**IMPLEMENTASI MESIN PENERJEMAH STATISTIK**

**PADA APLIKASI *CHATTING* BERBASIS ANDROID**

**DENGAN MOSES DECODER**

**SKRIPSI**

Program Studi Sarjana Informatika

Jurusan Informatika

Oleh:

**MUHAMMAD NUR AMIN**

NIM D1041151011



**FAKULTAS TEKNIK**

**UNIVERSITAS TANJUNGPURA**

**PONTIANAK**

**2021**

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Nur Amin

NIM : D1041151011

menyatakan bahwa dalam skripsi yang berjudul “Implementasi Mesin Penerjemah Statistik Pada Aplikasi Chatting Berbasis Android Dengan Moses Decoder” tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi manapun. Sepanjang pengetahuan Saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Saya sanggup menerima konsekuensi akademis dan hukum di kemudian hari apabila pernyataan yang dibuat ini tidak benar.

Pontianak, 31 Mei 2021

Muhammad Nur Amin

NIM. D1041151017

HALAMAN PENGESAHAN

**IMPLEMENTASI MESIN PENERJEMAH STATISTIK PADA APLIKASI CHATTING BERBASIS ANDROID DENGAN MOSES DECODER**

Jurusan Informatika

Program Studi Sarjana Informatika

Oleh:

Muhammad Nur Amin

NIM. D1041151011

Telah dipertahankan di depan Penguji Skripsi pada tanggal 13 April 2021

dan diterima sebagai salah satu persyaratan untuk memperoleh gelar sarjana.

Susunan Penguji Skripsi :

Dosen Pembimbing Utama : Dr Arif Bijaksana PN, S.T., M.T. (NIP. 197208081998021002)

Dosen Pembimbing Kedua : Anggi Perwitasari, S.T., M.T. (NIP 198908192019032012)

Dosen Penguji Utama : Dr. Herry Sujaini, S.T., M.T. (NIP 196806291997021001)

Dosen Penguji Kedua : Helen Sasty Pratiwi, S.T., M.Eng. (NIP 198601172012122004)

|  |  |
| --- | --- |
| Pontianak, 20 Juni 2021  Dekan  Dr.rer.nat. Ir. R. M. Rustamaji, M.T.,IPU  NIP. 196801161994031003 | Wakil Dekan Bidang Akademik  Dr.-Ing. Ir. Slamet Widodo, M.T.,IPM  NIP. 196712231992031002 |

*Bismillahirrahmanirrahim*

*Skripsi ini didedikasikan kepada kedua orang tua saya yaitu bapak saya Muhammad Syamsuddin dan ibu saya Kartina Suud, serta adik saya yang tercinta Nurdiana. Apa yang mereka telah berikan kepada saya melebihi daripada apa yang pernah saya inginkan, Serta saya mengucapkan syukur sebesar-besarnya kepada Allah Subhanahu Wa Ta’ala” karena telah memberi saya kekuatan dan kesehatan dalam menyelesaikan skripsi ini. Ucapan terimakasih juga kepada teman-teman di prodi Informatika baik senior maupun junior saya yang sudah membantu dan memberikan dukungan dalam penyusunan tugas akhir ini.*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah Subhanahu Wa Ta’ala yang senantiasa memberikan limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian ini yang berjudul “Implementasi Mesin Penerjemah Statistik Pada Aplikasi *Chatting* Berbasis Android Dengan Moses Decoder”.

Dalam penyelesaian penelitian ini penulis secara langsung maupun tidak langsung telah mendapatkan bimbingan, pengarahan dan bantuan dari beberapa pihak. Melalui laporan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Dr. Arif Bijaksana Putra Negara, S.T., M.T. selaku dosen pembimbing utama dan Ibu Anggi Perwitasari, S.T., M.T. selaku dosen pembimbing pendamping. Bapak Dr. Herry Sujaini, S.T., M.T. selaku dosen penguji utama dan Ibu Helen Sasty Pratiwi, S.T., M.Eng. selaku dosen penguji pendamping. Bapak Helfi Nasution, S.Kom., M.Cs. selaku dosen pembimbing akademik dan Bapak Dr. Yus Sholva, S.T., M.T. selaku ketua jurusan Informatika. Terima kasih juga kepada seluruh dosen pengajar dan staf administrasi, serta rekan-rekan di program studi Informatika Universitas Tanjungpura.

Tentunya penulis sadari terdapat banyak kekurangan dalam penyusunan skripsi ini, oleh karena itu kritik dan saran sangat dibutuhkan demi kesempurnaan dalam penyusunan skripsi. Semoga skripsi ini bermanfaat bagi perkembangan ilmu pengetahuan.

Pontianak, 31 Mei 2021

Penulis

Muhammad Nur Amin

ABSTRAK

*Chatting* adalah salah satu bentuk komunikasi yang umumnya dilakukan antara dua orang atau lebih secara langsung (*realtime*) yang memanfaatkan fasilitas jaringan (LAN atau internet). Keberagaman bahasa yang luar biasa dan keterbatasan manusia dalam menguasai berbagai macam bahasa dapat menghambat tersampainya pesan atau informasi pada aktivitas *chatting*. Karakteristik bahasa *chatting* adalah menggunakan bahasa yang tidak baku sehingga muncul banyak sekali variasi dalam penuturan satu bahasa bergantung pada kondisi orang yang menuturkan. Tujuan dari penelitian ini adalah untuk membuat sebuah aplikasi *chatting* berbasis android yang mampu memperbaiki bahasa *chatting* ke bahasa baku dan menerjemahkan bahasa tersebutmenggunakan mesin penerjemah statistik Moses Decoder, sehingga dapat digunakan pada komunikasi teks antar pengguna bahasa yang berbeda. Metodologi penelitian yang dilakukan meliputi analisis permasalahan, perancangan sistem, implementasi, dan pengujian sistem. Pada penelitian ini dilakukan pra proses pesan bahasa Melayu Sambas *chatting* dan penerjemahan bahasa Melayu Sambas ke bahasa Indonesia. Pengujian pada penelitian ini berupa pengujian akurasi hasil pra proses HMM, akurasi hasil terjemahan mesin penerjemah Moses Decoder, dan pengujian fungsionalitas aplikasi. Hasil pengujian pra proses HMM memperoleh nilai tertinggi 0,9 (*accuracy*), 0,834 (*precision*), 0,829 (*recall*), dan 0,832 (*f-measure*). Hasil pengujian akurasi terjemahan mesin translasi dengan BLEU memperoleh nilai tertinggi 53,14% sedangkan oleh ahli bahasa memperoleh nilai 86,86% dan 87,71%. Hasil pengujian *black box* menunjukkan aplikasi dapat mengirim, menerima, memperbaiki dan menerjemahkan pesan. Secara keseluruhan penelitian dinilai berhasil dalam membuat sebuah aplikasi *chatting* berbasis android yang dapat memperbaiki bahasa *chatting*, mengimplementasikan mesin penerjemah statistik Moses Decoder dan dapat digunakan sebagai media komunikasi teks.

**Kata Kunci:** Aplikasi *Chatting*, Android, HMM, Moses Decoder, Mesin Penerjemah Statistik, Bahasa Indonesia, Bahasa Melayu Sambas.

ABSTRACT

Chatting is a form of communication that is generally carried out between two or more people directly (realtime) using network facilities (LAN or internet). The extraordinary diversity of languages ​​and human limitations in mastering various languages ​​can hinder the delivery of messages or information in chat activities. The characteristic of chat language is that it uses non-standard language, so that there are many variations in the speech of one language depending on the condition of the person speaking. The purpose of this research is to create an android-based chat application that is able to improve chat language to standard language and translate the language using the Moses Decoder statistical translator engine, so that it can be used in text communication between users of different languages . The research methodology includes problem analysis, system design, implementation, and system testing. In this study, pre-processing of messages in Malay Sambas chat and translation of Malay Sambas into Indonesian was carried out. The tests in this study were in the form of testing the accuracy of the pre-processed HMM results, the accuracy of the translation results of the Moses Decoder machine translator, and testing the application functionality. The results of the pre-process HMM test obtained the highest values ​​of 0.9 (accuracy), 0.834 (precision), 0.829 (recall), and 0.832 (f-measure). The results of testing the accuracy of translation machine translation with BLEU obtained the highest score of 53.14%, while by linguists the score was 86.86% and 87.71%, respectively. The black box test results show that the application can send, receive, correct and translate messages. Overall the research is considered successful in making an android-based chat application that can improve chat language, implement the Moses Decoder statistical translation engine and can be used as a medium of text communication.

**Keywords:** Chat Application, Android, HMM, Moses Decoder, Statistical Translation Machine, Indonesian, Melayu Sambas.

DAFTAR ISI

[HALAMAN PERNYATAAN ii](#_Toc66975498)

[HALAMAN PERSETUJUAN iii](#_Toc66975499)

[KATA PENGANTAR v](#_Toc66975500)

[ABSTRAK vi](#_Toc66975501)

[ABSTRACT vii](#_Toc66975502)

[DAFTAR ISI viii](#_Toc66975503)

[DAFTAR TABEL xi](#_Toc66975504)

[DAFTAR GAMBAR xii](#_Toc66975505)

[DAFTAR KODE PROGRAM xiv](#_Toc66975506)

[DAFTAR LAMPIRAN xv](#_Toc66975507)

[BAB I PENDAHULUAN 1](#_Toc66975508)

[1.1 Latar Belakang 1](#_Toc66975509)

[1.2 Perumusan Masalah 3](#_Toc66975510)

[1.3 Tujuan Penelitian 3](#_Toc66975511)

[1.4 Pembatasan Masalah 3](#_Toc66975512)

[1.5 Sistematika Penulisan 3](#_Toc66975513)

[Bab II TINJAUAN PUSTAKA 5](#_Toc66975514)

[2.1 Kajian Terkait 5](#_Toc66975515)

[2.2 *Chatting* 8](#_Toc66975516)

[2.3 Normalisasi Teks 9](#_Toc66975517)

[2.4 Part of Speech (POS) Tagging 10](#_Toc66975518)

[2.5 Natural Language Toolkit (NLTK) 11](#_Toc66975519)

[2.6 Mesin Penerjemah Statistik 11](#_Toc66975520)

[2.7 Moses Decoder 12](#_Toc66975521)

[2.8 Korpus 13](#_Toc66975522)

[2.9 *Unified Modelling Language* (UML) 13](#_Toc66975523)

[2.10 Android 14](#_Toc66975524)

[2.10.1 Sistem Operasi Android 14](#_Toc66975525)

[2.10.2 Android Studio 14](#_Toc66975526)

[2.11 Java 16](#_Toc66975527)

[2.12 HyperText Preprocessor (PHP) 16](#_Toc66975528)

[2.13 Firebase 17](#_Toc66975529)

[2.14 K-Fold Cross Validation 17](#_Toc66975530)

[2.15 Pengujian 18](#_Toc66975531)

[2.15.1 Pengujian Hasil Pra Proses HMM 19](#_Toc66975532)

[2.15.2 Pengujian Hasil Penerjemahan Pesan 19](#_Toc66975533)

[2.15.3 Pengujian Fungsionalitas Aplikasi 20](#_Toc66975534)

[Bab III METODOLOGI PENELITIAN 21](#_Toc66975535)

[3.1 Data dan Alat Bantu Penelitian 21](#_Toc66975536)

[3.1.1 Data Penelitian 21](#_Toc66975537)

[3.1.2 Alat Bantu Penelitian 21](#_Toc66975538)

[3.2 Metodologi Penelitian 24](#_Toc66975539)

[3.3 Analisis Permasalahan 25](#_Toc66975540)

[3.3.1 Analisis Data dan Fitur WhatsApp 26](#_Toc66975541)

[3.3.2 Analisis Bahasa Melayu Sambas pada   
Aktivitas *Chatting* 33](#_Toc66975542)

[3.4 Perancangan Sistem 37](#_Toc66975543)

[3.4.1 Perancangan Arsitektur Sistem 38](#_Toc66975544)

[3.4.2 Unified Modelling Language (UML) 49](#_Toc66975545)

[3.4.3 Perancangan *Layout* Antarmuka Aplikasi 58](#_Toc66975546)

[3.5 Implementasi 64](#_Toc66975547)

[3.5.1 Pembuatan Server Penerjemah 64](#_Toc66975548)

[3.6 Pengujian Sistem 69](#_Toc66975549)

[3.6.1 Pengujian *Accuracy*, *Precision*, *Recall* dan   
*F-Measure* Hasil Pra Proses HMM dengan Metode   
*K-Fold Cross Validation* 69](#_Toc66975550)

[3.6.2 Pengujian Hasil Penerjemahan Mesin Translasi   
oleh BLEU dengan Metode *K-Fold Cross Validation* 70](#_Toc66975551)

