Mesin Penerjemah Situs Berita Online Bahasa Indonesia ke Bahasa Melayu Pontianak

Herry Sujaini

Staf Pengajar Program Studi Teknik Informatika, Jurusan Teknik Elektro,
Fakultas Teknik, Universitas Tanjungpura
e-mail: herry_sujaini@yahoo.com

Abstract– *Paper ini membahas salah satu potensi dari aplikasi mesin penerjemah, yaitu penerjemahan halaman situs.Halaman* situs *berbahasa Indonesia, diterjemahkan secara otomatis ke dalam bahasa Melayu Pontianak sehingga teks yang ada pada halaman sumber berubah menjadi teks dalam bahasa target.*
Cara kerja sistem ini adalah dengan mengambil seluruh halaman HTML dari sumber berbahasa Indonesia, selanjutnya memisahkan teks-teks yang berupa kalimat terhadap kode-kode HTML. Kalimat yang sudah dipisahkan selanjutnya diterjemahkan ke bahasa target (Melayu Pontianak). Kalimat hasil terjemahan ditampilkan pada halaman target dengan mengganti kalimat-kalimat pasangan terjemahannya.
Sistem ini mempergunakan mesin penerjemah berbasis statistik (MPS). Persoalan yang paling mendasar pada hasil kualitas terjemahan sistem ini adalah masih kecilnya kuantitas corpus. Sedangkan corpus merupakan merupakan data utama untuk membangun model - model yang digunakan pada MPS.

Keywords*- Penerjemahan halaman situs, mesin penerjemah statistik, bahasa Indonesia - Melayu Pontianak.*

1. Pendahuluan

Mesin penerjemah (MP) merupakan mesin yang dapat melakukan proses penerjemahan dari satu bahasa ke bahasa lainnya secara otomatis. MP memiliki kegunaan praktis sebab karena dapat membantu manusia untuk berkomunikasi satu sama lainnya yang memiliki bahasa yang berbeda. Masalah ini menjadi lebih penting pada era globalisasi sekarang ini, saat penerjemahan manual oleh manusia yang memiliki sumber daya terbatas dan mahal, MP memiliki potensi untuk dapat meningkatkan efisiensinya. Selain itu, media komunikasi seperti e-mail, sms, bbm, sosmed dan konferensi video saat ini telah menjadi semakin bervariasi dan serba instan dan telah menjadi bagian yang tidak terpisahkan dari aktivitas manusia.
Aktivitas penelitian pada bidang mesin penerjemah pertama kali dilakukan diGeorgetown University pada tahun1954. Mereka memiliki sasaran ideal ”terjemahan kualitas tinggi yang dilakukan sepenuhnya secara otomatis”(*fully auto matic high quality translation* (FAHQT)). Akan tetapi, proyek tersebut dinilai gagal oleh *Automatic Language Processing Advisory Committee* (ALPAC) sehingga para peneliti semakin realistis dan sadar dengan keterbatasan komputer sebagai alat penerjemah.
Salah satu pendekatan mesin penerjemah adalah dengan menggunakan pendekatan statistik yang menggunakan konsep probabilitas, biasanya disebut dengan mesin penerjemah statistik (MPS). Untuk setiap pasangan kalimat (*s*,*t*), diberikan sebuah *P(t|s)* yang diinterpretasikan sebagai distribusi probabilitas dimana MP akan menghasilkan *t* dalam bahasa tujuan saat diberikan *s* dalam bahasa sumber.
Mesin penerjemah secara luas telah dipergunakan dalam berbagai aplikasi, misalnya sebagai penerjemah umum multi bahasa seperti *Google Translator*, *Bing Translator* dan lain-lain. Pada paper ini, kami menyampaikan hasil rancang bangun mesin penerjemah halaman web secara otomatis dari bahasa Indonesia ke bahasa Melayu Pontianak.

