Multi-level cross-lingual attentive neural architecture for low resource name tagging. *Tsinghua Science and Technology*, 22(6), 633-645.

- Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, 75, 461-468.
- Jovita, J., Linda, L., Hartawan, A., & Suhartono, D. (2015). Using Vector Space Model in Question Answering System. *Procedia Computer Science*, 59(2015), 305-311.
- Kumar, M. A., Rajendran, S., & Soman, K. P. (2015). Cross-Lingual Preposition Disambiguation for Machine Translation. *Procedia Computer Science*, 54(2015), 291-300.
- Madankar, M., Chandak, M. B., & Chavhan, N. (2016). Information Retrieval System and Machine Translation: A Review. *Procedia Computer Science*, 78(2016), 845-850.
- Meade, G., Midgley, K. J., Dijkstra, T., & Holcomb, P. J. (2018). Cross-language Neighborhood Effects in Learners Indicative of an Integrated Lexicon. *Journal of Cognitive Neuroscience*, 30(1), 70-85.
- Oard, D., & Wang, J. (1999). Effects of term segmentation on Chinese/English cross-language information retrieval. *6th International Symposium on String Processing and Information Retrieval*. Cancun, Mexico, Mexico: IEEE.
- Qu, J., Nguyen, L. M., & Shimazu, A. (2016). Cross-language information extraction and auto evaluation for OOV term translations. *China Communications*, 13(12), 277-296.
- Sharma, V. K., & Mittal, N. (2016). Exploiting Wikipedia API for Hindi-english Cross-language Information Retrieval. *Procedia Computer Science*, 89(2016), 434-440.
- Tran, K.-N., & Christen, P. (2015). Cross-Language Learning from Bots and Users to Detect Vandalism on Wikipedia. *IEEE Transactions on Knowledge and Data Engineering*, 27(3), 673-685.
- Zhang, J., Zhou, Y., & Zong, C. (2016). Abstractive Cross-Language Summarization via Translation Model Enhanced Predicate Argument Structure Fusing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(10), 1842-1853.
- Zhou, G., Xie, Z., He, T., Zhao, J., & Hu, X. T. (2016). Learning the Multilingual Translation Representations for Question Retrieval in Community Question Answering via Non-negative Matrix Factorization. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(7), 1305-1314.