[3.6.3 Pengujian Hasil Penerjemahan Mesin Translasi   
oleh Ahli Bahasa 70](#_Toc66975552)

[3.6.4 Pengujian *Black Box* 71](#_Toc66975553)

[Bab IV HASIL IMPLEMENTASI DAN PENGUJIAN 73](#_Toc66975554)

[4.1 Hasil Implementasi Aplikasi 73](#_Toc66975555)

[4.1.1 Halaman Verifikasi Nomor Telepon 75](#_Toc66975556)

[4.1.2 Halaman Verifikasi OTP 76](#_Toc66975557)

[4.1.3 Halaman *Update* Profil 76](#_Toc66975558)

[4.1.4 Halaman Daftar Obrolan 77](#_Toc66975559)

[4.1.5 Halaman Daftar Kontak 78](#_Toc66975560)

[4.1.6 Halaman Ruang Obrolan 79](#_Toc66975561)

[4.1.7 Halaman Profil Saya 80](#_Toc66975562)

[4.1.8 Halaman Profil Teman 81](#_Toc66975563)

[4.2 Hasil Pengujian 82](#_Toc66975564)

[4.2.1 Pengujian *Accuracy*, *Precision*, *Recall* dan   
*F-measure* Pra Proses HMM 83](#_Toc66975565)

[4.2.2 Pengujian Hasil Penerjemahan Mesin Translasi   
oleh BLEU 84](#_Toc66975566)

[4.2.3 Pengujian Hasil Penerjemahan Mesin Translasi   
oleh Ahli Bahasa 88](#_Toc66975567)

[4.2.4 Pengujian *Black Box* 90](#_Toc66975568)

[4.3 Analisis Hasil Perancangan dan Pengujian 93](#_Toc66975569)

[Bab V PENUTUP 95](#_Toc66975570)

[5.1 Kesimpulan 95](#_Toc66975571)

[5.2 Saran 95](#_Toc66975572)

[DAFTAR PUSTAKA 97](#_Toc66975573)

DAFTAR TABEL

[Tabel 2.1 Tabel Perbandingan Penelitian 6](#_Toc62757235)

[Tabel 3.1 Jenis Pesan pada Aplikasi WhatsApp 28](#_Toc66092270)

[Tabel 3.2 Tabel Penggunaan Abreviasi 34](#_Toc66092271)

[Tabel 3.3 Tabel Penggunaan Slang 35](#_Toc66092272)

[Tabel 3.4 Tabel Penggunaan Pemanjangan Kata 37](#_Toc66092273)

[Tabel 3.5 Definisi Aktor 50](#_Toc66092274)

[Tabel 3.6 Definisi *Use Case* 50](#_Toc66092275)

[Tabel 3.7 Hasil Pengujian Black Box 71](#_Toc66092276)

[Tabel 4.1 Hasil Implementasi Rancangan *Class Diagram* 73](#_Toc75524555)

[Tabel 4.2 Hasil PengujianPra Proses HMM Skenario Pertama 84](#_Toc75524556)

[Tabel 4.3 Hasil PengujianPra Proses HMM Skenario Kedua 85](#_Toc75524557)

[Tabel 4.4 Hasil Pengujian BLEU Penerjemahan Bahasa Melayu Sambas   
(*chatting*) ke Bahasa Indonesia 87](#_Toc75524558)

[Tabel 4.5 Hasil Pengujian BLEU Penerjemahan Bahasa Melayu Sambas   
(Sudah Diperbaiki) ke Bahasa Indonesia 87](#_Toc75524559)

[Tabel 4.7 Hasil Pengujian BLEU Dengan Metode K-Fold Cross Validation   
Pada Fold 3 (450 Kalimat) 89](#_Toc75524560)

[Tabel 4.8 Hasil Pengujian Terjemahan Mesin Penerjemah Oleh Ahli Bahasa 89](#_Toc75524561)

[Tabel 4.9 Hasil Pengujian Proses *Login* 90](#_Toc75524562)

[Tabel 4.10 Hasil Pengujian Proses Memperbarui Profil 91](#_Toc75524563)

[Tabel 4.11 Hasil Pengujian Pengiriman Pesan 92](#_Toc75524564)

DAFTAR GAMBAR

[Gambar 2.1 Komponen Mesin Penerjemah Statistik 12](#_Toc66092341)

[Gambar 2.2 Contoh Skema 10-*Fold Cross Validation* 18](#_Toc66092342)

[Gambar 3.1 Metodologi Penelitian 24](#_Toc75524627)

[Gambar 3.2 Contoh Data Hasil Ekspor Percakapan Aplikasi Whatsapp 32](#_Toc75524628)

[Gambar 3.3 Arsitektur Sistem 38](#_Toc75524629)

[Gambar 3.4 Dokumen Teks *Chatting* WhatsApp 39](#_Toc75524630)

[Gambar 3.5 Korpus Teks *Chatting* 40](#_Toc75524631)

[Gambar 3.6 Korpus Bahasa Melayu Sambas 40](#_Toc75524632)

[Gambar 3.7 Korpus *Tagging* 41](#_Toc75524633)

[Gambar 3.8 Korpus Bahasa Indonesia 41](#_Toc75524634)

[Gambar 3.9 Alur Pembuatan Mesin HMM 43](#_Toc75524635)

[Gambar 3.10 Perbandingan Korpus 44](#_Toc75524636)

[Gambar 3.11 Isi Korpus *Tagging* Pra Proses Skenario Pertama 45](#_Toc75524637)

[Gambar 3.12 Isi Korpus *Tagging* Pra Proses Skenario Kedua 45](#_Toc75524638)

[Gambar 3.13 Proses Penerjemahan Pesan 46](#_Toc75524639)

[Gambar 3.14 Arsitektur Mesin Penerjemah Statistik 47](#_Toc75524640)

[Gambar 3.15 Proses Penerjemahan oleh Moses 48](#_Toc75524641)

[Gambar 3.16 *Use Case* Diagram Aplikasi *Chatting* 49](#_Toc75524642)

[Gambar 3.17 *Class Diagram* Aplikasi *Chatting* 52](#_Toc75524643)

[Gambar 3.18 *State Diagram User* 53](#_Toc75524644)

[Gambar 3.19 *State Diagram Chat* 54](#_Toc75524645)

[Gambar 3.20 *State Diagram Contact* 55](#_Toc75524646)

[Gambar 3.21 *Sequence Diagram Login* 56](#_Toc75524647)

[Gambar 3.22 *Sequence Diagram* Memuat *Chat* 56](#_Toc75524648)

[Gambar 3.23 *Sequence Diagram* Mengirim *Chat* 57](#_Toc75524649)

[Gambar 3.24 *Sequence Diagram* Memuat Kontak 57](#_Toc75524650)

[Gambar 3.25 *Sequence Diagram* Profil Teman 58](#_Toc75524651)

[Gambar 3.26 *Sequence Diagram* Profil 58](#_Toc75524652)

[Gambar 3.27 Perancangan Halaman Verifikasi Nomor Telepon 59](#_Toc75524653)

[Gambar 3.28 Perancangan Halaman Verifikasi OTP 60](#_Toc75524654)

[Gambar 3.29 Perancangan Halaman *Update* Profil 60](#_Toc75524655)

[Gambar 3.30 Perancangan Halaman Daftar Obrolan 61](#_Toc75524656)

[Gambar 3.31 Perancangan Halaman Daftar Kontak 62](#_Toc75524657)

[Gambar 3.32 Perancangan Halaman Ruang Obrolan 62](#_Toc75524658)

[Gambar 3.33 Perancangan Halaman Profil Saya 63](#_Toc75524659)

[Gambar 3.34 Perancangan Halaman Profil Teman 64](#_Toc75524660)

[Gambar 3.35 Perintah *Training* Korpus HMM 65](#_Toc75524661)

[Gambar 3.36 Perintah Menjalankan *Tagging* HMM 66](#_Toc75524662)

[Gambar 3.37 Perintah Instalasi Moses Decoder 66](#_Toc75524663)

[Gambar 3.38 Perintah *Preprocessing* Korpus Paralel 67](#_Toc75524664)

[Gambar 3.39 Perintah *Training* Korpus Paralel Bahasa Melayu Sambas   
ke Bahasa Indonesia 68](#_Toc75524665)

[Gambar 3.40 Perintah Menjalankan Moses Decoder pada Server 68](#_Toc75524666)

[Gambar 3.41 Proses Evaluasi Secara Manual oleh Ahli Bahasa 70](#_Toc75524667)

[Gambar 4.1 Halaman Verifikasi Nomor Telepon 75](#_Toc63430082)

[Gambar 4.2 Halaman Verifikasi OTP 76](#_Toc63430083)

[Gambar 4.3 Halaman *Update* Profil 77](#_Toc63430084)

[Gambar 4.4 Halaman Daftar Obrolan 78](#_Toc63430085)

[Gambar 4.5 Halaman Daftar Kontak 79](#_Toc63430086)

[Gambar 4.6 Halaman Ruang Obrolan 80](#_Toc63430087)

[Gambar 4.7 Halaman Profil Saya 81](#_Toc63430088)

[Gambar 4.8 Halaman Profil Teman 82](#_Toc63430089)

[Gambar 4.9 Contoh Perintah Membuat Hasil Terjemahan 86](#_Toc63430090)

[Gambar 4.10 Contoh Perintah Menggunakan BLEU 86](#_Toc63430091)

DAFTAR KODE PROGRAM

[Kode Program 3.1. *Source Code* Program Java *Tagging* 65](#_Toc63429954)

[Kode Program 4.1. Kode Pengujian *Accuracy*, *Precision*, *Recall*,   
dan *F-Measure* 83](#_Toc66046662)

DAFTAR LAMPIRAN

[LAMPIRAN A-1 Contoh Korpus *Tagging* HMM Skenario 1 A-1](#_Toc62811915)

[LAMPIRAN B-1 Contoh Korpus *Tagging* HMM Skenario 2 B-1](#_Toc62811915)

[LAMPIRAN C-1 Contoh Korpus Paralel Bahasa Indonesia −   
Bahasa Melayu Sambas C-1](#_Toc62811858)

[LAMPIRAN D-1 *Service* Penerjemah PHP D-1](#_Toc62811955)

[LAMPIRAN E-1 Contoh Hasil Terjemahan Langsung Moses Decoder   
(Bahasa Sambas *Chatting* ke Bahasa Indonesia) E-1](#_Toc62811858)

[LAMPIRAN F-1 Contoh Hasil Terjemahan Moses Decoder Dengan Pra Proses HMM (Bahasa Sambas ke Bahasa Indonesia) F-1](#_Toc62811858)

# BAB I PENDAHULUAN

## Latar Belakang

Komunikasi adalah pengiriman dan penerimaan pesan atau berita antara dua orang atau lebih sehingga pesan yang dimaksud dapat dipahami. Bahasa merupakan alat komunikasi yang digunakan oleh manusia, dengan bahasa seseorang dapat menyampaikan ide, mengenalkan diri, dan menceritakan pengalamannya kepada orang lain. Dalam pemakaiannya, keragaman bahasa dapat berupa bahasa lisan atau bahasa tulis bergantung pada kebutuhan dan tujuan komunikasi (Cenderamata and Sofyan, 2019). Indonesia sendiri memiliki keragaman bahasa dan budaya yang luar biasa, sebanyak 726 bahasa daerah dengan 719 bahasa daerah diantaranya masih aktif digunakan sehari-hari di seluruh pelosok Indonesia (Ginting and AZ, 2012).

Berkembangnya teknologi informasi berdampak pada akses komunikasi yang menjadi lebih luas. *Chatting* merupakan bentuk komunikasi yang biasanya dilakukan antara dua orang atau lebih secara langsung atau *realtime* dengan memanfaatkan fasilitas jaringan (Sahara, 2018). *Chatting* bukan lagi hal yang diminati, melainkan bagian dari kebutuhan sosial manusia untuk memudahkan komunikasi baik jauh maupun dekat (Hardianto and Handaga, 2015). Dalam melakukan komunikasi selama ini di kalangan masyarakat umum ketika berkomunikasi lewat *chatting* mengalami kendala jika akan berkomunikasi dengan orang yang berbeda bahasanya (Syaukani, 2010).

Mesin penerjemah statistik adalah mesin yang dapat melakukan proses penerjemahan dari satu bahasa ke bahasa lainnya secara otomatis dengan hasil terjemahan diperoleh atas dasar model statistik yang parameter-parameternya diambil dari hasil analisis korpus teks bilingual (korpus paralel), dan Moses Decoderadalahsalah satu sistem mesin penerjemah statistik yang sering digunakan (Pratiwi, Sujaini and Nyoto, 2017). Beberapa penelitian seperti “Aplikasi Penerjemah Dua Arah Bahasa Indonesia – Bahasa Melayu Sambas Berbasis Web Dengan Menggunakan Decoder Moses” oleh Andri Hidayat, “Implementasi Mesin Penerjemah Statistik Berbasis Android Dengan Moses Decoder” oleh Try Wahyudinata, dan “Pengembangan Antarmuka Mesin Penerjemah Statistik Multibahasa Berbasis Web” oleh Ninda Fitria Pratiwi membuktikan bahwa mesin penerjemah statistik berhasil untuk diterapkan dalam bentuk aplikasi penerjemah, namun belum digunakan sebagai media komunikasi.