Menurut Riza H. (2008), jumlah bahasa daerah yang dimiliki Indonesia mencapai 742 ragam yang menempatkan Indonesia pada urutan ke-2 sedunia sebagai laboratorium keanekaragaman bahasa setelah Papua Nugini dengan 867 ragam bahasa. Seiring dengan itu Darwis, M (2011) menjelaskan bahwa di Kalimantan, satu bahasa telah terancam punah dari lebih 50 bahasa daerah yang digunakan di pulau tersebut.Dari 13 bahasa daerah yang ada di Sumatera, satu diantaranya telah punah dan dua lainnya sudah terancam punah.Namun, di Jawa tidak ada bahasa daerah yang terancam punah. Adapun di Sulawesi, dari 110 bahasa yang ada, 36 bahasa terancam punah dan satu sudah punah, di Maluku, dari 80 bahasa yang ada 22 terancam punah dan 11 sudah punah, di daerah Timor, Flores, Bima dan Sumba 8 bahasa terancam punah dari 50 bahasa yang ada. Di daerah Papua dan Halmahera dari 271 bahasa, 56 bahasa terancam punah.Dikatakan lebih lanjut bahwa 9 bahasa dinyatakan telah punah di tanah Papua.208 bahasa terancam punah dan 32 bahasa segera punah.

Bahasa Melayu Pontianak merupakan salah satu bahasa yang dituturkan oleh masyarakat asli kota Pontianak. Penduduk asli kota Pontianak berdomisili di beberapa kecamatan yang letaknya berdekatan dengan Keraton Qadariah, dimana keraton tersebut merupakan pusat kerajaan Melayu di Pontianak. Di daerah-daerah tersebut bahasa Melayu Pontianak yang digunakan menggunakan kosakata bentuk asli bahasa Melayu Pontianak (Novianti E, 2011).

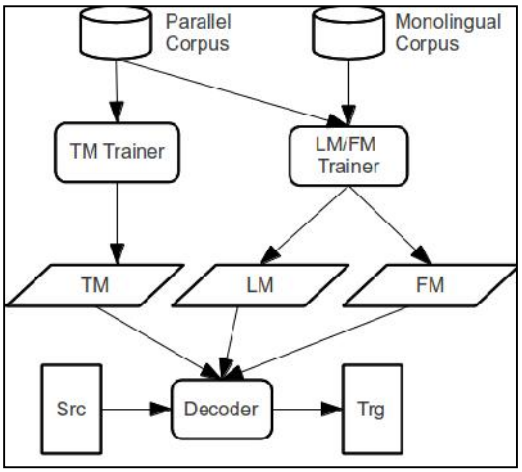
Berbagai cara dilakukan untuk melestarikan bahasa daerah, dari mulai memasukkannya ke dalam kurikulum sekolah, mengadakan seminar-seminar bahasa daerah, membuat dokumen-dokumen dalam bahasa daerah dan lain-lain. Salah satu cara yang belum banyak ditempuh adalah dengan membangun mesin penerjemah. Berdasarkan hal tersebut di atas, penelitian ini dilakukan untuk merancang dan mengimplementasikan mesin penerjemah situs berbahasa Indonesia ke salah satu bahasa daerah, yaitu bahasa Melayu Pontianak.

2. Mesin Penerjemah Statistik

Mesin penerjemah merupakan mesin yang melakukan penerjemahan secara otomatis, dimana sebuah komputer mengambil alih semua pekerjaan penerjemahan. Jelas,komputer akan bekerja lebih cepat dan lebih murah dari pada manusia. Pada dua dekade terakhir ini, terlihat bahwa penelitian pada bidang MP mengarah pada model penerjemahan yang dibangun secara otomatis dari *parallel corpus*. Model yang biasanya disebut mesin penerjemah statistik (MPS) ini menggunakan pendekatan teknik statistik.

