

Pengembangan Antarmuka Mesin Penerjemah Statistik Multibahasa Berbasis Web

Ninda Fitria Pratiwi^{#1}, Herry Sujaini^{#2}, Rudy Dwi Nyoto^{#3}

[#]Program Studi Teknik Informatika Universitas Tanjungpura

Jl. Prof Dr H. Hadari Nawawi, Kota Pontianak, 78115

¹nindafitria.NFP@gmail.com, ²herry_sujaini@yahoo.com, ³rudydn@gmail.com

Abstrak - Indonesia memiliki keragaman bahasa yang luar biasa. Akan tetapi, keterbatasan dalam menguasai berbagai macam bahasa daerah di Indonesia merupakan kendala utama bagi masyarakat umum untuk bisa mengerti bahasa daerah yang bukan berasal dari daerah asalnya. Dengan pesatnya perkembangan teknologi, saat ini sedang dikembangkan mesin penerjemah untuk mengatasi masalah penerjemahan bahasa. Mesin penerjemah statistik (*Statistical Machine Translation*) merupakan sebuah pendekatan mesin penerjemah dengan hasil terjemahan yang dihasilkan atas dasar model statistik yang parameter-parameternya diambil dari hasil analisis korpus teks paralel. Saat ini telah banyak dilakukan penelitian mengenai penerjemahan bahasa daerah ke bahasa Indonesia ataupun sebaliknya. Pada beberapa penelitian sebelumnya, mesin penerjemah statistik ada yang belum memiliki antarmuka dan ada yang memiliki antarmuka namun hanya penerjemahan satu bahasa saja. Tujuan penelitian ini adalah mengembangkan antarmuka mesin penerjemah statistik multibahasa dengan fitur komunitas terjemahan dimana pengguna dapat berkontribusi dalam aplikasi translator bahasa daerah sehingga dapat diakses secara luas melalui internet oleh publik. Pengujian dilakukan dengan metode *Black Box* dengan teknik *robustness testing* dan kuesioner. Berdasarkan hasil *robustness testing* menyatakan aplikasi ini dapat *handle* proses input dengan baik pada masing-masing bahasa. Berdasarkan hasil kuesioner menggunakan skala *Likert's Summated Rating (LSR)* diperoleh nilai total 1219 yang berarti aplikasi ini sangat positif dan dinilai berhasil.

Kata Kunci : *Mesin Penerjemah Statistik, Aplikasi Translator Bahasa Daerah, Komunitas Terjemahan.*

I. PENDAHULUAN

Bahasa merupakan alat komunikasi yang digunakan oleh seseorang untuk menyampaikan ide, gagasan, pengalaman dan sebagainya kepada orang lain. Bahasa memiliki peranan yang besar dalam kehidupan manusia. Bahasa Indonesia adalah bahasa resmi Republik Indonesia dan bahasa persatuan bangsa Indonesia. Wilayah Indonesia yang luas terbentang dari Sabang sampai Merauke memiliki keragaman bahasa dan budaya yang luar biasa. Menurut catatan The Linguist terdapat 746 bahasa yang tersebar di wilayah nusantara.

Dengan pesatnya perkembangan teknologi, saat ini sedang dikembangkan mesin penerjemah untuk mengatasi masalah penerjemahan bahasa. Mesin penerjemah (MP) merupakan mesin yang dapat melakukan proses penerjemahan dari satu bahasa ke bahasa lainnya secara otomatis. MP memiliki kegunaan praktis karena dapat membantu manusia untuk

berkomunikasi satu sama lainnya yang memiliki bahasa yang berbeda [1]. Mesin penerjemah statistik merupakan sebuah pendekatan mesin penerjemah dengan hasil terjemahan dihasilkan atas dasar model statistik yang parameter-parameternya diambil dari hasil analisis korpus teks bilingual (korpus paralel) [2].