Sebagai model komunikasi verbal yang berwujud teks atau tulis dan digunakan oleh berbagai lapisan masyarakat, umumnya bahasa dalam aktivitas *chatting* menggunakan kata tidak baku sehingga muncul variasi dalam penuturan bahasa oleh pelaku aktivitas *chatting*. Perubahan bentuk atau variasi bahasa pada komunikasi teks dapat berupa abreviasi atau penyingkatan kata, penggunaan slang, maupun lain sebagainya. Mesin penerjemah statistik tidak dapat langsung diterapkan pada komunikasi teks (*chatting*) karena kata yang digunakan umumnya tidak baku dan memiliki variasi yang sangat banyak serta berubah-ubah bergantung kondisi orang yang menuturkan suatu bahasa, sehingga diperlukan perbaikan kata pada kalimat atau bahasa *chatting* menjadi bentuk baku pada suatu bahasa, baru kemudian dilakukan penerjemahan.

Beberapa penelitian normalisasi teks atau perbaikan kata tidak baku yang pernah dilakukan seperti penelitian penelitian Andri, Sunda Ariana, dan Margareta Andriani dengan judul “Aplikasi Koreksi Kesalahan Berbasis Pada Tulisan Berbahasa Indonesia Untuk Meningkatkan Kualitas Penulisan Karya Ilmiah” yang menggunakan metode *N-gram* untuk melakukan pengoreksian kesalahan-kesalahan terhadap kata dan kalimat yang tidak sesuai dengan EyD. Penelitian Jinhua dkk. dengan judul “Extended HMM and Ranking models for Chinese Spelling Correction” yang membuat suatu *framework* untuk melakukan perbaikan kata dengan metode HMM dan *rangking models*.

Berdasarkan masalah yang telah dipaparkan, penulis akan melakukan penelitian untuk membuat implementasi dari mesin penerjemah statistik berupa aplikasi *chatting* berbasis android dengan memanfaatkan Moses Decoder sebagai mesin penerjemah dan HMM dengan *trigram* untuk perbaikan bahasa *chatting* ke bahasa baku, sehingga dapat digunakan sebagai media komunikasi tulis antar pengguna yang berbeda bahasa. Pada penelitian ini akan dilakukan perbaikan pada bentuk kalimat *chatting* bahasa Melayu Sambas menjadi bentuk baku bahasa Melayu Sambas dan kemudian penerjemahan bahasa Melayu Sambas menjadi bahasa Indonesia.

## Perumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan, maka rumusan masalah dalam penelitian tugas akhir ini yaitu bagaimana menggunakan memperbaiki teks pada bahasa *chatting* dengan HMM dan mengimplementasikan mesin penerjemah statistik Moses Decoder pada aplikasi *chatting* berbasis android.

## Tujuan Penelitian

Adapun tujuan dari penelitian yang dilakukan adalah membuat sebuah aplikasi *chatting* berbasis android yang mampu memperbaiki bahasa *chatting* ke bahasa baku menggunakan HMM dan menerjemahkan bahasamenggunakan mesin penerjemah statistik Moses Decoder, sehingga dapat digunakan pada komunikasi teks antar pengguna bahasa yang berbeda.

## Pembatasan Masalah

Untuk menghindari meluasnya permasalahan yang ada, serta keterbatasan ilmu dan kemampuan yang dimiliki, maka batasan penelitian ini antara lain:

1. Penerjemahan dilakukan dengan menggunakan 4500 baris korpus teks paralel yaitu dokumen teks *chatting* bahasa Melayu Sambas dan dokumen bahasa Indonesia.
2. Penelitian ini berfokus untuk mengimplementasikan mesin penerjemah statistik Moses Decoder sebagai alat penerjemahan bahasa dan HMM untuk perbaikan bahasa *chatting*.
3. Aplikasi *chatting* dibangun berbasis android dengan *input* berupa pesan teks.
4. Pembahasan untuk penerjemahan hanya dilakukan satu arah yaitu dari bahasa Melayu Sambas ke bahasa Indonesia.

## Sistematika Penulisan

Sistematika dari penulisan tugas akhir ini disusun dalam 5 (lima) bab yang terdiri dari Bab I Pendahuluan, Bab II Tinjauan Pustaka, Bab III Metodologi Penelitian, Bab IV Hasil Implementasi dan Pengujian, serta Bab V Penutup.

**Bab I Pendahuluan** adalah bab yang berisi latar belakang, perumusan masalah, tujuan penelitian, pembatasan masalah, dan sistematika penulisan.

**Bab II Tinjauan Pustaka** adalah bab yang berisi berisi uraian sistematis yang berhubungan dengan penelitian yang akan dilakukan. Beberapa teori-teori yang terkait adalah pengertian mesin penerjemah statistik, *Unified Modelling Language* (UML), Java, PHP, Penelitian Terkait, dan Pengujian Perangkat Lunak.

**Bab III Metodologi Penelitian** adalah bab yang berisi Metodologi Penelitian, Alat dan Data Penelitian, Diagram Alir Penelitian, Perancangan Aplikasi, Perancangan UML, Perancangan Basis Data, Perancangan Antarmuka serta Rencana Pengujian.

**Bab IV Hasil Implementasi dan Pengujian** adalah bab yang berisi implementasi atau hasil perancangan yang telah dilakukan pada Bab III dalam bentuk tampilan aplikasi. Pada bab ini juga dilakukan pengujian, pada hasil dari pengujian dilakukan analisis untuk mengarah kepada suatu kesimpulan.

**Bab V Penutup** adalah bab yang berisi kesimpulan dari penelitian yang telah dilakukan dan saran/rekomendasi untuk perbaikan atau pengembangan dari penelitian yang telah dilakukan.

# TINJAUAN PUSTAKA

## Kajian Terkait

Dalam penelitian ini akan digunakan tiga tinjauan studi terkait implementasi mesin penerjemah statistik yang nantinya mendukung penelitian yang akan dilakukan. Berikut merupakan kajian terkait dari penelitian yang telah dilakukan:

Jinhua dkk. (2014) melakukan penelitian perbaikan kata bahasa Cina (Mandarin) dengan menerapkan HMM dan *ranking models*. Data penelitian diambil halaman *web* Taiwan yang mengandung teks tradisional Cina berkualitas tinggi, kamus SogouW, kamus kata dan idiom Cina, tabel kode *Pinyin* dan *Canjie*. Pada penelitian ini Jinhua dkk. memperkenalkan suatu kerangka kerja “HANSpeller” yang menerapkan konsep HMM dan *ranking models* dalam koreksi ejaan bahasa Mandarin. Hasil dari penelitian ini telah mereka evaluasi pada CLP-2014 Bake-Off dan menghasilkan performa atau kinerja yang baik dalam perbaikan ejaan bahasa Cina.(Xiong *et al.*, 2015)

Andri Hidayat (2015) melakukan penelitian tentang implementasi aplikasi penerjemah berbasis web. Penelitian ini dilakukan dengan menggunakan korpus paralel bahasa Melayu Sambas dan bahasa Indonesia yang bersumber dari cerita rakyat daerah Kabupaten Sambas. Dengan pengujian secara otomatis menggunakan BLEU, didapat nilai BLEU *score* sebesar 58,50% untuk penerjemahan bahasa Indonesia ke bahasa Melayu Sambas dan 63,76% untuk penerjemahan bahasa Melayu Sambas ke bahasa Indonesia. Penerjemahan dengan mesin penerjemah statistik (Moses Decoder) ini berhasil diintegrasikan ke dalam sebuah aplikasi berbasis web. Sehingga berdasarkan hasil UAT (*User Acceptance Test*) didapat penilaian dengan rata-rata 89,16%.

Try Wahyudinata (2016) melakukan penelitian tentang implementasi mesin penerjemah statistik berbasis android dengan Moses Decoder. Penelitian ini menghasilkan sebuah aplikasi berbasis android yang mampu memberikan keluaran (*output*) berupa kata/kalimat yang merupakan hasil terjemahan dari masukan (*input*) sebelumnya. Aplikasi ini diuji dengan metode *robustness testing* (*Black Box*), kompabilitas aplikasi, dan kuesioner terhadap 25 responden yang terdiri dari mahasiswa. Hasil dari *robustness testing* menyatakan aplikasi dapat meng*handle* (mengatasi) proses *input* dengan baik, terutama pada proses *input* data yang tidak sesuai dan kosong. Sedangkan hasil kuesioner menggunakan skala *Likert’s Summated Rating* (LSR) diperoleh nilai total 1477 yang berarti aplikasi diimplementasikan dengan sangat baik.

Ninda Fitria Pratiwi (2017) melakukan penelitian pengembangan antarmuka mesin penerjemah statistik multi bahasa berbasis *web*. Tujuan penelitian ini adalah mengembangkan antarmuka mesin penerjemah statistik multi bahasa dengan fitur komunitas terjemahan dimana pengguna dapat berkontribusi dalam aplikasi translator bahasa daerah sehingga dapat diakses secara luas melalui internet oleh publik. Pada penelitian ini peneliti melakukan pengujian menggunakan metode *Black Box* dengan teknik *robustness testing* dan kuesioner. Hasil *robustness testing* menyatakan aplikasi dapat meng*handle* (mengatasi)proses *input* dengan baik pada masing-masing bahasa. Sedangkan hasil kuesioner menggunakan skala *Likert’s Summated Rating* (LSR) diperoleh nilai total 1219 yang berarti aplikasi sangat positif dan dinilai berhasil.

Dari penelitian-penelitian sebelumnya dapat dilihat bahwa normalisasi teks pada kalimat tidak baku dan implementasi mesin penerjemah statistik Moses Decoder dapat dilakukan. Pada penelitian ini dilakukan pengembangan untuk membuat aplikasi *chatting* yang menggunakan mesin penerjemah statistik Moses Decoder untuk menerjemahkan bahasa dan menggunakan HMM untuk pra proses atau perbaikan bahasa *chatting* sebelum diterjemahkan.

Perbandingan penelitian-penelitian yang terdapat pada kajian terkait diperlihatkan pada Tabel 2.1.

Tabel 2.1 Tabel Perbandingan Penelitian

| **No** | **Penulis** | **Judul** | **Keterangan** |
| --- | --- | --- | --- |
| 1 | Jinhua dkk. | *Extended HMM and Ranking models for Chinese Spelling Correction* | Melakukan penelitan untuk perbaikan ejaan bahasa Cina dengan menerapkan HMM dan *ranking models*. |
| 2 | Andri Hidayat | Aplikasi Penerjemah Dua Arah Bahasa Indonesia – Bahasa Melayu Sambas Berbasis WEB Dengan Menggunakan Decoder Moses | Menggunakan korpus paralel bahasa Indonesia dan bahasa Melayu Sambas.  Melakukan pengujian pada *decoder moses* dengan BLEU dan pengujian WEB dengan metode *Black-Box*.  Membangun antarmuka (*interface*) berbasis WEB. |
| 3 | Try Wahyudinata | Implementasi Mesin Penerjemah Statistik Berbasis Android Dengan Moses Decoder | Menggunakan korpus paralel bahasa Indonesia dan bahasa Melayu Pontianak.  Melakukan pengujian sistem dengan metode *Black Box.*  Mengimplementasikan mesin penerjemah statistik berbasis Android.  Menggunakan konfigurasi pada aplikasi agar dapat diimplementasikan untuk multi bahasa. |
| 4 | Ninda Fitria Pratiwi | Pengembangan Antarmuka Mesin Penerjemah  Statistik Multi Bahasa Berbasis Web | Menggunakan korpus paralel 5 bahasa daerah, yaitu bahasa Melayu Ketapang, bahasa Melayu Pontianak, bahasa Melayu Sambas, bahasa Dayak Kanayant dan bahasa Jawa Kromo.  Melakukan pengembangan antarmuka mesin penerjemah statistik multi bahasa berbasis web.  Melakukan pengujian sistem dengan metode *Black Box.* |
| 5 | Muhammad Nur Amin | Implementasi Mesin Penerjemah Statistik Pada Aplikasi *Chatting* Berbasis Android Dengan Moses Decoder | Menggunakan korpus paralel bahasa Indonesia dan bahasa Melayu Sambas.  Dokumen korpus dibuat berdasarkan hasil ekspor percakapan pada aplikasi WhatsApp.  Membuat aplikasi *chatting* yang mengimplementasikan mesin penerjemah statistik.  Menggunakan HMM dengan *trigram* untuk perbaikan kata pada bahasa *chatting* sebelum diterjemahkan. |

## *Chatting*

Komunikasi merupakan aktivitas pengiriman dan penerimaan pesan atau berita antara dua orang atau lebih sehingga pesan yang dimaksud dapat dipahami. Hadirnya layanan komunikasi dan informasi yang canggih tadi telah memanjakan kehidupan manusia era sekarang dalam berbagai aspek, terutama dalam hal hidup bersosial dengan sesamanya. *Chatting* sebagai bagian dari media komunikasi di bidang teknologi informasi menjadi salah satu layanan yang paling disukai anak-anak muda, bahkan orang dewasa saat ini, dalam rangka menjalin hubungan sosial dan bermasyarakat di dunia maya. Aktivitas *chatting* sendiri merupakan bentuk komunikasi yang berwujud tuturan verbal teks (bahasa tulis) yang disertai dengan pemakaian simbol-simbol (bentuk *emoticon*) atau kode-kode bahasa sebagai ganti ujaran lisan yang tidak dapat diekspresikan ke dalam ungkapan verbal karena keterbatasan media yang digunakannya (Hariyanto, 2010).