Penelitian awal MPS dimulai oleh Brown dkk.(1993) dengan model berbasis kata yang sesuai dengan namanya, melakukan proses penerjemahan kata demi kata. Model ini sebagian besar telah diganti oleh model yang lebih kompleks, tetapi tetap digunakan sebagai basis untuk model lainnya seperti penyelarasan kata (*wordalignment*) (Al-Onaizan dkk., 2002). Zens dkk. (2002) yang kemudian diikutiKoehndkk. (2003) mengusulkan model berbasis frase yang menerjemahkan kalimat berdasarkan kata-kata yang berurutan dalam kalimat sumber untuk kata yang bersesuaian pada bahasa target. Ungkapan istilah frase dalam hal ini hanya berarti kata-kata yang berdekatan, bukan frase sebenarnya dalam istilah tata bahasa. Awalnya, model berbasis frase berakar dari hasil penelitian oleh Ochdan Weber (1998); Ochdkk. (1999);dan Ochdan Ney (2004). Penerjemahan dengan penggunaan frase juga telah diusulkan olehWang dan Waibel (1998); Venugopal dkk. (2003), danWatanabe dkk. (2003). Marcu (2001) mengusulkan penggunaan frase dalam model kata berbasis decoding. Sejalan dengan itu, penggunaan model *log-linier* diusulkan oleh Och dan Ney (2002).

Secara umum, arsitektur mesin penerjemah statistik seperti diperlihatkan pada Gambar 1. Sumber data utama yang dipergunakan adalah *parallel corpus* dan *monolingual corpus*. Proses training terhadap *parallelcorpus* menghasilkan *translationmodel* (TM). Proses training terhadap bahasa target pada *parallelcorpus* ditambah dengan *monolingualcorpus* bahasa target menghasilkan *language model* (LM), sedangkan *fitur model* (FM) dihasilkan dari bahasa target pada *parallelcorpus*yang setiap katanya sudah ditandai dengan fitur linguistik seperti Part of Speech (PoS), *lemma*, gender, proses pembentukan kata (morfem) dan lain-lain. TM, LM dan FM hasil proses di atas digunakan untuk menghasilkan *decoder*. Selanjutnya *decoder* digunakan sebagai mesin penerjemah untuk menghasilkan bahasa target dari input kalimat dalam bahasa sumber *(Sujaini dan Arif, 2014).*



Sumber : (Sujaini dan Arif, 2014)

Gambar 1. Arsitektur Mesin Penerjemah Statistik

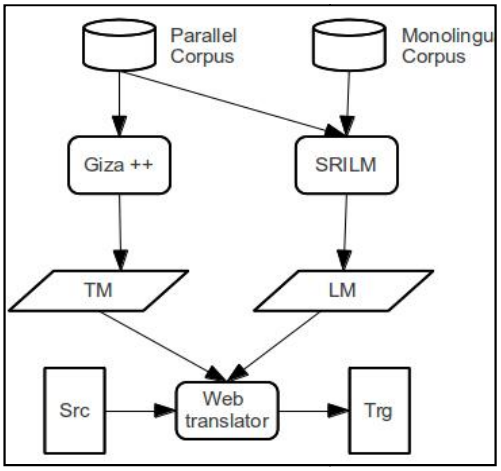
Jika dilihat dari arsitektur MPS,terlihat jelas bahwa bahan baku yang digunakan untuk menghasilkan model-model pada MPS adalah *parallel corpus*. *Monolingualcorpus* dapat diperoleh dari *parallelcorpus* pada sisi bahasa target walaupun biasanya diperbanyak lagi dari sumber-sumber lainnya.

3. Sistem Penerjemah Situs

Instrumen yang digunakan pada penelitian ini adalah SRILM (Stolcke, A. dkk., 2011) yang digunakan untuk membangun *language model*, Giza++(Och and Ney, 2003) yang digunakan untuk proses wordaligment dan membangun *translation model*, bahasa pemrograman PERL untuk membangun *decoder*, dan BLEU (Papineni dkk, 2002) untuk mengukur *modified n-gram precision score* antara hasil terjemahan otomatis dengan terjemahan rujukan dan menggunakan konstanta yang dinamakan *brevity penalty*.