Mesin penerjemah statistik saat ini seringkali digunakan oleh masyarakat dalam menerjemahkan suatu bahasa karena praktis untuk membantu masyarakat dalam berkomunikasi satu sama lainnya yang memiliki bahasa yang berbeda. Seperti yang kita ketahui mesin penerjemah yang umum digunakan sekarang ini adalah *Google Translate*. Tetapi bahasa yang terdapat pada *Google Translate* juga terbatas. Apabila masyarakat ingin menerjemahkan bahasa yang dimilikinya khususnya bahasa daerah, harus disiapkan mesin penerjemah khusus untuk bahasa tersebut. Salah satu sistem mesin penerjemah statistik yang sering digunakan adalah *Moses Decoder*. Saat ini telah dilakukan beberapa penelitian mengenai penerjemahan bahasa, seperti penelitian yang berjudul "Aplikasi Penerjemah Dua Arah Bahasa Indonesia - Bahasa Melayu Sambas Berbasis Web dengan Decoder Moses". Dalam penelitian tersebut hanya menguji satu bahasa saja, yaitu Bahasa Indonesia - Bahasa Melayu Sambas. Terdapat pula penelitian yang berjudul "Uji Akurasi Mesin Penerjemah Statistik Bahasa Indonesia ke Bahasa Melayu Sambas dan Mesin Penerjemahan Statistik Bahasa Melayu Sambas ke Bahasa Indonesia". Dalam penelitian tersebut hanya menguji nilai akurasi hasil terjemahan dari mesin yang dibangun, belum memiliki antarmuka dan belum dapat dipublikasikan melalui internet untuk diakses oleh publik.

Pada beberapa penelitian sebelumnya, mesin penerjemah statistik ada yang belum memiliki antarmuka dan ada yang memiliki antarmuka namun hanya penerjemahan satu bahasa saja. Untuk itu perlu dikembangkan antarmuka mesin penerjemah statistik bahasa Indonesia ke beberapa bahasa daerah atau sebaliknya, yang dibangun berbasis web agar dapat diakses oleh publik.

Berdasarkan masalah tersebut, penulis melakukan penelitian untuk mengembangkan antarmuka mesin penerjemah statistik multibahasa berbasis web dengan memanfaatkan *Moses Decoder* sebagai mesin penerjemah.

II. URAIAN PENELITIAN

A. Penerjemahan

Menurut Larson [3], penerjemahan adalah pengalihan makna dari bahasa sumber ke dalam bahasa sasaran. Selain itu, dalam penerjemahan terjadi penggantian bentuk bahasa sumber dengan bentuk bahasa sasaran. Dalam menerjemahkan bahasa sumber ke bahasa sasaran, bentuk boleh diubah tapi makna

harus dipertahankan. Larson menyatakan menerjemahkan berarti:

1. Mempelajari leksikon, struktur gramatikal, situasi komunikasi dan konteks budaya dari teks bahasa sumber.
2. Menganalisis teks bahasa sumber untuk menemukan maknanya.
3. Mengungkapkan kembali makna yang sama itu dengan menggunakan leksikon dan struktur gramatikal yang sesuai dalam bahasa sasaran dan konteks budayanya.

Menurut Nida dan Taber [4], penerjemahan adalah pengungkapan kembali pesan dari bahasa sumber di dalam bahasa sasaran dengan padanan terdekat dan wajar, pertama dalam hal makna dan kedua dalam hal gaya bahasa. Penerjemahan harus bertujuan utama pada pengungkapan kembali sebuah pesan. Untuk mengungkapkan kembali pesan dari teks sumber, seorang penerjemah harus membuat penyesuaian secara gramatikal dan leksikal dengan baik. Kesepadanan makna harus menjadi hal utama. Hal tersebut lebih merupakan penekanan sebuah reproduksi pesan daripada pemindahan bentuknya. Makna harus diutamakan dalam penerjemahan dan isi pesan merupakan keutamaan yang mendasar. Terjemahan yang paling baik adalah terjemahan yang tidak terlihat seperti terjemahan. Hal tersebut berarti bahwa terjemahan harus terasa sedekat mungkin dengan masyarakat bahasa sasaran. Penerjemahan harus menghasilkan padanan yang wajar sehingga tidak menunjukkan kekakuan dalam gramatikal dan gaya bahasa.

