



Penggunaan *Pivot Language* pada Mesin Penerjemah Statistik Bahasa Inggris ke Bahasa Melayu Sambas

Della Widya Ningtyas^{#1}, Herry Sujaini^{#2}, Novi Safriadi^{#3}

[#]Program Studi Informatika Universitas Tanjungpura
Jalan Prof. Dr. H. Hadari Nawawi, Pontianak, Kalimantan Barat

¹dellawidyaningtyas@gmail.com

²hs@untan.ac.id

³safriadi@informatics.untan.ac.id

Abstrak - Perbedaan kemampuan berbahasa dalam masyarakat dapat menjadi penghalang pertukaran informasi, salah satu cara untuk menanggulangnya adalah dengan memanfaatkan teknologi mesin penerjemah. Saat ini masih sulit menemukan korpus paralel bahasa Inggris – bahasa daerah, sehingga digunakan bahasa Indonesia sebagai bahasa perantara (*pivot language*) untuk mempermudah dalam pembuatan korpus paralel bahasa Inggris – bahasa daerah. Tujuan dari penelitian ini adalah untuk mengetahui pengaruh *pivot language* terhadap nilai akurasi mesin penerjemah statistik bahasa Inggris ke bahasa Melayu Sambas. Penelitian menggunakan korpus paralel sebanyak 3000 baris kalimat. Pengujian dilakukan dengan dua cara yaitu pengujian secara otomatis menggunakan BLEU (*Bilingual Evaluation Understudy*). Pengujian dilakukan dengan membandingkan nilai akurasi terjemahan dari mesin penerjemah Bahasa Inggris – Melayu Sambas dan mesin penerjemah Bahasa Inggris – Melayu Sambas dengan Bahasa Indonesia sebagai *pivot language*. Dari hasil pengujian yang telah dilakukan, penggunaan *pivot language* pada mesin penerjemah statistik Bahasa Inggris - Bahasa Melayu Sambas dapat mempengaruhi akurasi terjemahan, terlihat terjadi peningkatan nilai akurasi sebesar 5,015% dari nilai pengujian semula yang didapatkan dari mesin penerjemah tanpa *pivot* dengan metode BLEU. Berdasarkan data tersebut penggunaan *pivot language* pada mesin penerjemah statistik bahasa Inggris – bahasa Melayu Sambas dapat meningkatkan nilai akurasi terjemahan.

Kata kunci— *Pivot Language*, BLEU, Mesin Penerjemah Statistik

I. PENDAHULUAN

Bahasa merupakan alat komunikasi yang digunakan seseorang untuk menyampaikan ide, gagasan, konsep atau perasaan kepada orang lain. Ragam bahasa yang dimiliki setiap orang berbeda, hal ini yang terkadang menghalangi pertukaran informasi karena orang lain tidak memahami maksud dan tujuan yang ingin disampaikan.

Sebagai salah satu bahasa daerah yang ada di Kalimantan Barat, bahasa Melayu Sambas adalah bahasa

Melayu yang dituturkan oleh masyarakat asli kabupaten Sambas. Bahasa Melayu Sambas sendiri sampai saat ini memang jauh dari kepunahan, dikarenakan penutur bahasa ini relatif masih banyak walaupun tidak diketahui secara pasti berapa penutur bahasa Melayu Sambas.

Perbedaan bahasa dapat menjadi penghalang pertukaran informasi, maka saat ini sedang dikembangkan teknologi mesin penerjemah. Telah banyak dilakukan penelitian mengenai pengembangan mesin penerjemah statistik bahasa Inggris ke bahasa Indonesia dan bahasa Indonesia ke bahasa daerah sehingga sudah banyak tersedia korpus paralel bahasa Inggris – bahasa Indonesia dan korpus paralel bahasa Indonesia – bahasa daerah. Namun saat ini masih sulit menemukan korpus paralel bahasa Inggris – bahasa daerah, sehingga digunakan bahasa Indonesia sebagai bahasa perantara (*pivot language*) untuk mempermudah dalam memperoleh korpus paralel bahasa Inggris – bahasa daerah. Terdapat beberapa penelitian yang dilakukan berkaitan dengan *pivot language*, diantaranya penelitian tentang perbandingan metode *pivot* untuk mesin penerjemah statistik berbasis frasa [1], penelitian tentang penggunaan bahasa Inggris sebagai *pivot language* pada mesin penerjemah statistik bahasa Arab - Cina [2], penelitian tentang pentingnya pemilihan *pivot language* untuk mesin penerjemah statistik dengan meneliti bahasa selain bahasa Inggris sebagai *pivot language* [3], penelitian tentang menggunakan *pivot language* dalam mesin penerjemah statistik bahasa Cina – Spanyol dengan bahasa Inggris, bahasa Arab dan bahasa Prancis sebagai *pivot language* [4], dan penelitian tentang pendekatan *pivot* untuk memisahkan pola parafrase dari korpus *bilingual* [5].