Parallelcorpus yang digunakan pada penelitian ini adalah *parallelcorpus* Indonesia - Melayu Pontianak sebesar 4 K pasang kalimat bahasa Indonesia - Melayu Pontianak.

Arsitektur mesin penerjemah situs berbasis statistik yang digunakan pada penelitian ini seperti diperlihatkan pada Gambar 2. Sama seperti MPS pada umumnya, sumber data utama yang dipergunakan adalah *parallel corpus* dan *monolingual* yang diproses menggunakan GIZA++ dan SRILM untuk menghasilkan *translationmodel* (TM) dan *language model* (LM), sedangkan fitur model (FM) tidak digunakan pada sistem ini. Sumber input pada sistem ini berupa halaman situs berbahasa Indonesia, sedangkan target otuputnya berupa halaman situs berbahasa Melayu Pontianak.



Gambar 2. Arsitektur Mesin Penerjemah Situs

Setelah melakukan pra proses terhadap *corpus*, secara umum, proses-proses yang dilakukan pada penerjemah situs ini terdiri atas tiga langkah, yaitu (1) pengambilan kode HTML, (2) penerjemahan, dan (3) pengiriman kode HTML baru.

3.1. Pra Proses

Data utama dari sistem penerjemah situs ini adalah *parallel corpus* yang berisi data-data contoh kalimat dalam bahasa sumber beserta terjemahannya dalam kalimat target. Contoh *parallel corpus* yang digunakan pada sistem ini dapat dilihat pada gambar 3. Data yang lainnya adalah *mono corpus* yang berisi contoh kalimat-kalimat dalam bahasa target.

Pada sistem ini, digunakan Giza++ untuk proses word aligment dan membangun *translation model* yang berisi model translasi berbasis frase. Adapun potongan *translation model* dapat dilihat pada gambar 4. Selain itu sistem ini menggunakan SRILM untuk menghasilkan *languagemodel* (LM) yang terdiri dari unigram, dwigram dan trigram. Contoh trigram seperti terlihat pada gambar 5.

sudah lama juga wak cemuk mendanbakan posisi	udah lamak ugak wak cemuk mendanbekan
empuk , artinya tempat yang basah	posisi empuk , ertinya tempat nang basah
angan-angannya , akhirnya datang juga	angan-angan die , ahérnye tibe ugak
kata orang , ibarat buah durian jatuh dari langet	kate urang , ibarat buah durian jatok dari langet
rasanya , nasib suani ngah tinah ini akan berubah kulit	rasenya , nasèb lakinye ngah tinah nak berobah kulit
yang semula , keluargnya temasuk keluarga	nang semule , keluargnye temasok keluarge
dari tahapan pra sejahtera (pra ks) ,	dari tahapan pra sejahtera (pra ks) ,
tentulah akan jadi keluarga ks plus , artinya	tentulah nak jadi keluarge ks plus ,
tegolong orang yang kaya	ertinya tegolong urang nang kaye
" hei kawan , tentulah kau nanti akan necis ,	" hei bat , tentulah kau karang nak lawar ,
akan jadi manusia kalou sudah dilantik , "	nak jadi mausle kalok udah dilanték , "
sapa sahabat lamanya , mat kojeng , saat	tegor sahabat lamaknye , mat kojeng ,
bejunpa wak cemuk di warung kopi si aneng	ketike bejunpe wak cemuk di warung kuptinye
disapa seperti itu , wak cemuk hanya tesenyum-	si aneng
senyum. biasalah , kalou sudah dipuji ,	kenak sapa nacan tu , wak cemuk ganal-ganal
disanjung , apapun diberikannya	tesenyum-senyum. biaselah , kalok udah
sementara bang syokur sang penylar dangdut ,	kenak pujlan , kenak sanjong , ape jak
hanya bisa tesipu-sipu melihat tingkah orang	diberekannye
tua yang mau dilantik jadi orang	sementare bang syokur nang penylar
kebetulan mat kojeng dan bang syukur sedang	dangdut , ganal-ganal bise tesipu-sipu
menunggu si madon dan berahem , katenya ade	ningok péél urang tue nang maok dilanték
pekerjaan nau mengurus tanah kapling di pulau	jadi urang
hanyut	kebetolan mat kojeng ngan bang syukor
tapi , anehnya sapaan mat kojeng dijawab wak	tengah nunggu si madon ngan berahem ,
cemuk agak lain	katenye ade can maok ngoros tanah kapléng