Aplikasi *chatting* sendiri merupakan alat komunikasi dua arah antara satu orang atau beberapa orang baik teks, suara, maupun video. Pemanfaatan aplikasi *chatting* ini akan menghemat waktu, tenaga, dan biaya karena tidak perlu melakukan perjalanan jauh dan melelahkan untuk menjalin komunikasi (Putri, 2011).

Sebagai model komunikasi verbal yang berwujud teks atau tulis dan digunakan oleh berbagai lapisan masyarakat dalam berbagai situasi dan konteks sosial, umumnya bahasa dalam aktivitas *chatting* menggunakan kata tidak baku yaitu kata yang digunakan pada situasi tidak resmi atau informal. Karena bahasa yang digunakan bersifat tidak baku maka muncul variasi dalam penuturan bahasa oleh pelaku aktivitas *chatting*. Faktor-faktor sosial yang memengaruhi variasi bahasa dapat diketahui berdasarkan latar belakang penutur bahasa, beberapa faktor tersebut seperti latar belakang pekerjaan, usia, pendidikan, gender, tingkat kebangsawanan, keadaan sosial ekonomi, dan sebagainya (Utami, 2010).

## Normalisasi Teks

Dalam aktivitas *chatting* atau komunikasi teks penggunaan kata tidak baku menyebabkan perubahan bentuk pada suatu bahasa seperti terjadinya abreviasi atau penyingkatan kata, penggunaan slang, maupun lain sebagainya sehingga muncul banyak variasi dalam suatu bahasa. Tantangan terbesar dalam model komputasi bahasa diantaranya ambiguitas makna kata dan ambiguitas sintaksis, sementara penulisan atau ejaan yang tidak baku/sempurna merupakan salah satu penyebabnya (Priansya, 2017). Karena setiap orang dapat memiliki bentuk atau variasi yang berbeda-beda (yang tidak baku) dalam menuturkan satu bahasa pada komunikasi teks, maka diperlukan normalisasi teks.

Normalisasi teks atau normalisasi kata adalah mengubah kata yang tidak baku menjadi kata baku dalam suatu bahasa (Nugraha and Rizqullah, 2019). Normalisasi teks pada penelitian ini akan dilakukan dengan metode *Hidden Markov Model* (HMM) dengan rangkaian *trigram* (3 gram).

## *Hidden Markov Model* (HMM)

*Hidden Markov Model* atau HMM merupakan model statistik dimana suatu sistem yang dimodelkan diasumsikan sebagai *markov* proses dengan dengan kondisi yang tidak terobservasi (tidak dapat diamati). Suatu HMM dapat dianggap sebagai jaringan Bayesian dinamis yang sederhana (*simplest dynamic Bayesian network*)(Prasetyo, 2010).

Dalam model *Markov* normal, pengamat dapat melihat semua keadaan secara langsung. Oleh karena itu, probabilitas kondisi transisi adalah satu-satunya parameter yang teramati. Dalam HMM, keadaan tidak terlihat secara langsung, namun keluaran yang bergantung terhadap keadaan tersebut terlihat.

Setiap kondisi memiliki distribusi probabilitas pada semua kemungkinan keluar. Oleh karena itu, urutan langkah yang dihasilkan oleh HMM memberikan informasi tentang urutan keadaan. Harus dipahami bahwa sifat *hidden* (tersembunyi) mengacu pada keadaan bagian yang dilalui model, bukan parameter model. Model tetap tersembunyi bahkan jika parameter model diketahui.

*Hidden Markov Model* dapat digunakan untuk aplikasi dibidang *temporal pattern recognition* “pengenalan pola temporal” seperti pengenalan suara, tulisan, gestur, bioinformatika, kompresi kalimat, *computer vision*, ekonomi, finansial, dan pengenalan not balok.

Pada penelitian ini model HMM digunakan untuk pra proses atau perbaikan kalimat *chatting* bahasa Melayu Sambas ke kalimat baku bahasa Melayu Sambas.

## *N-gram*

Pada dasarnya model *N-gram* merupakan model probabilistik, yang kemudian dikembangkan untuk memprediksi item berikutnya dalam urutan item. Tergantung pada aplikasinya, item dapat berupa huruf/karakter, kata, atau lainnya. Salah satunya, model *N-gram* berbasis kata digunakan untuk memprediksi kata berikutnya menurut urutan kata tertentu. Dalam arti tertentu, *N-gram* hanyalah kumpulan kata, dan setiap kata memiliki panjang n kata. Misalnya, *N-gram* ukuran 1 disebut “*unigram*”; ukuran 2 adalah “*bigram*”; ukuran 3 adalah “*trigram*”, dan seterusnya.

Dalam pembuatan kata, metode *N-gram* digunakan untuk mengekstrak n fragmen kata dari rangkaian kata (kalimat, paragraf, bacaan), yang dibaca terus menerus dari teks sumber hingga akhir dokumen. Misalnya, kalimat: “Saya dapat melihat cahaya itu” dapat dipecah menjadi *N-gram* berikut:

* uni-gram: saya, dapat, melihat, cahaya, itu
* bi-gram: saya dapat, dapat melihat, itu ada
* tri-gram: saya dapat melihat, dapat melihat dia dan seterusnya.

Pada penelitian ini normalisasi atau perbaikan kata dari bahasa *chatting* dilakukan menggunakan HMM dengan rangkaian *N-gram* ialah *trigram* (3 gram).

## Natural Language Toolkit (NLTK)

Natural Language Toolkit (NLTK) merupakan *library*, modul atau serangkaian program yang dibangun untuk mempermudah dalam pengelolaan dan pemrosesan bahasa alami. NLTK untuk Bahasa Inggris awalnya dikembangkan pada tahun 2001 oleh Steven Bird dan Edward Loper dari Departemen Ilmu Komputer dan Informasi di Universitas Pennsylvania untuk mendukung penelitian dan pengajaran di NLP atau bidang terkait, seperti komputasi bahasa, kecerdasan buatan, pembelajaran mesin, dan lain-lain (Loper and Bird, 2002).

NLTK dibangun dengan bahasa pemrograman python dan pada penelitian ini digunakan untuk analisis teks yaitu pengujian *accuracy*, *precision*, *recall*, dan *f-measure* dari pra proses HMMyang dilakukan.

## Mesin Penerjemah Statistik

Mesin penerjemah statistik adalah suatu mesin penerjemah dimana penerjemahan dilakukan berbasis model statistik dengan parameter-parameter yang diambil dari analisis korpus paralel. Pendekatan statistik yang digunakan adalah konsep probabilitas, dimana setiap pasangan kalimat (S,T) akan diberikan sebuah P(T|S) yang diinterpretasikan sebagai distribusi probabilitas yang sebuah penerjemah akan menghasilkan T dalam bahasa tujuan ketika diberikan S dalam bahasa sumber (Tanuwijaya and Manurung, 2009).

Pada mesin penerjemah statistik terdapat 3 komponen yaitu *language model*, *translation model*, dan *decoder*. Ketiga komponen inilah yang terlibat dalam proses penerjemahan bahasa dari bahasa asal ke bahasa target atau dari satu bahasa ke bahasa lainnya (Hadi, 2014). Ketiga komponen pada mesin penerjemah statistik diperlihatkan pada gambar 2.1 berikut.



Gambar 2.1 Komponen Mesin Penerjemah Statistik

Pada penelitian ini mesin penerjemah yang menggunakan pendekatan statistik atau mesin penerjemah statistik dipilih sebagai alat penerjemahan bahasa. Penerjemahan dengan mesin penerjemah statistik menghasilkan terjemahan yang lebih baik dibandingkan dengan hanya terjemahan kata demi kata (*interlinear translation*), dengan syarat kualitas korpus paralel (padanan kalimat-kalimat bahasa sumber dan bahasa tujuan) yang dimasukkan ke dalam sistem mempunyai kualitas baik dan cukup banyak jumlahnya (Ginting and AZ, 2012).

## Moses Decoder

Moses Decoder merupakan implementasi mesin penerjemah dengan pendekatan statistik (*statistical machine translation*) yang menggunakan konsep probabilitas dalam melakukan penerjemahan bahasa berdasarkan hasil analisis dari korpus paralel. Korpus paralel adalah kumpulan dari kalimat-kalimat dalam dua bahasa yang berbeda, yang merupakan kalimat-kalimat selaras, setiap kalimat dalam satu bahasa cocok dengan kalimat yang diterjemahkan dalam bahasa lain (Wahyudinata, Sujaini and Nyoto, 2016).

Korpus paralel dapat digunakan oleh Moses Decoder setelah melakukan beberapa proses, dimulai dari *pre-processing* (pra-proses) berupa *Tokenize training data*, *Filter out long* *sentences* dan *Lowercase training data* hingga membangun model bahasa (*language modelling*) dan *training* pada sistem (Ginting and AZ, 2012).

Penelitian ini menggunakan Moses Decoder untuk menghasilkan bahasa target yang memiliki nilai probabilitas tertinggi sebagai hasil dari penerjemahan bahasa sumber dengan menggunakan korpus paralel yang sudah diproses sebelumnya. Selain itu juga akan dibangun model bahasa dengan korpus monolingual bahasa terget menggunakan KenLM, lalu model translasi dan susunan kata untuk penerjemahan dibangun dengan GIZA++.

## Korpus

Saat ini konsepsi korpus telah berkembang dan bermakna luas. Sekarang, korpus juga dapat dikumpulkan secara elektronik. Ada tiga aspek yang menjadi pertimbangan dalam memahami konsep. Pertama, korpus utamanya merupakan kumpulan teks yang dihasilkan secara elektronik dan dapat dianalisis secara otomatis atau semi otomatis. Kedua, korpus tidak hanya berisi kumpulan teks bentuk tulis, tetapi juga mencakup ujaran. Ketiga, korpus mungkin juga mencakup sejumlah besar teks yang berasal dari beragam sumber, misalnya dari beragam penulis dan penutur dan dalam berbagai topik (Baker, 1995).

Korpus didefinisikan sebagai koleksi atau sekumpulan contoh teks tulis atau lisan dalam bentuk yang dapat dibaca dengan menggunakan seperangkat mesin dan dapat diberi catatan berupa berbagai bentuk informasi linguistik (McEnery, Xiao and Tono, 2006).

Korpus yang digunakan pada penelitian ini adalah korpus teks *chatting* bahasa sambas yang sudah dilakukan *tagging* perbaikan kata dan korpus paralel bahasa Melayu Sambas dan bahasa Indonesia. Korpus paralel (*parallel corpus*) sendiri adalah pasangan korpus yang berisi kalimat-kalimat dalam suatu bahasa dan terjemahannya. Korpus paralel akan digunakan untuk membangun mesin penerjemah statistik Moses Decoder, sedangkan korpus *tagging* digunakan untuk membangun mesin pra proses HMM.

## *Unified Modelling Language* (UML)

*Unified Modelling Language* (UML) adalah suatu alat untuk memvisualisasikan dan mendokumentasikan hasil analisa dan desain yang berisi sintak dalam memodelkan sistem secara visual, dan juga merupakan satu kumpulan konvensi pemodelan yang digunakan untuk menentukan atau menggambarkan sebuah sistem *software* yang terkait dengan objek (Ropianto, 2016). UML adalah sebuah bahasa standar untuk pengembangan sebuah *software* yang dapat menyampaikan bagaimana membuat dan membentuk model-model, tetapi tidak menyampaikan apa dan kapan model yang seharusnya dibuat sehingga merupakan salah satu proses implementasi pengembangan *software* (Hidayat, Sujaini and Dwinyoto, 2015). UML mendefinisikan beberapa diagram diantaranya *use case diagram*, *class diagram*, *activity diagram*, *sequence diagram*, dan diagram lainnya (Dharwiyanti and Wahono, 2003).