Berdasarkan permasalahan di atas, maka akan dilakukan penelitian dengan mengimplementasikan *pivot language* pada mesin penerjemah statistik bahasa Inggris ke bahasa Melayu Sambas, dimana bahasa Indonesia digunakan sebagai *pivot language*. Selanjutnya dilakukan uji akurasi penerjemahan untuk mengetahui seberapa

besar pengaruh dari penggunaan bahasa Indonesia sebagai *pivot language* terhadap hasil terjemahan.

Berikut merupakan komponen yang digunakan dalam pembangunan sistem ini.

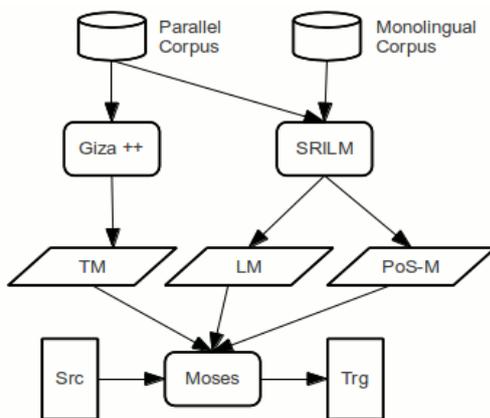
A. Mesin Penerjemah Statistik

Mesin penerjemah statistik (MPS) atau *Statistical machine translation* (SMT) adalah suatu paradigma dari mesin penerjemah dimana penerjemahan dilakukan berbasis model statistik dengan parameter-parameter yang diturunkan dari analisis paralel korpus [6].

Ada beberapa pendekatan untuk *machine translation* seperti pendekatan dengan menggunakan aturan (*rule-based machine translation*), pendekatan dengan menggunakan contoh (*example-based machine translation*), dan pendekatan dengan menggunakan model statistik (*statistical machine translation*). Dalam mesin penerjemah statistik, terdapat 3 komponen yang terlibat dalam proses penerjemahan dari satu bahasa ke bahasa lain yaitu: *language model*, *translation model*, dan *decoder* [7].

Moses merupakan *software open source* yang merupakan implementasi dari Mesin Penerjemah Statistik. Moses digunakan untuk melatih model statistik teks terjemahan dari bahasa sumber ke bahasa target. Saat melakukan penerjemahan bahasa, Moses membutuhkan korpus dalam dua bahasa, bahasa sumber dan bahasa target[8].

Secara umum, arsitektur mesin penerjemah statistik Moses ditunjukkan pada Gambar 1.



Gambar. 1. Arsitektur mesin penerjemah statistik Moses [9]

Sumber data utama yang dipergunakan adalah *parallel corpus* (korpus paralel) dan *monolingual corpus* (monolingual korpus). Proses *training* terhadap korpus paralel menggunakan GIZA++ menghasilkan *translation model* (TM). Proses *training* terhadap bahasa target pada korpus paralel ditambah dengan monolingual korpus bahasa target menggunakan SRILM menghasilkan *language model* (LM), sedangkan *PoS model* (PoS-M) dihasilkan dari bahasa target pada korpus paralel yang setiap katanya sudah ditandai dengan PoS. *TM*, *LM* dan *PoS-M* digunakan untuk menghasilkan *decoder* Moses.

Selanjutnya Moses digunakan sebagai mesin penerjemah untuk menghasilkan bahasa target dari *input* kalimat dalam bahasa sumber[10].

B. Case Folding

Case folding adalah tahapan mengubah huruf dalam dokumen menjadi huruf kecil. Karakter selain huruf dihilangkan dan dianggap pembatas (*delimiter*) [11].