Berdasarkan uraian di atas disimpulkan bahwa pada dasarnya penerjemahan itu adalah mengalihkan pesan tulis yang ada dalam bahasa sumber ke bahasa target yang berbeda sesuai dengan isi pesan bahasa sumber, dan dengan menggunakan cara-cara pengungkapan atau pengekspresian yang wajar pada bahasa target. Terlepas dari perbedaan-perbedaan dalam hal pendefinisian istilah “penerjemahan” itu, para pakar sepakat atas prinsip dasar penerjemahan bahwa masalah makna merupakan hal pokok yang harus dipertimbangkan.

B. Mesin Penerjemah Statistik

Statistical Machine Translation (SMT) yang pertama diperkenalkan pada tahun 1991 oleh para ahli Thomas J. Watson Research Center - IBM. Dengan kemampuan komputer yang telah meningkat pesat pada sekitar tahun 90-an maka dimungkinkan untuk melakukan pengembangan dengan teknik statistik pada sistem komputer. Hasil riset tim IBM ini memberukan kontribusi yang signifikan dalam peningkatan ketertarikan dan optimisme para peneliti dan komunitasnya pada kemampuan mesin penerjemah dengan komputer. *Statistical Machine Translation* (SMT) adalah suatu paradigma dari mesin penerjemah dimana penerjemahan dilakukan berbasis model statistik dengan parameter-parameter yang diturunkan dari analisis *parallel corpus*. Pendekatan statistik berbeda dengan pendekatan berbasis aturan (*rule-based*) dan berbeda dari translasi berbasis contoh kalimat. Penerjemahan dengan metode SMT menghasilkan terjemahan yang lebih baik dibandingkan dengan hanya terjemahan kata demi kata (*interlinear translation*), dengan syarat yang dimasukkan ke dalam sistem mempunyai kualitas baik dan cukup banyak jumlahnya.

C. Unified Modelling Language (UML)

Unified Modeling Language adalah standar pemodelan dalam pembuatan aplikasi berorientasi objek yang diajukan oleh *Object Management Group* (OMG) pada tahun 1996.

1. Use Case Diagram

Use case diagram mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibuat. Dengan kata lain, use case diagram digunakan untuk mengetahui fungsi-fungsi apa saja yang terdapat di dalam sistem dan siapa saja yang berhak mengakses fungsi tersebut [5].

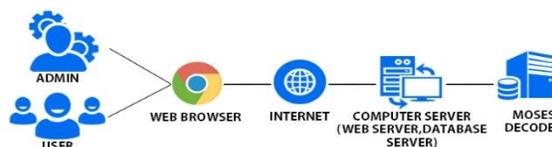
2. Activity Diagram

Activity diagram adalah sebuah diagram alur kerja yang menjelaskan berbagai kegiatan pengguna atau sistem, orang yang melakukan masing-masing aktivitas dan aliran sekuensial dari aktivitas-aktivitas tersebut [6].

III. PERANCANGAN SISTEM

A. Arsitektur Sistem

Desain arsitektur sistem akan ditunjukkan pada Gambar 1,

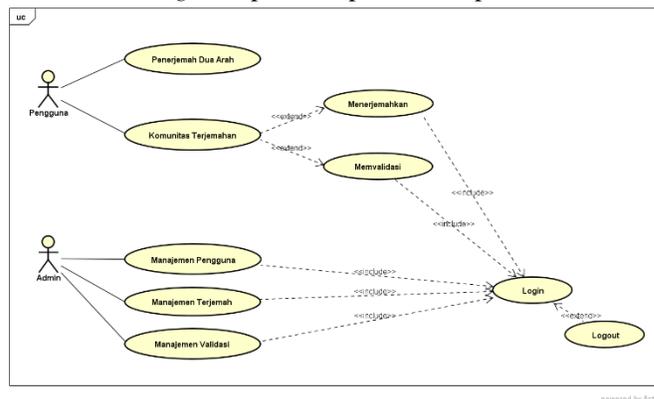


Gambar 1. Desain Arsitektur Sistem

Aplikasi yang dibangun berbasis web, yang dapat diakses oleh *user* atau pengguna untuk melakukan penerjemahan. Proses penerjemahan dimulai dengan *user* atau pengguna meminta terjemahan, *user* atau pengguna akan memasukkan masukan berupa kata, kalimat atau paragraf melalui *web server* dengan menggunakan aplikasi translator bahasa daerah pada *web browser* yang terhubung oleh internet, kemudian diteruskan ke *moses decoder* yang akan melakukan proses terjemahan. Hasil terjemahan yang dihasilkan oleh *moses decoder* akan dikembalikan lagi ke *user* atau pengguna melalui *web server*.