*Unified Modelling Language* (UML) sendiri akan digunakan untuk merancang atau menggambarkan model dari aplikasi *chatting* yang akan dibangun. Beberapa diagram yang akan digunakan seperti *use case diagram* adalah untuk menggambarkan interaksi antara pengguna dengan sistem yang akan dibuat. *Class diagram* digunakan untuk membantu menggambarkan atribut, operasi dan relasi antara kelas-kelas dari sistem yang akan dibangun. *Statechart diagram* digunakan untuk menggambarkan perilaku (*behavior*) dinamis sistem, transisi atau perubahan keadaan sistem dalam merespon suatu aksi yang diterima. Sedangkan *sequence diagram* akan digunakan untuk menggambarkan tahapan kejadian pada sistem berdasarkan interaksi pengguna pada *use case diagram*.

## Android

### Sistem Operasi Android

Android adalah sistem operasi berbasis *linux* untuk *mobile* seperti *smartphone* dan *tablet* PC. Android termasuk sistem operasi *open source* sehingga semua pengguna dapat mengembangkan dan membuat aplikasi sendiri (Hardianto and Handaga, 2015).

Sistem operasi android dipilih dalam membuat aplikasi *chatting* karena pertukaran data dan komunikasi antar proses yang mudah dengan adanya *Intent* dan *Content Provider*, tidak adanya perbedaan antara aplikasi yang dikembangkan oleh pihak ketiga dengan aplikasi inti bawaan android, serta memungkinkan sebuah aplikasi untuk berjalan di *background* dengan aplikasi lainnya.

### Android Studio

Android Studio adalah [Integrated Development Enviroment](https://id.wikipedia.org/wiki/Lingkungan_pengembangan_terpadu) (IDE) untuk sistem operasi [Android](https://id.wikipedia.org/wiki/Android_(sistem_operasi)), yang dibangung diatas perangkat lunak IntelliJ IDEA dan didesain khusus untuk pengembangan Android. IDE ini merupakan pengganti dari Eclipse Android Development Tools (ADT) yang sebelumnya merupakan IDE utama untuk pengembangan aplikasi android. Android studio sendiri pertama kali diumumkan di Google I/O Conference pada tanggal 16 Mei 2013.

Android Studio menawarkan banyak fitur yang meningkatkan produktivitas penggunanya dalam membuat aplikasi Android, seperti (Android Developer, no date):

1. Sistem build berbasis Gradle yang fleksibel.
2. Emulator yang cepat dan kaya fitur.
3. Lingkungan terpadu tempat Anda dapat mengembangkan aplikasi untuk semua perangkat Android.
4. Instant Run untuk mengirim perubahan ke aplikasi Anda yang sedang berjalan tanpa membuat APK baru.
5. *Template* kode dan integrasi GitHub untuk membantu Anda membuat fitur aplikasi umum dan mengimpor kode sampel.
6. *Framework* dan fitur pengujian yang lengkap.
7. Fitur *lint* untuk merekam performa, kegunaan, kompatibilitas versi, dan masalah lainnya.
8. Dukungan C++ dan NDK.
9. Dukungan bawaan untuk [Google Cloud Platform](https://cloud.google.com/tools/android-studio/docs/?hl=id), yang memudahkan integrasi *Google Cloud Messaging* dan *App Engine*.

Android Studio akan digunakan untuk membangun aplikasi *chatting* berbasis android dan dalam penelitian ini Android Studio akan digunakan bersama dengan bahasa pemrograman java. Ada dua jenis *file* yang nantinya akan banyak digunakan dalam membangun sebuah aplikasi android, yaitu *file* java dan *file* XML. *File* java berisi perintah-perintah dan logika-logika yang akan dieksekusi oleh aplikasi android. Sedangkan *file* XML digunakan untuk membangun tampilan dari aplikasi android seperti mengatur tata letak, menambahkan tombol, dan sebagainya. XML (*eXtensible Markup Language*) sendiri merupakan bahasa *web* turunan dari SGML (*Standard Generalized Markup Language*) yang ada sebelumnya, dan hampir sama dengan HTML (*Hypertext Markup Language*) dimana keduanya sama-sama turunan dari SGML (Hardianto and Handaga, 2015).

## Java

Java merupakan bahasa pemrograman *multi platform* dan *multi device* sehingga ketika suatu program ditulis dengan bahasa java maka program tersebut dapat berjalan pada *platform* dengan sistem operasi yang berbeda karena sifatnya yang umum dan non-spesifik. Aplikasi berbasis Java ini dikompilasikan ke dalam *p-code* dan bisa dijalankan dengan *Java Virtual Machine* (Hardianto and Handaga, 2015).

Pada penelitian ini bahasa java akan digunakan sebagai bahasa pemrograman untuk membuat aplikasi *chatting* berbasis android. Bahasa java akan digunakan untuk membuat perintah-perintah yang akan dieksekusi oleh aplikasi android. Bahasa java juga merupakan salah satu jenis bahasa pemrograman yang berorientasi pada objek (*Object Oriented Programming*), artinya fungsi dan variabel dibungkus dalam sebuah objek atau kelas yang dapat saling berinteraksi sehingga membentuk sebuah program.

## HyperText Preprocessor (PHP)

*HyperText Preprocessor* yang dikenal sebagai PHP adalah bahasa pemrograman *web* yang dapat ditanamkan ke dalam bahasa HTML yang diciptakan oleh Rasmus Lerdorf, seorang programer bahasa C yang sangat handal. Awalnya PHP merupakan kependekan dari *Personal Home Page* dan pada awalnya PHP digunakan untuk mengelola data formulir dari *web*. PHP merupakan bahasa pemrograman yang diciptakan secara gratis dan *open source* yang bersifat bebas (Wijayanto, 2012). PHP merupakan *Server Side Scripting*, dimana PHP selalu membutuhkan *web server* dalam menjalankan aksinya. Ketika sebuah halaman PHP diakses, kode PHP yang dibaca oleh *server* halaman tersebut. *Output* dari fungsi PHP pada halaman biasanya dikembalikan sebagai kode HTML yang dapat dibaca oleh browser.

Dalam membangun aplikasi *chatting* nantinya, aplikasi akan berhubungan dengan *server* penerjemah melalui *web service*. *Web service* sendiri adalah sekumpulan *application logic* beserta objek-objek dan metode-metode yang dimilikinya, yang terletak di suatu *server* yang terhubung ke internet sehingga dapat diakses menggunakan *protocol* HTTP dan SOAP (*Simple Object Access Protocol*) (Andi Juansyah, 2015). *Web service*  disini akan dibuat menggunakan bahasa pemrograman PHP. Pada *file* PHP berisi perintah-perintah untuk menjalankan mesin pra proses HMM dan mesin penerjemah Moses Decoder.

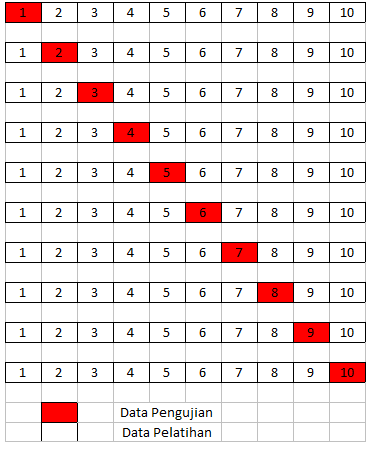
## Firebase

Firebase adalah sebuah penyedia layanan berupa *database realtime* dan *backend* yang dapat digunakan pada berbagai *platform*. *Backend* sendiri adalah sebuah bagian dalam kode aplikasi yang berhubungan langsung dengan isi *database*. Dengan Firebase, pengembang aplikasi tidak perlu membuat *backend* sendiri melainkan memakai API yang telah disediakan oleh Firebase sehingga pengembangan aplikasi dapat dipersingkat. Firebase Cloud Messaging (FCM) adalah solusi perpesanan lintas platform yang memungkinkan anda mengirimkan pesan tanpa biaya dengan andal. Dengan menggunakan FCM, anda dapat memberitahu aplikasi klien bahwa *email* baru atau data lain tersedia untuk disinkronkan. Anda dapat mengirim pesan pemberitahuan untuk mendorong keterlibatan dan retensi pengguna. Untuk kasus penggunaan seperti pesan instan, pesan dapat mentransfer muatan hingga 4KB ke aplikasi klien (Aditya, 2018).

Firebase sendiri merupakan suatu layanan dari Google yang digunakan untuk mempermudah para pengembang aplikasi dalam mengembangkan aplikasi. Pada penelitian ini Firebase akan digunakan untuk mengirim pesan yang sudah diterjemahkan, digunakan sebagai *realtime database*, digunakan sebagai otentikasi pengguna, dan digunakan untuk membuat notifikasi pesan.

## K-Fold Cross Validation

*Cross validation* (CV) merupakan metode statistik yang dapat digunakan untuk mengevaluasi kinerja suatu model atau algoritma yang membagi data menjadi dua subset (yaitu data *training*/pembelajaran dan data uji/validasi/evaluasi). Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi. Pemilihan tipe CV dapat didasarkan pada ukuran dataset atau kumpulan data. CV *k-fold* biasanya digunakan karena dapat mengurangi waktu perhitungan/komputasi dengan tetap menjaga keakuratan estimasi. Contoh skema 10 *fold cross validation* diperlihatkan pada gambar 2.2 berikut.



Gambar 2.2 Contoh Skema 10-*Fold Cross Validation*

Dalam penelitian ini skema 10-*fold cross validation* diterapkan pada data atau korpus dengan jumlah sebanyak 4.500 kalimat, sehingga diperoleh pembagian data uji sebagai berikut :

|  |  |
| --- | --- |
| 1. Data Uji = Kalimat 1 – 450 | 1. Data Uji = Kalimat 2.251 – 2.700 |
| 1. Data Uji = Kalimat 451 – 900 | 1. Data Uji = Kalimat 2.701 – 3.150 |
| 1. Data Uji = Kalimat 901 – 1.350 | 1. Data Uji = Kalimat 3.151 – 3.600 |
| 1. Data Uji = Kalimat 1.351 – 1.800 | 1. Data Uji = Kalimat 3.601 – 4.050 |
| 1. Data Uji = Kalimat 1.801 – 2.250 | 1. Data Uji = Kalimat 4.051 – 4.500 |

Perhitungan nilai rata-rata atau nilai akhir pada pengujian dengan metode *k-fold cross validation* diperoleh dengan persamaan 2.1 sebagai berikut:

(2.1)

## Pengujian

Secara umum dalam penelitian ini akan dilakukan tiga jenis pengujian yaitu pengujian hasil pra proses pesan, pengujian hasil penerjemahan pesan, dan pengujian fungsional aplikasi.

### Pengujian Hasil Pra Proses HMM

Pengujian pada hasil pra proses atau perbaikan kata dengan HMM meliputi pengujian *accuracy*, *precision*, *recall*, dan *f-measure*. Nilai *accuracy* menunjukkan tingkat kedekatan antara nilai prediksi dengan nilai aktual. Nilai *precision* menunjukkan tingkat ketepatan prediksi sistem dengan menghitung prediksi benar dari total data yang diprediksi sistem termasuk prediksi salah. Nilai *recall* menunjukkan tingkat keberhasilan sistem dalam mengenali suatu kelas yang harus dikenali. Sedangkan nilai *f–measure* adalah penggabungan *precision* dan *recall*, dan merupakan nilai yang mewakili secara keseluruhan tingkat kinerja sistem (Antinasari, Perdana and Fauzi, 2017).

Sehingga dapat dikatakan bahwa *accuracy* adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. *Recall* merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Dan *f-measure* sebagai gabungan dari *precision* dan *recall* adalah tingkat keakuratan sistem dalam pengujian. Pengujian *accuracy, precision, recall*, dan *f-measure* dilakukan untuk mengetahui nilai akurasi dari hasil pra proses HMM.

### Pengujian Hasil Penerjemahan Pesan

Pengujian hasil penerjemahan tanpa melibatkan responden atau sistem evaluasi otomatis disebut BLEU (*Bilingual Evaluation Understudy*). 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. BLEU mengukur *modified n-gram precision score* antara hasil terjemahan otomatis dengan terjemahan rujukan dasn menggunakan konstanta yang dinamakan *brevity penalty* (Hadi, 2014).

Skor BLEU didapat dari hasil perkalian antara *brevity penalty* dengan rata-rata geometri dari *modified precision score*. *Brevity penalty* digunakan untuk mencegah kalimat pendek memperoleh nilai yang tinggi. Semakin tinggi skor BLEU, maka semakin akurat dengan rujukan. Skor dari BLEU berada pada rentang 0 sampai 1. Suatu terjemahan dikatakan identik dengan terjemahan rujukan jika nilai BLEU sama dengan 1. Rumus diperlihatkan pada persamaan 2.2, 2.3 dan 2.4 berikut (Tanuwijaya and Manurung, 2009).