Gambar 3. Potongan *Parallel Corpus*

```
" balas wak ||| " balas wak ||| 1 1 1 1 2.718 ||| ||| 1 1
" balas ||| " balas ||| 1 1 1 1 2.718 ||| ||| 1 1
" biasalah jeng ||| " biaselah jeng ||| 1 1 1 1 2.718 ||| ||| 1 1
" biasalah ||| " biaselah ||| 1 1 1 1 2.718 ||| ||| 1 1
" emang betul ||| " emang betol ||| 1 0.166667 1 1 2.718 ||| ||| 1 1
" emang ||| " emang ||| 1 0.333333 1 1 2.718 ||| ||| 1 1
" hei kawan ||| " hei bat ||| 1 1 1 1 2.718 ||| ||| 1 1
" hei ||| " hei ||| 1 1 1 1 2.718 ||| ||| 1 1
" kata wak ||| " kate wak ||| 1 0.833333 1 0.909091 2.718 ||| ||| 2 2
" kata ||| " kate ||| 1 0.833333 1 0.909091 2.718 ||| ||| 2 2
" katenya lagi ||| " katenye agik ||| 1 0.9 1 0.857143 2.718 ||| ||| 1 1
" katenya ||| " katenye ||| 1 1 1 1 2.718 ||| ||| 1 1
" pokoknya ada ||| " pokoknye ade ||| 1 0.342105 1 0.866667 2.718 ||| ||| 1 1
" pokoknya ||| " pokoknye ||| 0.5 0.5 1 1 2.718 ||| ||| 2 1
" pokoknye sudah ||| " pokoknye udah ||| 1 0.403226 1 0.714286 2.718 ||| ||| 1 1
" pokoknye ||| " pokoknye ||| 0.5 0.5 1 1 2.718 ||| ||| 2 1
" sana juga ||| " sane ugak ||| 1 0.5 1 0.8 2.718 ||| ||| 1 1
" sana ||| " sane ||| 1 0.5 1 0.8 2.718 ||| ||| 1 1
" sapa sahabat ||| " tegor sahabat ||| 1 1 1 1 2.718 ||| ||| 1 1
" sapa ||| " tegor ||| 1 1 1 1 2.718 ||| ||| 1 1
" tanya mat ||| " tanyak mat ||| 1 1 1 1 2.718 ||| ||| 1 1
" tanya ||| " tanyak ||| 1 1 1 1 2.718 ||| ||| 1 1
" ||| " ||| 1 1 1 1 2.718 ||| ||| 12 12
& ibas serahkan ||| ngan ibas ngasikkan ||| 1 0.015873 1 1 2.718 ||| ||| 1 1
& ibas ||| ngan ibas ||| 1 0.0238095 1 1 2.718 ||| ||| 1 1
```

Gambar 4. Potongan *Traslation Model*

```
\3-grams:
-1.053107       , " kate
-1.137602       <s> " pokoknye
-0.7010821      itu , die
-0.8575251      tangan , tadak
-0.5718167      sampai 6,1 persen
-0.04621392     awak ? </s>
-0.02878714     ke ? </s>
-0.7700987      <s> aek ini
-0.5130484      diubah agik </s>
-0.6983579      siape agik kalok
-0.4210106      jadi anggota dprd
-0.5734094      betombok antar budak
-0.871398       <s> ape name
-1.520732       <s> ape ndaknye
-1.066885       <s> awak tu
-0.8220639      <s> bal ini
-0.6390275      ngan bang syukor
-0.5754972      <s> betombok antar
-0.9481648      <s> budak itu
-0.5136191      antar budak sekolah
-1.154333       wak cemuk ,
-0.7081262      , die langsung
-1.379072       <s> die memang
-1.356042       <s> die suke
-0.5789095      maok dilantek jadi
```