C. Tokenizing

Tahap *tokenizing* adalah tahap pemotongan string *input* berdasarkan tiap kata yang menyusunnya [12].

D. Pivot Language

Pivot language adalah bahasa yang digunakan sebagai bahasa perantara terjemahan antara banyak bahasa. *Pivot language* digunakan sebagai bahasa perantara terjemahan antara bahasa yang berbeda. Untuk menerjemahkan antara setiap pasangan bahasa A dan B, salah satu diterjemahkan A ke bahasa pivot P, selanjutnya dari P ke B [13].

E. Bilingual Evaluation Understudy (BLEU)

BLEU adalah sebuah algoritma yang berfungsi untuk mengevaluasi kualitas dari sebuah hasil terjemahan yang telah diterjemahkan oleh mesin dari satu bahasa alami ke bahasa lain. Ide utama dibalik ini adalah “semakin dekat terjemahan sebuah mesin dengan terjemahan manusia, maka akan semakin baik[14].

BLEU mengukur *modified n-gram precision score* antara hasil terjemahan otomatis dengan terjemahan rujukan dan menggunakan konstanta yang dinamakan *brevity penalty* [15]:

$$BP_{BLEU} = f(x) = \begin{cases} 1, & \text{if } c > r \\ e^{(1-r/c)}, & \text{if } c \leq r \end{cases}$$

$$P_n = \frac{\sum_{c \in \text{corpus } n\text{-grams}} \text{count}_{clip}(n\text{-gram})}{\sum_{c \in \text{corpus } n\text{-grams}} \text{count}(n\text{-gram})}$$

$$BLEU = BP_{BLEU} \cdot e^{\sum_{n=1}^N w_n \log P_n}$$

Gambar. 2. Rumus yang digunakan dalam perhitungan nilai BLEU

Keterangan:

BP = *brevity penalty*

c = jumlah kata dari hasil terjemahan otomatis

r = jumlah kata rujukan

P_n = *modified precision score*

w_n = 1/N (standar nilai N untuk BLEU adalah 4)

p_n = jumlah *n-gram* hasil terjemahan yang sesuai dengan rujukan dibagi jumlah *n-gram* hasil terjemahan

II. PEMBAHASAN

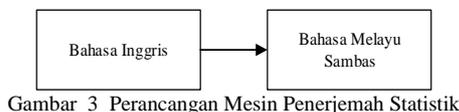
A. Data Penelitian

Data penelitian yang digunakan berupa berita *online* dua bahasa yaitu bahasa Indonesia dan bahasa Inggris dari *website* www.berita2bahasa.com. Bahasa Melayu Sambas diambil dari data bahasa Indonesia yang diterjemahkan ke bahasa Melayu Sambas. Data-data yang diperoleh tersebut selanjutnya diolah menjadi korpus teks paralel bahasa

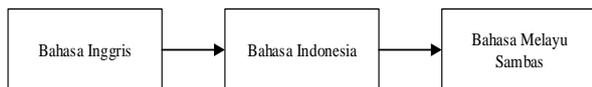
Inggris, bahasa Indonesia dan bahasa Melayu Sambas. Jumlah korpus paralel yaitu 3000 pasang kalimat.

B. Perancangan Mesin Penerjemah Statistik

Perancangan mesin penerjemah statistik yang akan dibuat dapat ditunjukkan pada Gambar 3 dan Gambar 4.



Gambar 3 Perancangan Mesin Penerjemah Statistik



Gambar 4 Perancangan Mesin Penerjemah Statistik Menggunakan Pivot Language

Pada penelitian ini terdapat tiga mesin penerjemah statistik yang akan dibuat. Tahap pertama akan dibuat mesin translasi bahasa Inggris ke bahasa Melayu Sambas seperti yang ditunjukkan pada Gambar 3. Tahap kedua menggunakan bahasa Indonesia sebagai *pivot language* akan dibuat dua mesin yaitu mesin translasi bahasa Inggris ke bahasa Indonesia dan mesin translasi bahasa Indonesia ke bahasa Melayu Sambas seperti yang ditunjukkan pada Gambar 4.