(2.2)

(2.3)

(2.4)

Keterangan:

BP = *brevity penalty*

c = jumlah kata dari hasil terjemahan otomatis

r = jumlah kata rujukan

𝑃𝑛 = *modified precission score*

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

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

Pengujian yang dilakukan dengan melibatkan responden yaitu pengujian hasil penerjemahan mesin translasi oleh ahli bahasa. Pengujian hasil penerjemahan mesin translasi oleh ahli bahasa digunakan untuk mengetahui apakah hasil terjemahan dari mesin penerjemah statistik mendekati atau bahkan sudah sesuai dengan terjemahan dari ahli bahasa. Pengujian secara manual merupakan metode pengujian yang memiliki tingkat akurasi paling baik, meskipun akan memakan waktu yang lama karena dilakukan secara manual oleh ahli bahasa. (Dharmawan, Sujaini and Muhardi, 2020)

### Pengujian Fungsionalitas Aplikasi

Pengujian fungsional aplikasi dilakukan dengan metode *black box* menggunakan teknik *robust testing*. *Black box* merupakan salah satu metode pengujian perangkat lunak yang digunakan untuk menemukan kesalahan dan mendemonstrasikan fungsional aplikasi saat dioperasikan, apakah *input* diterima dengan benar dan *output* yang dihasilkan telah sesuai dengan yang diharapkan (Hidayat, Sujaini and Dwinyoto, 2015). *Robustness Testing* adalah pengujian dengan data *input* dipilih di luar spesifikasi yang telah didefinisikan dan tujuan dari pengujian ini adalah membuktikan bahwa tidak ada kesalahan jika masukan tidak valid.

# METODOLOGI PENELITIAN

## Data dan Alat Bantu Penelitian

### Data Penelitian

Data penelitian yang digunakan berupa dokumen teks *chatting* berbahasa Melayu Sambas yang diambil aplikasi WhatsApp dan berisi percakapan bahasa Melayu Sambas. Dokumen teks *chatting* ini kemudian dilakukan pelabelan, yaitu menandai setiap kata dengan kata benar dan perbaikan pada kalimat teks *chatting* yang tidak benar (tidak baku), sehingga menghasilkan korpus *tagging* untuk pra proses HMM dan korpus bahasa Melayu Sambas.

Korpus bahasa Melayu Sambas diterjemahkan ke bahasa Indonesia sehingga menghasilkan Korpus bahasa Indonesia. Korpus bahasa Melayu Sambas dan korpus bahasa Indonesia ini dijadikan sebagai korpus paralel untuk mesin translasi Moses Decoder. Jumlah korpus paralel yang dibuat sebanyak 4500 pasang kalimat.

### Alat Bantu Penelitian

Alat yang digunakan pada penelitian ini terdiri dari alat perancangan penelitian, perangkat keras, dan perangkat lunak. Berikut akan dipaparkan apa saja alat perancangan penelitian, perangkat keras, dan perangkat lunak yang digunakan:

1. Alat Perancangan Penelitian

Alat penelitian yang digunakan dalam penelitian ini adalah *Unified Modelling Language* (UML). UML digunakan untuk menggambarkan batasan sistem dan fungsi-fungsi sistem secara umum, menggambarkan kegiatan atau proses yang dilaksanakan secara umum, menggambarkan representasi struktur sistem. Diagram yang akan digunakan dalam pengembangan sistem ini adalah *use case diagram*, *class diagram*, *statechart diagram*, dan *sequence diagram*.

1. Perangkat Keras

Adapun perangkat keras yang digunakan untuk merancang aplikasi yang akan dibangun adalah sebagai berikut:

1. Satu unit laptop Dell Latitude 3580 dengan spesifikasi, Intel Core i5-7200 2.50GHz, 8.00GB RAM DDR4, Intel HD Graphic 620, Harddisk 1TB.
2. Smartphone Samsung A5 2017 dengan spesifikasi, Octa-core 1.9 GHz Cortex-A53, Mali-T830MP3, 3GB RAM + 32GB ROM, Android versi 8.0.0, LTE Support.
3. Perangkat Lunak

Adapun perangkat lunak yang digunakan untuk merancang aplikasi yang akan dibangun adalah sebagai berikut:

1. Alibaba ECS (Elastic Cloud Server)

Alibaba ECS (Elastic Cloud Server) digunakan sebagai *cloud server* pada penelitian ini, dengan spesifikasi RAM 1 GiB dan penyimpanan 40 GiB.

1. Sistem Operasi Linux Ubuntu Server 16.04 LTS

Sistem operasi Linux Ubuntu Server 16.04 LTS digunakan sebagai sistem operasi pada server dan untuk instalasi serta konfigurasi mesin penerjemah pada server.

1. Windows 10 Pro 64-bit

Sistem operasi Windows 10 Pro 64-bit digunakan untuk pengerjaan aplikasi android, pengerjaan penulisan, dan perancangan *Unified Modelling Language* (UML).

1. Apache 2.4.41

Apache 2.4.41 digunakan sebagai *web server* dan digunakan pada sisi *server*.

1. PHP 7.4

PHP 7.4 digunakan sebagai *web service* yang menghubungkan antara aplikasi android dengan *server*. *Web service* digunakan agar pesan teks dapat dilakukan pra proses dan penerjemahan pesan oleh *server*.

1. MobaXterm v12.1

MobaXterm digunakan untuk akses ssh dan ftp pada *server* yang kemudian digunakan untuk mengunggah file korpus dan file PHP ke server. MobaXterm juga digunakan untuk melakukan instalasi dan konfigurasi pada *server*.

1. Firebase

Firebase digunakan sebagai *socket realtime* pada sisi aplikasi android, *realtime database*, otentikasi pengguna, dan digunakan untuk membuat notifikasi pesan masuk. Firebase ini yang mengirimkan pesan yang sudah diterjemahkan oleh mesin ke penerima pesan.

1. IPostagger

IPostagger digunakan untuk melakukan pra proses HMM dengan *trigram* terhadap pesan teks yang ditulis oleh pengirim pesan. Pesan yang sudah dilakukan pra proses kemudian akan dikirimkan ke *server* Moses Decoder.

1. NLTK (Natural Language Toolkit)

NLTK merupakan *tools* yang akan digunakan untuk mencari nilai *accuracy, precision, recall*, dan *f-measure* dari hasil pra proses *Ngram*.

1. PyCharm Community

PyCharm Community merupakan IDE khusus bahasa pemrograman python. PyCharm Community digunakan untuk menjalankan atau mengeksekusi perintah NLTK untuk mencari nilai akurasi pra proses HMM.

1. Moses Decoder

Moses Decoder digunakan sebagai mesin penerjemah yang akan menerjemahkan pesan teks berbahasa Melayu Sambas yang sudah dilakukan praposes sehingga menjadi pesan teks berbahasa Indonesia.

1. Kenlm

Kenlm digunakan sebagai bahasa pemodelan dalam men-*training* korpus bahasa Melayu Sambas dan bahasa Indonesia. Hasil dari pemodelan bahasa adalah probabilitas kemunculan kata yang nantinya akan dijadikan sebagai referensi dalam proses penerjemahan.

1. Giza++

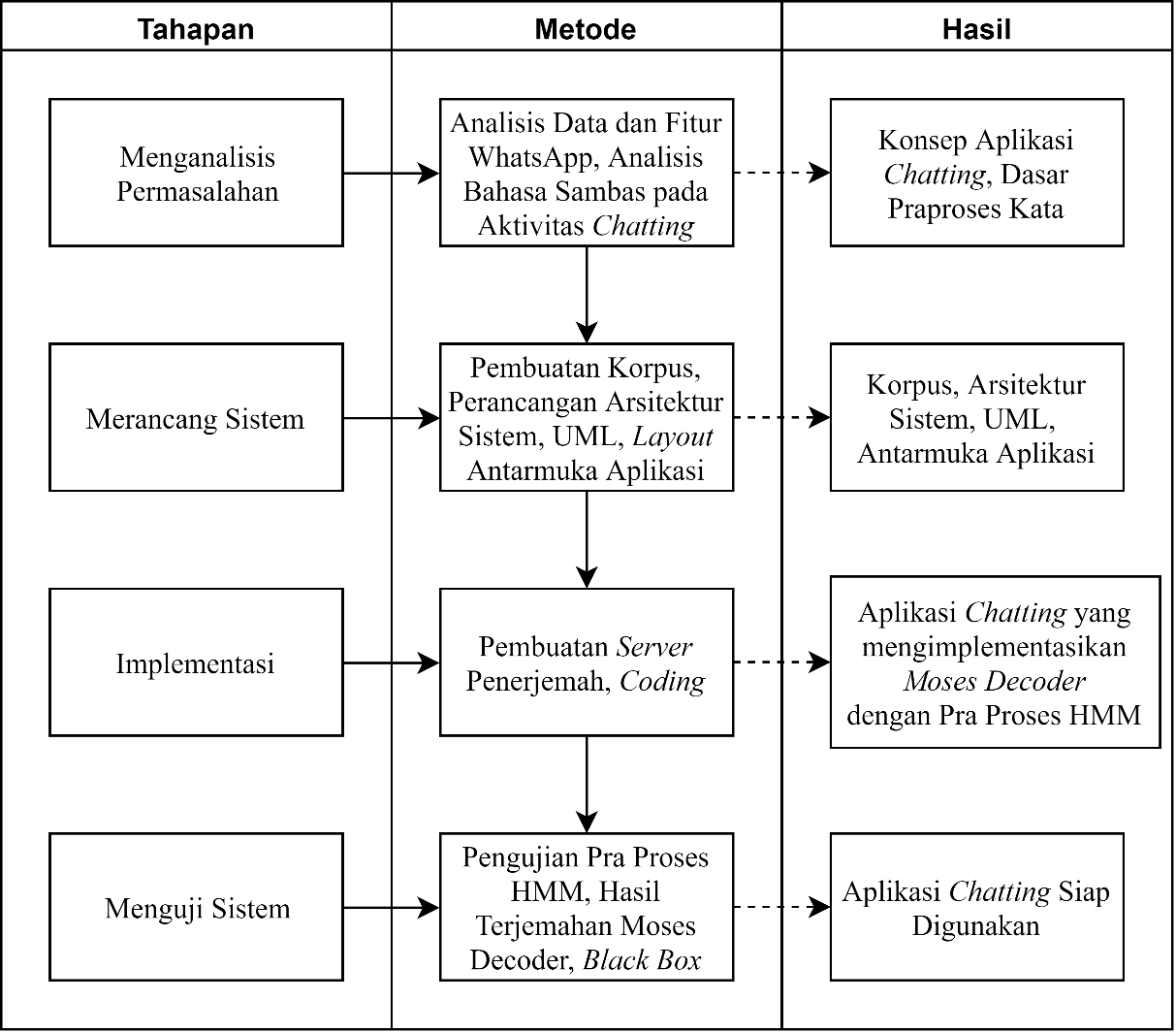
Giza++ digunakan untuk men-*training* korpus paralel sehingga menghasilkan vocabulary corpus, word alignment, dan tabel model translasi. Vocabulary corpus berisi pasangan kata-kata sumber dan target masing-masing beserta jumlah kemunculan kata dalam korpus. Word alignment berisi susunan kalimat sumber terhadap kalimat target. Sedangkan model translasi berisi padanan kata antara bahasa sumber dengan bahasa target yang memiliki nilai probabilitas.

1. Android Studio 3.5.1

Android Studio 3.5.1 digunakan untuk membangun aplikasi *chatting* berbasis Android yang mengimplementasikan mesin penerjemah statistik dengan Moses Decoder.

## Metodologi Penelitian

Metodologi penelitian dilakukan dengan tujuan untuk memberi gambaran terkait aktivitas penelitian yang dilakukan. Penelitian ini juga dilakukan dalam tiga proses yaitu tahapan, metode, dan hasil. Gambar 3.1 berikut menggambarkan metodologi penelitian yang akan dilakukan.



Gambar 3.1 Metodologi Penelitian

Berdasarkan metodologi penelitian pada Gambar 3.1 maka dapat dilihat beberapa tahapan dalam penelitian ini. Tahapan-tahapan tersebut terdiri dari:

1. Analisis Permasalahan

Analisis permasalahan merupakan tahapan untuk mengenal atau merinci elemen-elemen dalam penelitian guna dapat memperoleh informasi dan menentukan dasar-dasar dalam penelitian dalam hal ini dasar-dasar dalam penerjemahan bahasa. Analisis permasalahan ini akan dilakukan dengan mengidentifikasi karakteristik aplikasi *chatting*, dan kemudian mengidentifikasi karakteristik bahasa Melayu Sambas.