Gambar 5. Potongan *Language Model 3-gram*

3.2. Pengambilan Kode HTML

Kode HTML diambil oleh sistem secara online ke alamat situs yang akan diterjemahkan. Contoh kode HTML dari halaman sumber dapat dilihat pada gambar 6. Jika diperhatikan, HTML potongan dari halaman situs tersebut terkandung kalimat-kalimat bahasa alami seperti : “Destinasi Wisata Sebagai Aset Daerah” dan “Istilah Pendapatan Asli Daerah (PAD) bagi sebagian orang bukan sesuatu yang asing. Beragam pembangunan bermula dari sini.”.Dua kalimat tersebut yang harus dipisahkan dan diterjemahkan ke dalam bahasa target. Proses tersebut dilakukan berulang terhadap semua kalimat bahasa alami yang terdapat dalam halaman situs tersebut.

```
<div class="box_4 h0">
    <div class="title">
        <a href="http://news.detik.com?880005471"><strong>Opini<span>Anda</span></strong></a>
        <span class="l_grey1"><a href="http://news.detik.com?880065471">Index &raquo;</a></span>
    </div>
    <ul class="list_berita_3">
        <li>
    <span class="date">Jumat, 30/05/2014 11:05 WIB</span>
    <h5 class="l_blue2_detik"><a href="http://news.detik.com?880065471">Destinasi Wisata Sebagai
        Aset Daerah</a></h5>
        Istilah Pendapatan Asli Daerah (PAD) bagi sebagian orang bukan sesuatu yang asing.
        Beragam pembangunan bermula dari sini.
    <div class="clearfix"></div>
</li>
    </ul>
</div>
</div> <div class="box_tengah fl">
    <div class="box_4 h0">
                    <div class="title">
                        <strong>Detik<span>Forum</span></strong>
```

Gambar 6. Potongan HTML Halaman Sumber

3.3. Proses Penerjemahan

Proses penerjemahan dilakukan dengan pendekatan penerjemahan berbasis frase, dimana kalimat yang akan diterjemahkan dibagi menjadi beberapa frase dan kemudian dibandingkan dengan data yang ada pada TM berdasarkan TM yang telah dipersiapkan pada pra proses. Hasilnya adalah beberapa kandidat hasil terjemahan. Kandidat-kandidat terjemahan tersebut, kemudian dinilai untuk mencari kalimat dengan nilai tertinggi dengan melibatkan TM dan LM. Sistem penilaian yang dilakukan menggunakan standar penilaian pada MPS yaitu dengan model log-linier.

3.4. Pengiriman Kode HTML

Tahapan selanjutnya adalah mengganti semua kalimat yang telah diterjemahkan dengan hasil terjemahannya masing-masing, kemudian mengirimkan halaman web baru yang berisi HTML lengkap dengan kalimat-kalimat yang telah diterjemahkan.

4.Hasil Pengujian dan Analisis

Pengujian sistem dilakukan terhadap sebuah situs berita dalam bahasa Indonesia yang cukup populer yaitu situs http://www.detik.com. Sedangkan sistem penerjemah web di-posting pada domain http://cammane.untan.ac.id dengan mengakses halaman http://cammane.untan.ac.id/ detikmelayu.pl.

Sebagai contoh halaman sumber dalam bahasa Indonesia dapat dilihat pada gambar 7. Halaman target dalam bahasa Melayu Pontianak, dapat dilihat pada gambar 8.