Pada penelitian ini digunakan 3000 kalimat yang dibagi atas 5 fold yang masing-masing berjumlah 600 kalimat seperti yang diperlihatkan pada Tabel 1. Untuk sebuah percobaan diambil 1 buah fold untuk *testing* dan 4 buah fold sisanya untuk *training*. Dengan kata lain, dari 3000 kalimat korpus paralel dibagi menjadi 600 kalimat untuk *testing* dan 2400 kalimat untuk *training* seperti yang terlihat pada Tabel 2. Sehingga untuk sebuah data set korpus paralel dilakukan 5 kali percobaan dan dari hasil kelima percobaan tersebut diambil nilai rata-ratanya.

TABEL I. PEMBAGIAN FOLD PADA KORPUS

Fold	Jumlah
P	1-600
Q	601-1200
R	1201-1800
S	1801-2400
T	2401-3000

TABEL II. JUMLAH MESIN DENGAN PEMBAGIAN FOLD

Mesin	Testing	Training
1	P	Q R S T
2	Q	P R S T
3	R	P Q S T
4	S	P Q R T
5	T	P Q R S

C. Implementasi Mesin Penerjemah Statistik Bahasa Inggris ke Bahasa Melayu Sambas

Model bahasa digunakan sebagai sumber pengetahuan berbasis teks dengan nilai-nilai probabilitistik. Penelitian ini menggunakan n-gram sebagai *language model*. Model Bahasa dibangun dengan *tool open source* SRILM.

Berdasarkan Gambar 5 model bahasa menghasilkan n-gram data yang terdiri dari n gram 1, n gram 2, n gram 3. Unigram (n gram 1) mempunyai data satu token, bigram (n gram 2) mempunyai data dua token dan trigram (n gram 3) mempunyai data tiga token dan masing-masing data dari n gram disertakan nilai probabilitasnya.

```

\data\
ngram 1=9832
ngram 2=40417
ngram 3=2526
\1-grams:
-4.75694      abisnye
-4.15488      acare

\2-grams:
-1.522103     ade juak
-1.850887     ade urang

\3-grams:
-1.248756     banyak urang kasal
-0.9383614    banyak urang yang
    
```

Gambar. 5 Model bahasa dengan bahasa Melayu Sambas sebagai bahasa target

Model translasi digunakan untuk memasangkan teks *input* dalam bahasa sumber dengan teks *output* dalam bahasa target. Model translasi dibangun dengan *tool open source* Giza++. Proses pemodelan translasi oleh Giza++ menghasilkan dokumen *vocabulary corpus* dan *word alignment*. Dokumen-dokumen tersebut terdapat 4 file yaitu "*corpus*, giza Inggris-Melayu Sambas, giza Melayu Sambas-Inggris dan model". Pemodelan translasi oleh Giza++ dapat dilihat pada Gambar 6 dan Gambar 7.

1	UNK	0
2	the	3637
3	,	2601
4	of	2024
5	to	1624

Gambar. 6 Dokumen *vocabulary corpus* bahasa Inggris

Angka 1 sampai 5 pada dokumen *vocabulary corpus*

1	UNK	0
2	,	2598
3	dan	1411
4	di	1366
5	yang	1258

Gambar. 7 Dokumen *vocabulary corpus* bahasa Melayu Sambas

merupakan uniq id untuk setiap data token, sedangkan angka disebelah kanan token menunjukkan frekuensi kemunculan. *Vocabulary corpus* yang dihasilkan mesin penerjemah bahasa Inggris-bahasa Melayu Sambas terdiri dari 9369 token untuk korpus bahasa Inggris dan 10522

token untuk bahasa Melayu Sambas. Dokumen *alignment* dapat dilihat pada Gambar 8.

```
# Sentence pair (111) source length 10 target length 13 alignment score :
1.19316e-12
as many as 224 passengers were rescued and 23 bodies have been recovered
NULL ({} sebanyak ({} 1 2 3 {} 224 ({} 4 {} penumpang ({} 5 {} berhasil ({}
diselamatkan ({} 6 7 {} dan ({} 8 {} 23 ({} 9 {} mayat ({} 10 {} udah ({} 11 {}
ditemukan ({} 12 13 {})
```

Gambar. 8 Dokumen *Alignment* bahasa Inggris – bahasa Melayu Sambas

Dokumen alignment terdapat tiga baris kalimat. Baris pertama berisi letak kalimat target (111) dalam korpus, panjang kalimat sumber (10), panjang kalimat target (13) dan nilai alignment (1.19316e-12). Baris kedua merupakan bahasa sumber dan baris ketiga merupakan alignment kalimat bahasa target terhadap kalimat bahasa sumber. Berdasarkan gambar tersebut kalimat bahasa target di-align ke kalimat bahasa sumber. Makna kata "penumpang" ({} 5 {}) pada kalimat bahasa target, di-align ke kata kelima bahasa sumber yaitu "passengers".