1. Perancangan Sistem

Perancangan sistem merupakan tahapan dalam menggambarkan kebutuhan sistem dan aktivitas yang terjadi dalam sistem, yang dalam penelitian ini menggunakan diagram *Unified Modelling Language* (UML). Pada tahap ini juga melakukan desain antarmuka dari aplikasi *chatting*.

1. Implementasi

Tahapan selanjutnya setelah perancangan sistem adalah proses implementasi. Pada proses implementasi dilakukan pembuatan server penerjemah dan proses pemrograman (*coding*). Instalasi dan konfigurasi mesin HMM dan Moses Decoder dilakukan di server penerjemah, sedangkan proses pemrograman dilakukan dengan membuat perangkat lunak berbasis android yang akan digunakan sebagai aplikasi *chatting* serta membuat *web service* untuk mengakses server penerjemah.

1. Pengujian Sistem

Pengujian dilakukan untuk mengetahui apakah sistem yang telah dibuat dapat mengimplementasikan mesin penerjemah statistik berbasis Android dengan Moses Decoder. Pengujian dilakukan dengan menguji tingkat akurasi hasil perbaikan kata atau pra proses HMM, hasil terjemahan mesin translasi dan keberhasilan fungsionalitas aplikasi.

## Analisis Permasalahan

Dalam pengembangan aplikasi *chatting* yang mengimplementasikan mesin penerjemah bahasa terdapat beberapa unsur atau elemen yang harus dilakukan pendalaman sebelumnya seperti identifikasi karakteristik aplikasi *chatting* dan karakteristik bahasa yang digunakan. Pada penelitian ini peneliti melakukan identifikasi karakteristik dari salah satu aplikasi *chatting* yaitu aplikasi WhatsApp, sedangkan bahasa yang digunakan sebagai objek penerjemahan adalah bahasa Melayu Sambas.

### Analisis Data dan Fitur WhatsApp

WhatsApp merupakan sebuah aplikasi perpesanan instan dan lintas platform yang digunakan untuk melakukan obrolan antara dua orang atau lebih dalam aplikasi. Secara umum WhatsApp hampir sama dengan SMS (*Short Message Service*) hanya saja aplikasi WhatsApp tidak mengirim pesan menggunakan pulsa melainkan mengirimkan pesan menggunakan koneksi internet pada perangkat. Selain itu pada aplikasi WhatsApp sendiri memiliki beberapa karakteristik seperti:

1. Dapat berjalan di berbagai jenis perangkat

Aplikasi WhatsApp dapat berjalan di beberapa jenis perangkat baik pada perangkat android dan iOS maupun perangkat desktop seperti Windows dan macOS. Aplikasi WhatsApp memiliki versi yang berjalan di *website* yang disebut dengan WhatsApp Web yang mana memiliki fungsi yang sama dengan versi aplikasi utama sebagai aplikasi untuk berkirim pesan namun dengan fitur yang lebih terbatas. WhatsApp Web sendiri berjalan beriringan dengan aplikasi WhatsApp yang terinstal pada perangkat utama (*smartphone*), sehingga pengguna harus tetap terhubung dengan aplikasi WhatsApp utama untuk dapat menggunakan WhatsApp Web. Namun kehadiran WhatsApp Web sendiri sangat mempermudah pengguna karena memungkinkan pengguna untuk mengakses akun WhatsApp pada perangkat *smartphone* melalui *browser* pada perangkat desktop.

Untuk aplikasi WhatsApp yang dapat di install pada perangkat desktop sendiri sebenarnya merupakan versi yang sama dengan WhatsApp Web baik dari segi tampilan dan juga fitur, sehingga harus tetap terhubung dengan aplikasi WhatsApp utama untuk dapat digunakan.

1. Memiliki dua jenis ruang obrolan

Pada aplikasi WhatsApp terdapat dua jenis ruang obrolan yang dapat dipilih atau digunakan oleh pengguna yaitu ruang obrolan pribadi (*private chat*) dan ruang obrolan grup (*group chat*). Pada ruang obrolan pribadi satu pengguna hanya dapat berinteraksi langsung atau berkirim pesan dengan satu pengguna lain dalam satu waktu. Namun ketika melakukan panggilan baik berupa panggilan suara atau panggilan video, pengguna selain dapat melakukan panggilan pribadi ke satu pengguna lain juga dapat menambahkan beberapa kontak lain untuk bergabung pada panggilan.

Sedangkan pada ruang obrolan grup, satu pengguna dapat berinteraksi dengan beberapa pengguna lain sekaligus dalam satu waktu. Jumlah maksimal peserta atau pengguna yang dapat bergabung pada satu ruang obrolan grup sendiri adalah sejumlah 256 orang pengguna aplikasi WhatsApp, sedangkan untuk ruang obrolan grup sendiri dapat dibuat dengan jumlah peserta minimal 2 orang pengguna aplikasi WhatsApp. Peserta dalam ruang obrolan grup dapat dipilih langsung oleh pembuat grup, ditambahkan oleh admin grup setelah ruang obrolan grup dibuat, ataupun melalui undangan via tautan yang dapat dibuat oleh admin grup.

Orang yang membuat ruang obrolan grup secara otomatis akan menjadi admin grup dan memiliki beberapa kewenangan khusus dalam ruang obrolan grup. Beberapa kewenangan khusus tersebut seperti menambahkan peserta ke dalam grup, mengeluarkan peserta dari dalam grup, mengatur setelan grup, serta menjadikan peserta sebagai admin grup dan menjadikan admin grup sebagai peserta biasa. Pada setelan grup, admin dapat membatasi agar hanya admin yang dapat mengirim pesan di grup dan membatasi agar hanya admin yang dapat mengubah informasi dasar. Admin grup pertama atau orang yang membuat grup tidak dapat dikeluarkan dari grup oleh admin grup lain dan sebaliknya admin pertama dapat menjadikan admin grup sebagai peserta biasa bahkan mengeluarkannya dari grup.

Secara umum peserta yang tergabung dalam sebuah ruang obrolan grup dapat melakukan tindakan seperti berinteraksi dengan peserta lain di dalam grup ataupun memilih keluar dari dalam grup.

Selain memiliki kedua ruang obrolan yang sudah disebutkan sebelumnya, aplikasi WhatsApp dapat melakukan pesan siaran/*broadcast* yang mana pengguna dapat mengirimkan pesan yang sama ke beberapa kontak sekaligus dalam satu waktu. Untuk satu ruang *broadcast* atau ruang siaran pengirim dapat menambahkan kontak paling banyak hingga 256 kontak dan yang dapat menerima pesan *broadcast* hanyalah kontak/pengguna yang juga menyimpan/memiliki nomor pengirim pada kontak mereka. Pesan yang dikirim melalui ruang *broadcast* juga secara otomatis akan masuk ke ruang obrolan pribadi sehingga jika penerima *broadcast* membalas pesan maka pesan balasan hanya akan dikirim ke ruang obrolan pribadi.

1. Mengirim berbagai jenis pesan

Aplikasi WhatsApp mampu melakukan beberapa jenis pengiriman pesan yaitu berupa pengiriman pesan teks dan pengiriman file seperti gambar/foto, video, audio, dokumen, kontak, hingga pengiriman pesan berupa lokasi yang dapat terhubung dengan aplikasi peta pada perangkat. Selain itu pada percakapan di ruang obrolan pengguna juga dapat menggunakan emoji, animasi (GIF), dan stiker untuk dapat lebih mengekspresikan diri dalam ruang obrolan. Beberapa jenis pesan yang dapat dikirim dengan aplikasi WhatsApp diperlihatkan pada tabel 3.1 berikut.

Tabel 3.1 Jenis Pesan pada Aplikasi WhatsApp

| **Jenis Pesan** | **Ukuran Maksimal**  **Pesan per Pengiriman** | **Keterangan** |
| --- | --- | --- |
| Pesan teks | 65.536 karakter | * Menggunakan pengkodean karakter Unicode * Mendukung penggunaan tulisan bahasa lain seperti tulisan bahasa Arab dan tulisan bahasa Mandarin * Mendukung penggunaan *emoticon* * Teks berupa tautan memiliki aksi untuk dapat diakses melalui *browser* atau melalui aplikasi terkait |
| Pesan file media | 16 MB (*megabyte*) | * File media berupa audio, video, foto/gambar, stiker dan animasi |
| Pesan file dokumen | 100 MB (*megabyte*) | * Dapat mengirim file berbagai jenis file diantaranya dokumen seperti file *word*, *excel*, *pdf* dan file dokumen lainnya * Dapat mengirim file media berupa audio, video dan foto/gambar * Dapat mengirim file aplikasi (apk) dan file kompres (*.rar* atau *.zip*) |
| Pesan kontak | 257 kontak | * Daftar kontak yang bisa dikirim berdasarkan kontak pada penyimpanan perangkat pengirim |
| Pengiriman lokasi | 1 lokasi | * Pesan lokasi ditautkan dengan aplikasi peta/*maps* dari perangkat |

Pada pengiriman pesan teks pengguna dapat mengirimkan pesan berupa huruf, angka, simbol, dan emoji dengan panjang tulisan maksimal hingga 65.536 karakter dalam satu pesan. Aplikasi WhatsApp menggunakan pengkodean karakter Unicode sehingga mendukung penggunaan tulisan bahasa asing menggunakan simbol, seperti tulisan bahasa Arab dan tulisan bahasa Mandarin. Pada pesan teks yang berupa alamat *website* maka aplikasi WhatsApp dapat mengarahkan pengguna ke *browser* tertentu pada perangkat untuk mengakses atau membuka situs *website* tersebut.

Untuk pengiriman gambar/foto yang dikirim maka secara otomatis aplikasi WhatsApp akan mengkompres gambar sehingga kualitas gambar akan menurun dari gambar aslinya. Hal ini dilakukan agar ukuran gambar yang dikirimkan juga semakin kecil sehingga penggunaan kuota data atau kuota internet yang diperlukan oleh penerima gambar untuk menerima atau mengunduh gambar tersebut juga lebih kecil, di samping juga menghemat memori penyimpanan perangkat. Untuk pengiriman gambar/foto dapat dilakukan dengan memilih gambar yang akan dikirim pada galeri penyimpanan perangkat atau dengan mengambil gambar/foto secara langsung. Untuk pengiriman gambar/foto dengan ukuran dan kualitas asli/orisinil, pengguna dapat mengirimkan gambar/foto sebagai file dokumen agar gambar/foto tidak dilakukan penurunan kualitas dan ukuran.

Pengiriman file berupa audio dan video ke ruang obrolan pada aplikasi WhatsApp dibatasi dengan ukuran maksimal file sebesar 16 MB (*megabyte*). Seperti hal nya pengiriman gambar/foto, untuk pengiriman audio dan video pengguna dapat mengirimkan file yang tersimpan pada penyimpanan perangkat pengguna ataupun dapat dengan melakukan pengambilan audio dan video secara langsung melalui aplikasi WhatsApp.

Pada pengiriman dokumen ke ruang obrolan, ukuran maksimal satu file dokumen yang dikirim melalui aplikasi WhatsApp adalah sebesar 100 MB (*megabyte*). Selain dapat mengirimkan file dokumen seperti file *word*, *excel*, *pdf* dan file dokumen lainnya, pengiriman dokumen juga dapat mengirimkan bentuk file aplikasi (apk), file kompres (*.rar*/*.zip*), bahkan file media seperti gambar/foto, audio, dan video.

Jenis pengiriman lain yang dapat dilakukan oleh aplikasi WhatsApp adalah pengiriman kontak dan pengiriman lokasi. Untuk pengiriman kontak pengguna dapat mengirimkan kontak teman yang tersimpan pada perangkat pengguna ke pengguna lain sehingga penerima dapat menghubungi atau menyimpan kontak yang sudah dikirimkan. Sedangkan untuk pengiriman pesan lokasi, aplikasi dapat mengirimkan lokasi berupa lokasi pengguna pengguna, rekomendasi lokasi di sekitar pengguna, dan pengiriman berupa lokasi pengguna secara *realtime*. Pesan lokasi tersebut nantinya dapat diakses menggunakan aplikasi peta bawaan dari perangkat pengguna ataupun dapat menggunakan aplikasi lain yang terpasang pada perangkat pengguna.

Semua jenis pengiriman pesan baik berupa teks, gambar/foto, audio, video, dokumen, kontak, ataupun lokasi pada aplikasi WhatsApp menggunakan pengiriman melalui koneksi internet pada perangkat pengguna sehingga tidak akan mengurangi kuota pulsa pengguna secara langsung, kecuali pengguna membebankan penggunaan koneksi internetnya pada kuota pulsa.

