

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Saat ini sudah banyak sekali ditemukan sistem untuk melakukan translasi dari suatu bahasa ke dalam bahasa lainnya, seperti *Google Translate*, *Bing Translator*, dan lain-lain. Hal ini terjadi akibat semakin banyaknya informasi dan *resource* yang tersedia dalam berbagai bahasa sehingga mengakibatkan tuntutan akan pengalihan informasi menjadi sesuatu yang sangat penting dan diperlukan, terutama untuk bahasa Indonesia. Namun, sistem-sistem translasi tersebut menggunakan metode *crowd sourcing* yang memerlukan evaluasi dalam suatu domain tertentu. Metode *crowd sourcing* adalah suatu metode yang berupaya untuk memperbaiki hasil translasi dalam mesin pencari dengan melibatkan *user* dalam melakukan evaluasi. Melalui cara seperti ini dimungkinkan terdapat banyak *noise* atau kata-kata yang tidak sesuai dengan hasil translasi. Oleh karena itu, pada penelitian ini akan dibuat sebuah model yang berguna untuk mengubah bahasa Indonesia ke dalam bahasa Inggris dengan menggunakan *Statistical Machine Translation* (SMT) dan akan dianalisa berdasarkan data yang terdapat pada Alkitab. Alasan penggunaan Alkitab dalam penelitian ini adalah karena Alkitab mengandung ayat-ayat yang merupakan kata baku sehingga *resource* sumber dan *resource* tujuannya sudah diketahui dengan pasti.

SMT merupakan proses pengubahan (*translate*) suatu bahasa ke dalam bahasa lainnya dengan memanfaatkan probabilitas dari kata-kata dari bahasa sumber dengan bahasa tujuan. Hasil perbandingan probabilitas yang diperoleh akan dianalisis secara statistik dan menjadi acuan untuk mencari solusi translasi terbaik. Pada penelitian ini juga akan ditekankan mengenai IBM Model yang terdiri dari: (1) *lexical translation*, (2) *model of alignment* atau *reordering* model, (3) *fertility model*, (4) *relative reordering model*. IBM Model digunakan sebagai penentu performa dari model translasi. IBM Model digunakan oleh SMT, dimana SMT sangat cocok untuk melakukan translasi dengan memanfaatkan dan mempelajari *resource* sumber dan *resource* tujuan. (S, 2013) Model GIZA ini diharapkan dapat menghasilkan hasil translasi yang sesuai dengan faktor-faktor IBM Model dan mendapatkan nilai kecocokan padanan kata terbesar saat dilakukan evaluasi dengan menggunakan

algoritma BLEU (*Bilingual Evaluation Understudy*). Algoritma BLEU dianggap sebagai algoritma evaluasi yang cocok pada penelitian ini. (Papineni, Roukos, Ward, & Zhu, 2001) Hal ini dikarenakan pada penelitian ini akan dicari kaitan kecocokan antara padanan kata per kata, bukan konteks kalimat secara keseluruhan.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, masalah-masalah yang muncul dapat diidentifikasi sebagai berikut:

1. Faktor-faktor apa saja yang dapat menentukan performa dari model translasi?
2. Bagaimana cara melakukan pengujian terhadap model translasi yang telah dibuat?

## 1.3. Tujuan

Berdasarkan rumusan masalah yang ada, maka tujuan dari pembuatan laporan ini diantaranya:

1. Membuat model translasi dengan kata baku dan mempertimbangkan faktor-faktor yang dapat menentukan performa model translasi, seperti: *lexical translation*, *model of alignment* atau *reordering model*, *fertility model*, dan *relative reordering model*.
2. Membandingkan hasil translasi model GIZA dengan hasil translasi *Bing Translator* menggunakan algoritma BLEU untuk dievaluasi.

## 1.4. Batasan Masalah

Batasan dari tugas akhir ini adalah:

1. Data untuk pembentukan model translasi diperoleh dari Alkitab bahasa Indonesia dan bahasa Inggris.
2. Pengujian akan dilakukan pada 31 ayat Alkitab dari renungan harian e-RH (PSM) 1.2.1 bulan Juli 2014 dan edisi tahunan dari e-RH (PSM) 1.2.1 tahun 2010 untuk ekspansi eksperimen.
3. Batasan pembandingan untuk hasil translasi diambil dari hasil translasi situs [sabda.org](http://sabda.org).
4. Metode translasi yang digunakan adalah metode IBM model 1-4.

5. Mesin translasi yang akan digunakan sebagai pembanding adalah *Bing Translator*.
6. Evaluasi model akan dilakukan dengan cara membandingkan hasil translasi yang dibuat menggunakan model GIZA dengan hasil translasi dari *Bing Translator*, yaitu dengan mencari nilai kecocokan padanan kata terbesar dengan menggunakan algoritma BLEU.

### **1.5. Sistematika Penyajian**

Sistematika pembahasan dari penyusunan laporan ini adalah sebagai berikut:

#### **BAB I. Pendahuluan**

Bab ini berisi tentang pendahuluan yang terdiri dari latar belakang, rumusan masalah, tujuan, batasan masalah, dan sistematika penyajian laporan tugas akhir.

#### **BAB II. Landasan Teori**

Bab ini berisi tentang penjelasan teori-teori yang berkaitan dengan pembuatan dan pendukung perangkat lunak.

#### **BAB III. Analisis dan Desain**

Bab ini berisi tentang penjelasan analisis, gambaran arsitektur keseluruhan, dan disain perangkat lunak.

#### **BAB IV. Pengembangan Perangkat Lunak**

Bab ini berisi tentang penjelasan perencanaan tahap implementasi modul, penjelasan mengenai analisis dari algoritma yang digunakan, dan implementasi antarmuka.

#### **BAB V. Testing dan Evaluasi Sistem**

Bab ini berisi tentang penjelasan rencana pengujian perangkat lunak yang akan dilakukan.

#### **BAB VI. Simpulan dan Saran**

Bab ini berisi kesimpulan dari pembahasan pada perancangan perangkat lunak serta analisa pengujian aplikasi yang telah dibuat.