D. Pengujian Hasil Terjemahan

Pengujian hasil terjemahan dilakukan dengan pengujian otomatis BLEU. Mesin penerjemah statistik dibangun dengan dua tahap. Tahap pertama merupakan mesin penerjemah statistik bahasa Inggris ke bahasa Melayu Sambas. Tahap kedua merupakan mesin penerjemah menggunakan bahasa Indonesia sebagai pivot language. Pada tahap kedua ini dibangun dua mesin yaitu mesin penerjemah statistik bahasa Inggris ke bahasa Indonesia dan mesin penerjemah statistik bahasa Indonesia ke bahasa Melayu Sambas.

Setelah membuat output berupa hasil terjemahan otomatis dari mesin penerjemah statistik, langkah selanjutnya adalah mendapatkan nilai BLEU dari output dengan cara membandingkan output tersebut dengan korpus bahasa target yang telah dibuat sebelumnya. Hasil pengujian terjemahan otomatis dapat dilihat pada Tabel 3.

TABEL III.
HASIL PENGUJIAN BLEU

Mesin	MPS Inggris - Melayu Sambas Tanpa Pivot Language	MPS Inggris - Melayu Sambas dengan Pivot Language
1	21,16	22,55
2	23,09	24,68
3	24,79	25,04
4	15,84	16,8
5	14,41	15,2
Rata-rata	19,858	20,854

Berdasarkan Tabel 3 diperoleh nilai BLEU pada mesin penerjemah statistik Bahasa Inggris-bahasa Melayu Sambas dengan nilai rata-rata 19,858%. Nilai rata-rata BLEU pada mesin penerjemah statistik bahasa Inggris-bahasa Melayu Sambas dengan pivot language adalah 20,854%.

Hasil perhitungan penilaian otomatis terhadap hasil terjemahan seluruh kalimat uji pada mesin penerjemah

statistik bahasa Inggris – bahasa Melayu Sambas mengalami peningkatan nilai BLEU sebesar 5,015%.

E. Analisis Hasil Pengujian

Berikut adalah contoh hasil terjemahan mesin penerjemah statistik.

- Kalimat masukan : among the dead were two teenage girls
 - di antare yang tewas adalah duak remaja perempuan
 - para tewas ade di antare duak teenage perempuan
 - 12,98%
 - di antare korban tewas adalah duak teenage perempuan
 - 88,25%
- Kalimat masukan : president jokowi at the end of september 2015 has reviewed 500 hectares of planting land in karawang regency , west java
 - presiden jokowi pade akhir september 2015 telah meninjau lahan uji coba tanam seluas 500 hektar di kabupaten karawang , jawa barat
 - presiden jokowi pade akher september 2015 udah reviewed 500 hektar kabon tanam di karawang regency , jawa barat
 - 79,38%
 - presiden jokowi pade akher september 2015 udah reviewed 500 hektar tanam lahan di kabupaten karawang , jawa barat
 - 81,64%
- Kalimat masukan : this surplus condition that makes indonesia not import onion in 2016 , said amran sulaiman
 - keadaan surplus itok yang muat indonesia daan mengimpor bawang merah pade 2016 , nye amran sulaiman
 - itok surplus condition yang muat indonesia daan impor bawang merah pade 2016 , jinye amran sulaiman
 - 90,76%
 - itok surplus kondisi yang muat indonesia daan impor bawang merah pade 2016 , jinye amran sulaiman
 - 93,65%
- Kalimat masukan : two other hostages have been released earlier , and two are still detained
 - duak sandera laingnye dah kanak bebaskan sebalomnye , dan duakk maseh ditahan
 - duak laing sandera dah merilis awal , dan duak maseh ditahan
 - 72,47%
 - duak sandera laingnye dah merilis awal , dan duak maseh ditahan
 - 82,11%
- Kalimat masukan : he is believed to have organized smuggling 1,500 people into australia since 1999
 - die diyakinek dah mengorganisir penyelundupan 1.500 urang ke australia sejak 1999