Gambar 7. Contoh Halaman Situs Sumber Berita

Sumber : http://news.detik.com/read/2015/01/29/213247/2818393/10/2-mobil-innova-yang-keluar-istana-masuk-rumah-megawati-jokowikah

Gambar 8. Contoh Halaman Hasil Terjemahan

Sumber : http://cammane.untan.ac.id/detikmelayu.pl?l=swen2015/01/29/213247/2818393/10/2-mobil-innova-yang-keluar-istana-masuk-rumah-megawati-jokowi-kah?nd772205mr

Akurasi hasil terjemahan sistem ini diukur dengan menggunakanmetodeBLEU.Pada eksperimen ini digunakan 300 kalimat yang dibagi atas 4 *fold* yaitu :*fold* 1 : kalimat no 1-75, *fold* 2 : kalimat no 76-150, *fold* 3 : kalimat no 151-225 dan *fold* 4 : kalimat no 226-300. Dari keempat *fold*, dibentuk 6 grup uji yang terdiri atas : Grup A (kalimat no 1-150), Grup B (kalimat no 1-75 dan 151-225), Grup C (kalimat no 1-75 dan 226-300), Grup D (kalimat no 76-225), Grup E (kalimat no 76-150 dan 226-300), dan Grup F (kalimat no 151-300). Hasil pengujian terhadap masing-masing grup uji dapat dilihat pada tabel 1.

Tabel 1. Nilai BLEU Pengujian Akurasi Terjemahan

Grup Uji	BLEU Score (%)
A	63,38
B	57,12
C	59,38
D	68,12
E	68,44
F	61,68

Dari tabel di atas, dapat dihitung rata-rata nilai BLEU yang merepresentasikan akurasi dari sistem penerjemah yaitu sebesar 63,02 %. Walau belum memuaskan, nilai ini bisa dikatakan cukup besar untuk penggunaan korpus yang relatif kecil. Hal ini dikarenakan bahasa Indonesia merupakan bahasa yang diadopsi dari bahasa Melayu sehingga tidak terlalu jauh perbedaannyadengan bahasa Melayu Pontianak.

5. Kesimpulan

Darihasil implementasi yang telah dilakukan terhadap sistem penerjemah situs dari bahasa Indonesia ke bahasa Melayu Pontianak, terlihat bahwa sistem telah dapat menghasilkan halaman situs baru yang persis sama dengan halaman situs aslinya, dengan kalimat-kalimat yang baru berbahasa Melayu Pontianak. Meskipun sistem telah dapat menghasilkan halaman situs terjemahan, namun hasil terjemahnnya sendiri masih belum mencapai tingkat akurasi yang memuaskan yaitu sebesar 63,02%. Hal tersebut dikarenakan kuantitas *corpus* yang masih kecil.

Akurasi hasil terjemahan tersebut dapat ditingkatkan dengan cara menambah kuantitas *corpus* yang memiliki kalimat-kalimat yang berkualitas. Selain itu juga dapat ditambahkan model-model fitur bahasa seperti *Part of Speech* (PoS), *lemma*,morfem dan lain-lain.

Saran yang dapat disampaikan antara lain agar dapat dilakukan penelitian yang lebih mendalam untuk meningkatkan akurasi terjemahan dengan penambahan jumlah *corpus* dan penambahan fitur-fitur linguistik. Selain itu juga perlu dilakukan penelitian terhadap bahasa-bahasa daerah lainnya.

Referensi

[1] Al-Onaizan, Y., Germann, U., Hermjakob, U., Knight, K., Koehn, P., Marcu, D. dan Yamada, K. (2002) : Translation with scarce bilingual resources. *Journal Machine Translation*. 17(1), 1 – 17.

[2] Brown, P. F., Pietra, V.J.D., Pietra, S.A.D., dan Mercer, R. L. (1993) :Themathematic sofstatistical machine translation.*Computational Linguistics*, 19(2), 263–313.

[3] Darwis, M (2011) :*Nasib Bahasa Daerah Di Era Globalisasi: Peluang Dan Tantangan*, Workshop Pelestarian Bahasa Daerah Bugis Makassar, Balitbang Agama , Parepare.