B. Use Case Diagram

Use case diagram aplikasi diperlihatkan pada Gambar 2,



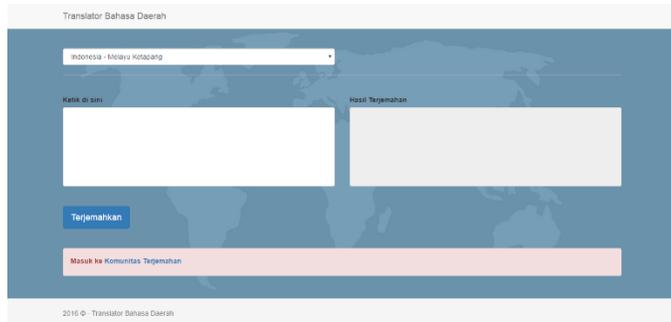
Gambar 2. Use Case Diagram Aplikasi

C. Pengujian Aplikasi

Pengujian dilakukan dengan menggunakan metode *Black Box* dengan teknik *robustness testing* dan kuesioner.

D. Hasil Aplikasi

Berikut beberapa tampilan hasil perancangan aplikasi, yang diperlihatkan pada Gambar 3 sampai dengan Gambar 6.



Gambar 3. Tampilan Penerjemah Dua Arah

Gambar 3 merupakan tampilan untuk menerjemahkan bahasa secara dua arah dari bahasa A ke bahasa B dan sebaliknya.



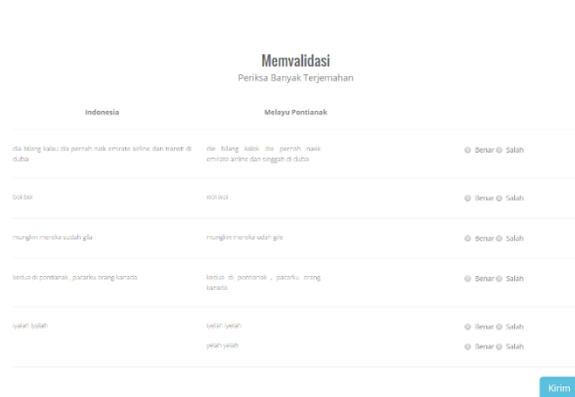
Gambar 4. Tampilan Menu Komunitas Terjemahan

Gambar 4 merupakan tampilan menu komunitas terjemahan, fitur dari aplikasi translator bahasa. Pengguna bisa berkontribusi dalam sistem penerjemah melalui komunitas terjemahan.



Gambar 5. Tampilan Menerjemahkan pada Komunitas Terjemahan

Gambar 5 merupakan tampilan fitur menerjemahkan pada komunitas terjemahan untuk menerjemahkan kata atau frasa layaknya pengguna mengucapkan atau menuliskannya.



Gambar 6. Tampilan Memvalidasi pada Komunitas Terjemahan

Gambar 6 merupakan tampilan fitur memvalidasi pada komunitas terjemahan untuk memperbaiki hasil terjemahan kata atau frasa dengan menandai terjemahan yang benar atau salah.

E. Hasil Pengujian

1. Robustness Testing

Pengujian dilakukan menggunakan metode *Black Box* dengan teknik *robustness testing* yang dilakukan pada aplikasi translator bahasa daerah dengan tujuan apakah aplikasi dapat mengambil hasil terjemahan yang dihasilkan pada moses decoder dan menampilkannya dengan baik dari bahasa-bahasa yang terdapat pada aplikasi. Pada Tabel 1, memperlihatkan pengujian pada penerjemahan bahasa Indonesia ke bahasa Melayu Ketapang dalam pengujian dilakukan dengan data kosong, satu kata, dua kata dan satu kalimat.

