

## IMPLEMENTASI MESIN TRANSLASI INGGRIS-INDONESIA BERBASIS STATISTIK (UNTUK LITERATUR KARYA ILMIAH TEKNIK TELEKOMUNIKASI)

Lela Manarusana<sup>1</sup>, M.ramdhani<sup>2</sup>, Koredianto Usman<sup>3</sup>

<sup>1</sup>Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

### Abstrak

Penelitian tentang mesin translasi telah banyak dilakukan di Indonesia, mulai dari mesin translasi Inggris-Indonesia, Jawa-Indonesia, maupun China-English. Namun, sedikit sekali penelitian di bidang ini yang berdasarkan Ilmu statistik, padahal penelitian mesin translasi dengan metode statistik telah banyak dilakukan di luar negeri.

Dalam penelitian ini akan dilakukan implementasi mesin translator berbasis statistik untuk kalimat yang berasal dari literatur karya ilmiah teknik, khususnya teknik telekomunikasi. Penelitian ini dilakukan untuk mengetahui nilai akurasi dari sistem apabila dilakukan pengujian terhadap kalimat yang berasal dari literatur karya ilmiah.

Pengujian pada sistem dilakukan menggunakan dua corpus (kumpulan kalimat) yang berjumlah masing-masing 100 kalimat, dimana corpus pertama terdiri dari 100 kalimat yang telah dilatihkan terlebih dahulu. Dan corpus kedua terdiri dari 100 kalimat yang belum pernah dilatihkan. Setelah dilakukan pengujian, hasil terjemahan sistem akan dievaluasi dengan dua cara yaitu, secara otomatis menggunakan BLEU metrik dan secara manual oleh manusia. Setelah dilakukan analisis terhadap hasil evaluasi, bahwa mesin translator statistik dapat menerjemahkan kalimat yang berasal dari karya ilmiah teknik telekomunikasi dengan cukup baik. Dimana skor berdasarkan evaluasi otomatis untuk kalimat yang dilatih bernilai 0.9102 dan secara manual bernilai 0.754. Untuk kalimat yang dilatih memiliki skor yang tidak terlalu besar, hal ini dikarenakan kalimat yang diujikan merupakan kalimat yang diambil secara acak dari sumber yang berbeda dengan sumber kalimat yang dilatihkan.

**Kata Kunci :** Literatur karya ilmiah, Mesin translator

### Abstract

Research on translation machine has been done in Indonesia, from translation machine English-Bahasa, Javanese-Bahasa, to Chinese-English. But, very few researches is done in this field based on statistical science, although many researches on translation machine using statistical method has been done abroad.

In this research, it will be implemented a translation machine based on statistics for sentences from science engineering journal literature, especially telecommunication engineering. This research is done in order to know accuracy rate of the system if it will be tested on sentences from science journal literature.

Testing on system will be done using two corpuses (groups of sentence) which have 100 sentences each, where first corpus consists of 100 sentences that has not been trained before. After testing, system's output will be evaluated using two ways, that are automatically using BLEU metrics and manually by human.

After analization is done to evaluation results, that statistical translation machine can translate sentences from telecommunication engineering science journal well. The score from automatic evaluation for trained sentence has value of 0.9102 and manually 0.754. Trained sentence has not very large value, because tested sentences were randomly selected from different resources with the trained sentences.

**Keywords :** Science journal literature, translator machine

---

---

# BAB I

## PENDAHULUAN

### 1.1 LATAR BELAKANG

Bahasa dapat dibedakan menjadi bahasa alami dan bahasa buatan. Bahasa alami adalah bahasa yang digunakan sehari-hari untuk berkomunikasi, sedangkan bahasa buatan adalah bahasa yang dibuat khusus untuk tujuan tertentu.

Bahasa manusia yang berbeda-beda menyebabkan manusia mencoba untuk mengungkapkannya dengan berbagai cara. Salah satunya adalah dengan menggunakan komputer untuk menerjemahkan satu bahasa ke bahasa lainnya. Perangkat demikian dikenal sebagai "Mesin Penerjemah" atau "Mesin *Translasi*".

Dalam menerjemahkan kalimat dari satu bahasa ke bahasa yang lain, mesin translasi memiliki beberapa metode diantaranya adalah *example-based*, *case-based*, *rule-based*, dan *statistic-based*. Metode *example-based* yaitu penerjemah yang menghasilkan terjemahan terbaru mengacu kepada hasil terjemahan sebelumnya yang tersimpan di dalam basis data sistem.