[4] Koehn, P., Och, F. J., dan Marcu, D. (2003) : Statistical phrase based translation. *Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL)*, Edmonton.

[5] Marcu, D. (2001) : Towards A Unified Approach To Memory And Statistical-Based Machine Translation, *Proceedings of the 39th Annual Meeting of the Association of Computational Linguistics (ACL*), Toulouse,378-385.

[6] Novianti,E.(2011):MenilikNasibBahasaMelayuPontianak, *InternationalSeminar "Language Maintenance and Shift"*,Semarang,70-74.

[7] Och, F. J., Tillmann, C., and Ney, H. (1999) : Improved Alignment Models For Statistical Machine Translation. *Proceedings of the Joint Conference of Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP-VLC)*, 20–28.

[8] Och, F. J. dan Ney, H. (2002) : Discriminative Training and Maximum Entropy Models for Statistical Machine Translation, *Proceedings of the 40th Annual Meeting of the Association of Computational Linguistics (ACL)*, Philadelphia, 295-302.

[9] Och, F. J. and Ney, H. (2003) :*A Systematic Comparison Of Various Statistical Alignment Models*, Journal Computational Linguistics, 29(1), 19-51.

[10] Och, F. J. dan Ney, H. (2004) :*The Alignment Template Approach to Statistical Machine Translation*, Journal Computational Linguistics, 30(4), 417–449.

[11] Och, F. J. danWeber, H. (1998) : Improving Statistical Natural Language Translation With Categories And Rules. *Proceedings of the 36th Annual Meeting of the Association of Computational Linguistics (ACL)*,Montreal.

[12] Papineni, K., Roukos, S., Ward, T., dan Zhu, W.-J. (2002) :BLEU: A Method For Automatic Evaluation of Machine Translation,*In Proceedings of the 40th Annual Meeting of the Association of Computational Linguistics (ACL)*, Pennsylvania, 311-318.

[13] RizaH.(2008):ResourcesReportonLanguagesofIndonesia, *The6thWorkshoponAsianLanguaeResources*.Hyderabad,93-94.

[14] Stolcke, A., Zheng, J., Wang, W., dan Abrash, V. (2011) :SRILM at Sixteen: Update and Outlook, *IEEE Automatic Speech Recognition and Understanding Workshop*, Waikoloa.

[15] Sujaini, H. dan Arif B.P.N. (2014) : Strategi Memperbaiki Kualitas Korpus untuk Meningkatkan Kualitas Mesin Penerjemah Statistik, *Seminar Nasional Teknologi Informasi XI*. Jakarta. 47-51.

[16] Venugopal, A., Vogel, S., dan Waibel, A. (2003) : Effective Phrase Translation Extraction From Alignment Models. Hinrichs, E. and Roth, D., editors, *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, 319–326.

[17] Wang, Y.Y. dan Waibel, A. (1998) : Modeling With Structures In Statistical Machine Translation. *Proceedings of the 36th Annual Meeting of the Association of Computational Linguistics (ACL),* Montreal, 1357-1363.

[18] Watanabe, T., Sumita, E., dan Okuno, H. G. (2003) : Chunk-Based Statistical Translation, *Proceedings of the 41st Annual Meetingon Association for Computational Linguistics*, Sapporo,Volume1,303-310.

[19] Zens, R., Och, F. J., dan Ney, H. (2002) :*Phrase-based statistical machine translation, Proceedings of the German Conference on Artificial Intelligence (KI 2002),* Heidelberg, 48-54.

Biografi

Herry Sujaini, lahir di Pontianak pada tanggal 29 Juni 1968. Memperoleh gelar ST dari Jurusan Teknik Elektro Universitas Tanjungpura tahun 1995. Kemudian memperoleh gelar MT dan Dr. dari Institut Teknologi Bandung tahun 2001dan 2014.Saat ini sebagai Staf Pengajar program studi Teknik Informatika Universitas Tanjungpura.