Tabel 1

Tabel Pengujian Penerjemahan Bahasa

Input	Contoh Data	Hasil Eksekusi	Keterangan
Data Kosong		Tidak Berhasil	Pesan kesalahan : Please fill out this field.
Satu Kata	Raja	Berhasil. Terjemahan : Raje	
Dua Kata	Raja muda	Berhasil. Terjemahan : Raje muda	
Satu Kalimat	Raja muda semakin cinta	Berhasil. Terjemahan : Raje mude semakin cinte	

Pada Tabel 2, memperlihatkan pengujian pada penerjemahan bahasa Indonesia ke bahasa daerah dalam pengujian dilakukan dengan satu kalimat dari bahasa Indonesia ke masing-masing bahasa daerah.

Tabel 2

Tabel Pengujian Penerjemahan Bahasa

Input	Contoh Data	Hasil Eksekusi	Keterangan
Satu Kalimat (Bahasa Indonesia-Bahasa Melayu Ketapang)	Tapi kamu cantik dan banyak yang suka sama kamu	Berhasil. Terjemahan : Tapi kau cante dan banyak yang suke same kau	

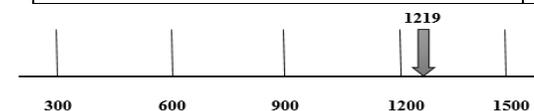
Satu Kalimat (Bahasa Indonesia-Bahasa Melayu Pontianak)	Tapi kamu cantik dan banyak yang suka sama kamu	Berhasil. Terjemahan : Tapi kau lawar dan banyak yang suke same kau
Satu Kalimat (Bahasa Indonesia-Bahasa Melayu Sambas)	Tapi kamu cantik dan banyak yang suka sama kamu	Berhasil. Terjemahan : Tapi kau canteck dan banyak nang sukke same kau
Satu Kalimat (Bahasa Indonesia-Bahasa Dayak Khanayatn)	Tapi kamu cantik dan banyak yang suka sama kamu	Berhasil. Terjemahan : Tapi kao miah man manyang nang suka sama kao
Satu Kalimat (Bahasa Indonesia-Bahasa Jawa Kromo)	Tapi kamu cantik dan banyak yang suka sama kamu	Berhasil. Terjemahan : Tapi panjenengan cantik kalih katah lingkang sami panjenengan ngremen

2. Kuesioner

Kuesioner berisi 10 pertanyaan yang digunakan dalam pengujian aplikasi. Kuesioner dibagikan kepada 30 responden yang merupakan mahasiswa dan mahasiswi Program Studi Teknik Informatika Fakultas Teknik Universitas Tanjungpura Pontianak. Untuk melihat skor terbesar dan terkecil dari satu orang responden dan total semua responden dari kuesioner ditunjukkan pada Tabel 3 berikut.

Tabel 3
Total Skor Responden Dari Kuesioner

Responden	Item										Total
	1	2	3	4	5	6	7	8	9	10	
1	4	3	4	3	3	3	3	3	3	3	32
2	3	4	4	4	3	5	5	3	3	3	37
3	4	4	3	2	3	4	4	4	3	3	34
4	5	4	4	5	4	4	4	5	4	4	43
5	4	4	4	4	5	5	4	3	3	4	40
6	3	4	4	4	4	3	4	3	4	4	37
7	3	4	3	4	5	4	4	3	4	4	38
8	4	5	5	4	4	4	5	4	4	4	43
9	4	4	4	5	4	4	4	5	4	5	43
10	4	5	5	4	4	4	5	4	5	4	44
11	4	3	4	3	5	3	5	3	4	4	38
12	4	4	5	3	3	5	5	2	3	4	38
13	4	4	4	3	4	5	5	4	3	3	39
14	3	4	3	4	4	4	4	2	4	4	36
15	5	5	5	5	3	3	5	5	3	4	43
16	5	5	5	5	5	5	4	5	5	5	49
17	4	4	4	4	3	3	4	4	3	4	37
18	4	4	4	4	4	4	5	3	4	4	40
19	5	5	5	5	5	3	2	5	3	4	42
20	4	4	3	4	3	5	3	3	4	4	37
21	4	5	5	4	5	4	5	4	4	4	44
22	5	4	3	5	5	3	5	3	4	5	42
23	4	5	5	4	5	4	5	4	4	4	44
24	4	5	5	4	4	4	4	5	5	4	44
25	4	5	4	4	3	4	4	5	4	4	41
26	4	5	4	4	3	4	4	5	4	4	41
27	4	5	5	4	4	4	5	4	4	4	43
28	4	5	5	5	5	4	4	4	3	3	42
29	4	5	5	5	5	4	4	4	4	4	44
30	4	4	5	5	5	4	5	4	4	4	44
Total Skor											1219