Metode kedua adalah *case-based*, konsep dari sistem ini adalah menerjemahkan kalimat dengan memakai pola/*case* yang serupa dengan pola yang disimpan dalam *set of case*. Metode *rule-based*, aturan yang ada di dalam sistem dibuat oleh pembuat sistem berdasarkan pengetahuan linguistik. Sedangkan metode *statistic-based* merupakan metode sistem *translasi* berdasarkan ilmu statistika dengan menggunakan model pembelajaran dari sekumpulan kalimat (*corpus*) sebagai data latihnya.

Di Indonesia, Penelitian mesin *translasi* berbasis statistik belum banyak berkembang. Berbeda dengan di luar negeri penelitian mesin *translasi* berbasis statistik telah mulai banyak dikembangkan. Hal inilah yang mendasari dilakukannya penelitian mesin *translasi* berbasis statistik.

Kesulitan yang dialami mahasiswa dalam menerjemahkan *textbook* khususnya mahasiswa jurusan teknik menjadi bahan pertimbangan dalam

---

penyusunan tugas akhir ini. Hal tersebut yang mendasari digunakannya kumpulan kalimat yang berasal dari teks karya ilmiah teknik.

## 1.2 TUJUAN

Tujuan dari penelitian ini adalah Membangun mesin *translasi* Inggris-Indonesia berbasis statistik untuk literatur karya ilmiah teknik telekomunikasi, Menganalisa ketepatan mesin *translasi* Inggris-Indonesia berbasis statistik dalam menerjemahkan kata-kata atau kalimat dalam karya ilmiah teknik.

## 1.3 RUMUSAN MASALAH

Dalam tugas akhir ini akan dirumuskan beberapa masalah, yaitu;

1. Bagaimana pemodelan sistem *translator* berbasis statistik ini
2. Bagaimana kinerja mesin *translator* berbasis statistik apabila diberikan masukan berupa teks karya ilmiah teknik
3. Apakah mesin *translator* berbasis statistik ini cukup efektif untuk menerjemahkan teks karya ilmiah teknik.
4. Apa yang dapat disimpulkan dari analisa yang dilakukan pada mesin *translator* ini

## 1.4 BATASAN MASALAH

Pembahasan tugas akhir ini meliputi beberapa hal, yaitu;

1. Proses penerjemahan hanya dilakukan satu arah, yaitu dari bahasa Inggris ke bahasa Indonesia.
2. Jumlah kata atau kalimat yang tersedia dibatasi hanya beberapa kata atau kumpulan kata.
3. Kalimat yang dilatihkan dan diujikan hanya berasal dari literatur karya ilmiah teknik telekomunikasi, khususnya bidang transmisi
4. Tingkat ketepatan dalam menerjemahkan data uji dievaluasi oleh *software BLEU metric* dan *human*
5. Proses Pengujian dilakukan menggunakan *corpus* yang telah disusun, dengan asumsi teks yang digunakan merupakan terjemahan yang *valid*

- 
6. Tidak membahas sistem dan algoritma yang digunakan untuk melakukan *translasi*

## 1.5 METODE PENELITIAN

Metode penelitian yang akan dilakukan adalah sebagai berikut;

1. Studi literatur

Berisikan pembahasan teoritis melalui studi literatur dari buku-buku atau jurnal ilmiah yang berkaitan dengan permasalahan sistem *translator* berbasis statistik.

2. Pemodelan Sistem dan Persiapan Tool

*Tools* yang akan digunakan untuk model *translasi* adalah GIZA++, model bahasa adalah SRI LM dan *decoder* adalah Moses.

3. Penyusunan kalimat-kalimat yang akan dilatih dan diuji

4. Pengujian dan Analisis

Analisis dilakukan dari hasil yang diperoleh dari proses pengujian.

## 1.6 SISTEMATIKA PENULISAN

Secara umum Tugas Akhir ini dibagi menjadi lima bab bahasan, ditambah dengan lampiran dan daftar istilah yang diperlukan. Penjelasan masing-masing bab adalah sebagai berikut:

### **BAB I : Pendahuluan**

Bab I berisi latar belakang masalah, perumusan masalah, batasan masalah, tujuan penulisan, metode penelitian, dan sistematika penulisan yang memuat susunan penulisan Tugas Akhir.