- b) die itok diyakinek ngelakukan organized penyelundupan 1.500 urang ke australia sejak 1999
- c) 88,25%
- d) die itok diyakinek untok udah mengorganisir penyelundupan 1,500 urang ke australia sejak 1999
- e) 92,99%

Keterangan :

- a) : kalimat referensi
- b) : terjemahan MPS tanpa pivot language
- c) : skor BLEU MPS tanpa pivot language
- d) : terjemahan MPS dengan pivot language
- e) : skor BLEU MPS dengan pivot language

Nomor satu frase “among the dead” diterjemahkan oleh mesin penerjemah statistik tanpa *pivot language* menghasilkan output “para tewas ade di antare” dan skor BLEU sebesar 12,98%. Mesin penerjemah statistik dengan *pivot language* memperbaiki hasil terjemahan tersebut dengan menghasilkan skor BLEU sebesar 88,25% dan *output* yang dihasilkan “di antare korban tewas”. Kalimat nomor dua frase “in karawang regency, west java” diterjemahkan oleh mesin penerjemah statistik tanpa *pivot language* menghasilkan *output* “di karawang regency , jawa barat” dan skor BLEU sebesar 79,38%. Mesin penerjemah statistik dengan *pivot language* memperbaiki hasil terjemahan tersebut dengan menghasilkan skor BLEU sebesar 81,64% dan *output* yang dihasilkan “di kabupaten karawang , jawa barat”. Kalimat nomor tiga frase “this surplus condition” diterjemahkan oleh mesin penerjemah statistik tanpa *pivot language* menghasilkan *output* “itok surplus condition” dan skor BLEU sebesar 90,76%. Mesin penerjemah statistik dengan *pivot language* memperbaiki hasil terjemahan tersebut dengan menghasilkan skor BLEU sebesar 93,65% dan *output* yang dihasilkan “itok surplus kondisi”. Kalimat nomor empat frase “two other hostages” diterjemahkan oleh mesin penerjemah statistik tanpa *pivot language* menghasilkan *output* “duak laing sandera” dan skor BLEU sebesar 72,47%. Mesin penerjemah statistik dengan *pivot language* memperbaiki hasil terjemahan tersebut dengan menghasilkan skor BLEU sebesar 82,11% dan *output* yang dihasilkan “duak sandera laingnye”. Kalimat nomor lima frase “he is believed to have organized” diterjemahkan oleh mesin penerjemah statistik tanpa *pivot language* menghasilkan *output* “die itok diyakinek ngelakukan organized” dan skor BLEU sebesar 88,25%. Mesin penerjemah statistik dengan *pivot language* memperbaiki hasil terjemahan tersebut dengan menghasilkan skor BLEU sebesar 92,99% dan *output* yang dihasilkan “die itok diyakinek untok udah mengorganisir”.

Skor BLEU pada mesin penerjemah statistik dengan *pivot language* lebih tinggi dari pada skor BLEU pada mesin penerjemah statistik tanpa *pivot language* karena kalimat hasil terjemahan memiliki terjemahan yang mendekati dengan kalimat referensi. Hal ini disebabkan mesin penerjemah statistik menggunakan bahasa

Indonesia sebagai bahasa perantara atau *pivot language*, sehingga kualitas mesin penerjemah menjadi lebih baik.

III. KESIMPULAN

A. Kesimpulan

Berdasarkan hasil implementasi dan hasil analisis pengujian terhadap mesin penerjemah statistik bahasa Inggris – bahasa Melayu Sambas disimpulkan bahwa:

1. Mesin penerjemah statistik dapat diimplementasikan untuk menerjemahkan bahasa Inggris ke bahasa Melayu Sambas.
2. Terjadi peningkatan nilai akurasi mesin penerjemah statistik dengan metode BLEU yaitu sebesar 5,015%.
3. Penggunaan *pivot language* dapat mempengaruhi peningkatan akurasi penerjemahan pada mesin penerjemah statistik bahasa Inggris – bahasa Melayu Sambas.
4. Bahasa Indonesia sebagai *pivot language* dapat direkomendasikan sebagai bahasa perantara pada mesin penerjemah statistik bahasa Inggris (bahasa asing lainnya) ke bahasa daerah.