Gambar 7. Hasil Kuesioner pada Implementasi LSR

Gambar 7 menunjukkan bahwa hasil penelitian berada di antara skor 1200 dan 1500 yang artinya, hasil kuesioner menandakan responden menilai aplikasi sangat positif dan dinilai berhasil.

F. Analisis Hasil Pengujian

Rincian hasil analisis pengujian aplikasi translator bahasa daerah berbasis web yang telah dilakukan adalah sebagai berikut:

1. Hasil *robustness testing* menyatakan aplikasi dapat meng-handle proses input dengan baik pada masing-masing bahasa, terutama pada proses input data yang tidak sesuai dan kosong.
2. Hasil kuesioner menunjukkan sebagian besar responden menerima dengan baik aplikasi ini.
3. Dengan mengisi kuesioner tersebut, secara tidak langsung responden telah ikut berkontribusi dalam aplikasi dengan cara daftar dan login serta menggunakan fitur komunitas terjemahan.
4. Responden memberikan beberapa masukan untuk perbaikan aplikasi seperti tambahan fitur aplikasi dan perbaikan pada tampilan aplikasi.
5. Hasil interpretasi skor dari *Likert's Summated Rating (LSR)* menunjukkan hasil 1219 dengan nilai sangat positif yang berarti aplikasi dinilai berhasil.

IV. KESIMPULAN/RINGKASAN

Berdasarkan hasil implementasi dan hasil analisis pengujian terhadap aplikasi translator bahasa daerah dapat disimpulkan bahwa:

1. Pengembangan antarmuka mesin penerjemah statistik multibahasa dapat berjalan dengan baik dalam proses input pada masing-masing bahasa yang terdapat di aplikasi.
2. Berdasarkan hasil kuesioner sebagian besar responden menerima dengan baik aplikasi ini, dengan mengisi kuesioner tersebut, secara tidak langsung responden telah ikut berkontribusi dalam aplikasi dengan cara daftar dan login serta menggunakan fitur komunitas terjemahan.
3. Berdasarkan hasil kuesioner menggunakan skala *Likert's Summated Rating (LSR)* diperoleh nilai total 1219 yang berarti aplikasi ini sangat positif dan dinilai berhasil.

DAFTAR PUSTAKA

- [1] Sujaini, Herry. (2014). *Mesin Penerjemah Situs Berita Online Bahasa Indonesia ke bahasa Melayu Pontianak*. Jurnal ELKHA Vol. 6. No 2
- [2] Hidayat, Andri. (2015). *Aplikasi Penerjemah Dua Arah Bahasa Indonesia - Bahasa Melayu Sambas Berbasis Web Dengan Menggunakan Decoder Moses*. Pontianak: Universitas Tanjungpura
- [3] Larson, Mildred L. (1988). *Penerjemah Berdasar Makna: Pedoman untuk Pemadanan Antarbahasa. Terjemahkan Kencanawati Taniran*. Jakarta: Arcan
- [4] Nida, E. dan C. Taber. (1969). *The Theory and Practice of Translation*. Leiden: E.J Brill
- [5] Sukamto, Rosa Ariani dan M. Shalahudin. 2013. *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Informatika
- [6] Haviluddin. (2011). *Memahami Penggunaan Uml (Unified Modelling Language)*. Jurnal Informatika Mulawarman Vol 6