### **BAB II : Teori Pendukung**

Pada Bab II ini berisi teori yang mendukung dan mendasari penulisan tugas akhir ini, yaitu teori tentang mesin *translator* berbasis statistik, teori probabilitas, pemodelan sistem, komponen penyusun mesin *translator* berbasis statistik.

### **BAB III : Perancangan dan Implementasi Sistem**

Pada Bab III ini berisi penjelasan tentang langkah-langkah perancangan model bahasa, model translasi dan penjelasan cara kerja sistem secara keseluruhan pada *software* yang sudah terintegrasi.

### **BAB IV : Pengujian dan Analisis Sistem**

---

Pada Bab IV ini berisi analisa terhadap hasil dari sistem yang telah dievaluasi oleh BLEU *metric* dan manusia.

### **BAB V : Penutup**

Pada Bab V ini berisi kesimpulan dari analisis yang telah dilakukan, serta rekomendasi atau saran untuk perbaikan dan pengembangan lebih lanjut



---

mesin translator statistik ini adalah jenis *alignment* kata *grow-diag-final*. Berbeda dengan penelitian sebelumnya yang menggunakan metode *alignment* kata *grow-diag-final-and* untuk masukan kalimat yang memiliki genre novel

- Hasil ketepatan untuk kalimat yang telah dilatihkan kepada sistem adalah sebesar 0.9102 untuk penilaian secara otomatis oleh *BLEU*
- Hasil ketepatan untuk kalimat yang telah dilatihkan adalah sebesar 0.754 untuk penilaian secara manual
- Sedangkan nilai ketepatan untuk kalimat yang belum pernah dilatihkan adalah 0.2663 untuk penilaian secara otomatis, dan 0.470 untuk penilaian secara manual
- Ketepatan yang terukur untuk kalimat *test* yang belum pernah dilatihkan secara signifikan berbeda jauh dengan yang telah dilatihkan, tetapi berdasarkan evaluasi yang dilakukan secara manual, hasil terjemahan sistem tidak terlalu mengurangi makna kalimat seharusnya

## SARAN

- Menggunakan lebih dari satu *human* dalam mengevaluasi hasil terjemahan secara manual
- Menggunakan lebih dari satu kalimat referensi dalam evaluasi secara otomatis
- Menggunakan kalimat uji yang lebih banyak
- Penerapan mesin translator statistik untuk aplikasi yang dapat langsung digunakan

## DAFTAR PUSTAKA

- [1] Axelrod, E.Ammittai.(2006). *Factored Language Models for Statistical Machine Translation*.Edinburgh, U.K: University of Edinburgh.
- [2] Callison-Burch, Chris.(2002). *Co-training for Statistical Machine Translation*.

---

mesin translator statistik ini adalah jenis *alignment* kata *grow-diag-final*. Berbeda dengan penelitian sebelumnya yang menggunakan metode *alignment* kata *grow-diag-final-and* untuk masukan kalimat yang memiliki genre novel

- Hasil ketepatan untuk kalimat yang telah dilatihkan kepada sistem adalah sebesar 0.9102 untuk penilaian secara otomatis oleh *BLEU*
- Hasil ketepatan untuk kalimat yang telah dilatihkan adalah sebesar 0.754 untuk penilaian secara manual
- Sedangkan nilai ketepatan untuk kalimat yang belum pernah dilatihkan adalah 0.2663 untuk penilaian secara otomatis, dan 0.470 untuk penilaian secara manual
- Ketepatan yang terukur untuk kalimat *test* yang belum pernah dilatihkan secara signifikan berbeda jauh dengan yang telah dilatihkan, tetapi berdasarkan evaluasi yang dilakukan secara manual, hasil terjemahan sistem tidak terlalu mengurangi makna kalimat seharusnya

## SARAN

- Menggunakan lebih dari satu *human* dalam mengevaluasi hasil terjemahan secara manual
- Menggunakan lebih dari satu kalimat referensi dalam evaluasi secara otomatis
- Menggunakan kalimat uji yang lebih banyak
- Penerapan mesin translator statistik untuk aplikasi yang dapat langsung digunakan

## DAFTAR PUSTAKA

- [1] Axelrod, E.Ammittai.(2006). *Factored Language Models for Statistical Machine Translation*.Edinburgh, U.K: University of Edinburgh.
- [2] Callison-Burch, Chris.(2002). *Co-training for Statistical Machine Translation*.