B. Saran

Saran yang dapat diberikan sebagai pengembangan dari penelitian ini adalah sebagai berikut.

1. Perlu adanya teknik penanganan *Out-of-Vocabulary* (OOV) untuk menerjemahkan kata-kata yang tidak terdapat dalam *vocabulary corpus* sehingga dapat meningkatkan kualitas terjemahan mesin penerjemah statistik.

REFERENSI

- [1] Utiyama, Masao., Isahara, Hitoshi. 2007. A Comparison of pivot Methods for Phrase-based Statistical Machine Translation. Jepang, Proceedings of NAACL HLT, pages 484–491.
- [2] Habash, Nizar., Hu, Jun. 2009. Improving Arabic-Chinese Statistical Machine Translation using English as Pivot Language. Proceedings of the Fourth Workshop on Statistical Machine Translation , pages 173–181.
- [3] Paul, Michael., Y, Hirofumi., S, Eiichiro., N,Satoshi. 2009. On the Importance of Pivot Language Selection for Statistical Machine Translation. Proceedings of NAACL HLT: Short Papers, pages 221–224.
- [4] Costa-jussa, Marta R., Henriquez, Carlos., Banchs, Rafael E. 2011. *Enhancing Scarce-Resource Language Translation Through Pivot Combinations*. Proceeding of the 5th International Joint Conference on Natural Language Processing, pages 1361-1265.
- [5] Zhao, Shiqi., W, Haifeng., Liu, Ting., Li, Sheng. 2008. Pivot Approach for Extracting Paraphrase Patterns from Bilingual Corpora. Proceedings of ACL-08: HLT, pages 780–788.
- [6] Hadi, Ibnu. 2014. *Uji Akurasi Mesin Penerjemah Statistik Bahasa Indonesia ke Bahasa Melayu Sambas dan Mesin Penerjemahan Statistik Bahasa Melayu Sambas ke Bahasa Indonesia*. Pontianak, JUSTIN Vol 3 No 1.
- [7] Manning, Christopher D. dan Schutze, Hinrich. 2000. *Foundations Of Statistical Natural Language Processing*. London : The MIT Press Cambridge Massachusetts.
- [8] Hasbiyansyah, Muhammad. 2016. Tuning for Quality Untuk Uji Akurasi Mesin Penerjemah Statistik (MPS) Bahasa Indonesia – Dayak Kanayatn. Pontianak , JEPIN Vol. 1 No. 1 2016.

- [9] Sujaini, Herry., Negara, Arif Bijaksana Putra. 2015. *Analysis of Extended Word Similarity Clustering based Algorithm on Cognate Language*. Gujarat: ESRSA Publications Pvt. Ltd.
- [10] Y.Jarob, H. Sujaini dan N. Safriadi. 2016. Uji Akurasi Penerjemahan Bahasa Indonesia – Dayak Taman dengan Penandaan Kata Dasar dan Imbuan. JEPIN, Vol. 2 No. 2.
- [11] Hasbiansyah , Muhammad. 2016. Tuning For Quality Untuk Uji Akurasi Mesin Penerjemah Statistik (MPS) Bahasa Indonesia - Bahasa Dayak Kanayatn. Pontianak, JEPIN Vol 1 No 1 2016.
- [12] Marlinda, Linda dan Rianto, Harsih. 2013. Pembelajaran Bahasa Indonesia Berbasis Web Menggunakan Metode Maximum Marginal Relevance. Manajemen Informatika. Jakarta: Sesindo
- [13] Triawati, Candra. 2009. Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia. Jakarta: IT TELKOM
- [14] Wu, Hua., Wang, Haifeng. 2007. *Pivot language Approach for Phrase-Based Statistical Machine Translation*. Proceedings of 45th Annual Meeting of the Association for Computational Linguistics, pages 856-863.
- [15] Mandira, Soni. 2016. Perbaikan Probabilitas Lexical Model untuk Meningkatkan Akurasi Mesin Penerjemah Statistik. Pontianak, JEPIN Vol 2 No 1 2016.