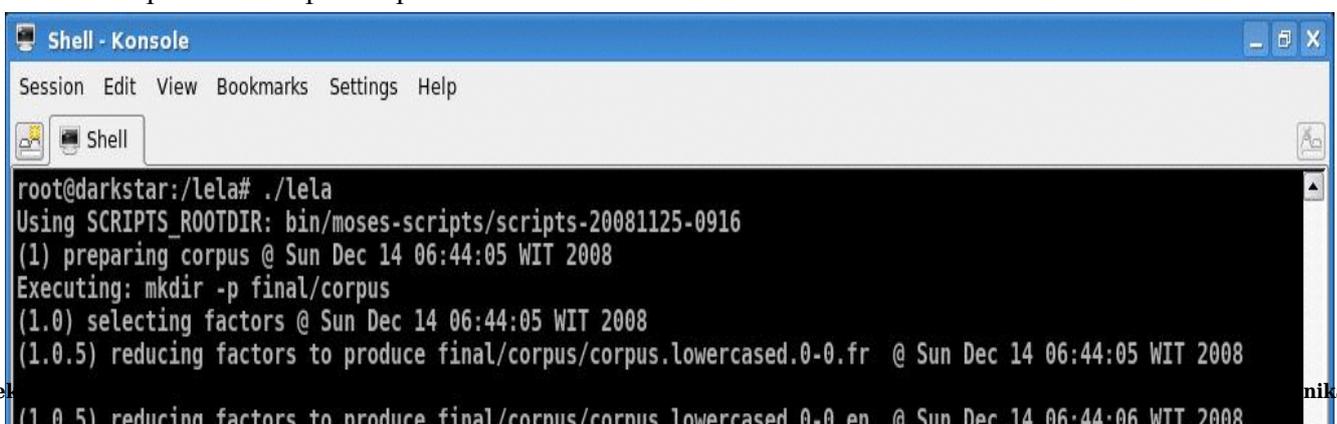
Edinburgh, U.K:University of Edinburgh.

- [3] Olteanu, M., Suriyentrakon, P., Moldovan, D.(2006). *Proceedings of the Workshop on Statistical Machine Translation*. pages 150-153. New York City:Association for Computational Linguistics.
- [4] Koehn, P., dkk. (2005). *Edinburgh System Description for the 2005 IWSLT Speech Translation Evaluation*. Scotland, U.K: University of Edinburgh.
- [5] Taryono, Amri. (2008). *Analisis Pengaruh Metode Word Alignment dan Reordering model pada Translator Berbasis Statistik Untuk Implementasi Translasi Bahasa Inggris ke Bahasa Indonesia*. Bandung:Institut Teknologi Telkom.
- [6] <http://www.iccs.inf.ed.ac.uk/~pkoehn/publications/tutorial2006.pdf>
- [7] <http://www.iccs.inf.ed.ac.uk/~pkoehn/publications/phrase2003.pdf>
- [8] <http://www1.cs.columbia.edu/nlp/sgd/bleu.pdf>
- [9] <http://www.speech.sri.com/projects/srilm/download.html>
- [10] <http://www.fjoch.com/GIZA++.html>
- [11] <http://www.statmt.org/moses>

Telkom  
University

## LAMPIRAN

### 1. Tampilan dalam proses pelatihan model translasi:



```
Shell - Konsole
Session Edit View Bookmarks Settings Help
Shell
root@darkstar:/lela# ./lela
Using SCRIPTS_ROOTDIR: bin/moses-scripts/scripts-20081125-0916
(1) preparing_corpus @ Sun Dec 14 06:44:05 WIT 2008
Executing: mkdir -p final/corpus
(1.0) selecting_factors @ Sun Dec 14 06:44:05 WIT 2008
(1.0.5) reducing_factors_to_produce_final/corpus/corpus.lowercased.0-0.fr @ Sun Dec 14 06:44:05 WIT 2008
(1.0.5) reducing_factors_to_produce_final/corpus/corpus.lowercased.0-0.en @ Sun Dec 14 06:44:06 WIT 2008
```