1. Penanganan percakapan

Aplikasi WhatsApp mengandalkan koneksi internet pada perangkat untuk dapat terhubung dengan pengguna aplikasi WhatsApp yang lain, meski begitu aplikasi tidak akan keluar ketika pada perangkat tidak terhubung dengan koneksi internet. Pengguna tetap dapat mengakses aplikasi, melihat kontak, maupun melihat perbincangan dengan pengguna lain walaupun tidak terhubung dengan koneksi internet. Sedangkan ketika perangkat pengguna tidak terkoneksi dengan dengan internet maka pengguna tidak dapat mengirim pesan maupun melakukan panggilan dan tidak dapat menerima pesan maupun panggilan.

Aplikasi WhatsApp memungkinkan pengguna untuk mengetahui status pesan apakah pesan tersebut berhasil dikirim, sudah sampai kepada penerima pesan, sudah dibaca oleh penerima pesan, masih dalam proses pengiriman, ataupun pesan gagal dikirimkan. Aplikasi WhatsApp juga memungkinkan pengguna untuk melihat kondisi dari teman obrolan apakah sedang dalam keadaan terhubung/menggunakan aplikasi (*online*) ataupun kapan terakhir teman terhubung/menggunakan aplikasi.

Dalam ruang obrolan pengirim pesan dapat menghapus pesan yang terlanjur terkirim, baik untuk pengirim sendiri maupun untuk semua pengguna yang ada pada ruang obrolan dengan batas waktu paling lama sekitar 1 jam setelah pesan dikirimkan. Selain itu pengguna dapat meneruskan pesan pada ruang obrolan pada kontak ataupun grup dengan batasan paling banyak 5 kontak atau grup yang dipilih. Selain itu juga pengguna dapat mencari kata pada pesan dalam ruang obrolan, menandai pesan tertentu untuk disimpan (pesan berbintang), mengarsipkan obrolan, serta menandai obrolan agar selalu berada di atas pada daftar obrolan (sematkan obrolan).

1. Akun dan kontak

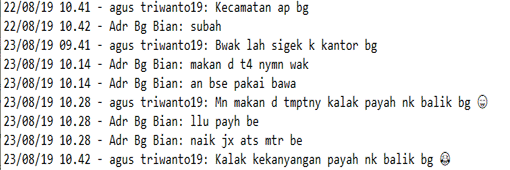
Untuk membuat akun pada aplikasi WhatsApp maka pengguna hanya perlu melakukan registrasi menggunakan nomor telepon aktif yang digunakan oleh pengguna. Pada nomor telepon aktif tersebut nantinya akan dikirimi kode verifikasi dari sistem WhatsApp. Setelah melalui proses verifikasi, pengguna diminta untuk mengatur foto profil dan nama pengguna. Foto profil dan nama pengguna yang sudah diatur dapat diatur kembali nantinya pada pengaturan aplikasi WhatsApp.

Setelah berhasil masuk ke halaman utama aplikasi WhatsApp maka aplikasi akan membaca daftar kontak telepon pada perangkat pengguna. Nomor telepon pada kontak yang terdaftar pada WhatsApp akan secara otomatis ditambahkan ke kontak aplikasi WhatsApp sehingga pengguna tidak perlu menambahkan kontak pada aplikasi WhatsApp secara manual.

1. Ekspor obrolan

Salah satu fitur yang terdapat pada aplikasi WhatsApp adalah fitur ekspor pesan, yaitu fitur untuk menyalin riwayat percakapan pengguna baik pada ruang obrolan pribadi maupun pada ruang obrolan grup. Hasil dari ekspor percakapan pada aplikasi WhatsApp berupa dokumen dengan format \*.txt dan file dengan format \*.vcf untuk pesan berupa kontak. Untuk satu kali ekspor pesan yang tidak menyertakan media (gambar/foto, audio, video, dan lainnya) WhatsApp dapat mengekspor pesan hingga 40.000 pesan dalam satu percakapan.

Pada penelitian ini sendiri penulis akan menggunakan fitur ekspor pesan untuk mengumpulkan data-data percakapan yang diperlukan dalam penelitian. Contoh data hasil ekspor percakapan pada aplikasi WhatsApp diperlihatkan pada gambar 3.2 berikut.



Gambar 3.2 Contoh Data Hasil Ekspor Percakapan Aplikasi Whatsapp

Gambar 3.2 menunjukkan dokumen hasil ekspor percakapan berupa teks yang memuat beberapa informasi yaitu informasi berupa tanggal/waktu pesan dikirim dan diterima, nama pengirim pesan, dan isi pesan yang dikirim dan diterima. Informasi berupa isi pesan yang dikirim dan diterima merupakan data yang akan digunakan peneliti dalam penelitian ini. Analisis informasi isi pesan pada percakapan dalam aplikasi WhatsApp dibahas pada bagian analisis bahasa Melayu Sambas pada aktivitas *chatting*.

### Analisis Bahasa Melayu Sambas pada Aktivitas *Chatting*

Bahasa tidak dapat dipisahkan dalam lingkungan sosial dikarenakan bahasa merupakan alat komunikasi yang digunakan oleh manusia, dengan bahasa seseorang dapat menyampaikan ide, mengenalkan diri, dan menceritakan pengalamannya kepada orang lain.

Bahasa Melayu Sambas adalah bahasa daerah Melayu berdialek Sambas yang digunakan oleh masyarakat etnis Melayu Sambas yang berdomisili di pesisir pantai Utara Kalimantan Barat. Bahasa Melayu Sambas digunakan dalam percakapan masyarakat Melayu Sambas sehari-hari baik di pedesaan maupun di perkotaan baik dalam komunikasi lisan maupun komunikasi tertulis.

Dalam penggunaan bahasa tulis khususnya di media sosial, gejala bahasa abreviasi merupakan terobosan baru untuk berkomunikasi. Berkomunikasi yang diwujudkan melalui abreviasi dalam media sosial telah menuntun pada perubahan pemakaian bahasa dalam suatu masyarakat. Sifat bahasa yang berupa arbitrer, konvensional, dan dinamis memungkinkan bahasa mengalami perubahan (Cenderamata and Sofyan, 2019). Hal ini juga terjadi pada penggunaan Bahasa Melayu Sambas di masyarakat yang cenderung memendekkan kata saat berkomunikasi melalui media sosial dengan tujuan menghemat waktu pengetikan.

Proses pemendekan kata atau abreviasi dapat berdampak positif dan dapat pula berdampak negatif. Bentukan-bentukan bahasa baru yang dihasilkan dari proses pemendekan dapat memiliki dampak positif pada satu sisi memperkaya khasanah kekayaan bahasa, seperti kosakata jika dalam praktiknya tidak menghambat proses komunikasi. Akan tetapi, tidak menutup kemungkinan dapat menimbulkan dampak negatif jika gejala tersebut malah menghambat proses komunikasi.

Selain penggunaan abreviasi atau penyingkatan kata pada penggunaan bahasa tulis atau aktivitas *chatting*, juga terdapat bentuk-bentuk lain yang digunakan seperti penggunaan slang dan pemanjangan kata. Contoh dari penggunaan abreviasi atau penyingkatan kata pada aktivitas *chatting* bahasa Melayu Sambas seperti pada tabel 3.2 berikut.

Tabel 3.2 Tabel Penggunaan Abreviasi

| **Pesan Teks** | **Pesan Asli** |
| --- | --- |
| Bwak lah sigek k kantor bg | Bawak lah sigek ke kantor bang |
| Mn makan d tmptny kalak payah nk balik bg | Mun makan di tempatnye kalak payah nak balik bang |
| Jak blm buat SK juak, jadi blm sah | Jak balom buat Surat Keterangan juak, jadi balom sah |
| Aku nk molah KTM tok | Aku nak molah Kartu Tanda Mahasiswa tok |
| Tpi aku rase becarek die msh sian lalu brek kabr | Tapi aku rase becarek die maseh sian lalu barek kabar |
| Aku tgl blsan kali brok blik | Aku tanggal belasan kali barok balik |
| Kmek pndah kn k hp Lok lh pn dh slsai tok | Kamek pindah kan ke *handphone* dolok lah pun dah selasai tok |
| Tp mn rmh lmak kmek pun dh masok aek k rmh y | Tapi mun rumah lamak kamek pun dah masok aek ke rumah ye |
| Tp barat2 ny kuliah iye lh perjuangan, mun mudah rmai dah yg kuliah | Tapi barat-barat nye kuliah iye lah perjuangan, mun mudah ramai dah yang kuliah |
| Sblm ny sih mseh jln juak, pling bnyk ny aek cmn nympai stngah bak jk | Sebalom nye sih maseh jaln juak, paling banyak nye aek cuman nyampai setangah bak jak |

Penggunaan abreviasi atau penyingkatan kata merupakan suatu yang paling umum terjadi pada penggunaan bahasa tulis dalam hal ini aktivitas *chatting* bahasa Melayu Sambas. Seperti pada tabel 3.2 dapat dilihat pola penyingkatan yang umum terjadi pada aktivitas *chatting* bahasa Melayu Sambas adalah :

Menghilangkan beberapa atau semua huruf-huruf vokal pada kata seperti pada kata “bwak” yang berarti “bawak” dan kata “blm” yang berarti “balom”.

Menghilangkan beberapa huruf baik di awal ataupun di akhir kata untuk mewakili satu kata seperti pada kata “lom” yang berarti “balom”.

Menggabungkan beberapa kata menjadi satu kata seperti pada kata “KTM” yang berarti “Kartu Tanda Mahasiswa”.

Penggunaan angka atau karakter sebagai ganti kata ulang seperti pada kata “barat2” yang berarti “barat-barat”.

Bentuk slang dapat dikatakan sebagai ragam bahasa nonformal yang biasa dipakai oleh kelompok sosial tertentu, dicirikan dengan kemunculan kosakata baru dan cepat berubah yang umumnya berupa satuan ekspresi atau kata-kata yang sudah mengalami berbagai jenis perubahan bentuk dan makna (Cenderamata and Sofyan, 2019). Contoh penggunaan slang pada aktivitas *chatting* bahasa Melayu Sambas seperti pada tabel 3.3 berikut.

Tabel 3.3 Tabel Penggunaan Slang

| **Pesan Teks** | **Pesan Asli** |
| --- | --- |
| Pakai bis x | Pakai bis kali |
| Sx sampi rmh an llu berisek blong | Sekali sampai rumah daan lalu berisek blong |
| Kite serate agek jk lah jem | Kite serate agek jak lah jem (nama/panggilan untuk lawan bicara) |
| Makan d t4 nymn wak | Makan di tempat nyaman wak (panggilan untuk lawan bicara) |
| Malah dah nak otw ke rumah drek tok | Malah udah nak *on the way* ke rumah direk tok |
| Bloem tau ngpe drek nk maing juak ke, kite bile futsal tok | Balom tau ngape direk nak maing juak ke, kite bile futsal tok |
| Usah balik dlok Jak ndx an narik ye bhe | Usah balik dolok Jak indak an narik ye be |
| Tapi pun direk maok daan ape sitok ajak ho | Tapi pun direk maok daan ape sitok ajak ho (panggilan untuk lawan bicara) |
| Tdek aq dah nak pagi u lamak balasnya e asrma agek aq | Tadek aku dah nak pagi kau lamak balasnya ke asrama agek aku |
| Mane ku wa bpk d read ny nk ny | Mane aku *whatsapp* bapak di bace nye nak nye |

Selain abreviasi, penggunaan slang juga merupakan sesuatu yang umum atau sering terjadi pada suatu aktivitas *chatting*, dan setiap variasi bentuk slang dapat berbeda-beda dan berubah-ubah bergantung pada penutur yang menggunakan bahasa. Berdasarkan tabel 3.3 dapat diketahui beberapa variasi penggunaan slang pada bahasa Melayu Sambas yang biasa digunakan dalam aktivitas *chatting* seperti :

Merupakan bentuk abreviasi atau penyingkatan kata yang disertai dengan penggunaan huruf atau angka sebagai pengganti suatu kata atau pengganti beberapa huruf dalam kata, seperti pada kata “kali” yang diganti dengan “x”, kata “Sekali” diganti “Sx”, kata “aku” diganti “aq”, ataupun pada kata “tempat” yang diganti dengan “t4”.

Penggunaan nama atau panggilan tertentu terhadap lawan bicara yang mana panggilan tersebut sama-sama dimengerti oleh pelaku komunikasi/*chatting* seperti penggunaan kata “jem” yang merupakan nama panggilan lawan bicara maupun penggunaan kata panggilan “ho” atau “wak” yang berarti “kawan”.

Penggunaan bahasa/istilah asing baik yang mengalami penyingkatan kata ataupun tidak seperti kata “otw” yang merupakan singkatan dari “*on the way*” yang berarti “sedang dalam perjalanan”, kata “wa” yang berarti “*whatsapp*” (nama aplikasi perangkat lunak), maupun kata “*read*” yang berarti “baca”.

Penyisipan huruf dalam suatu kata dengan maksud untuk menambahkan bunyi pada kata ketika dibaca, seperti penambahan huruf “h” pada kata “be” sehingga menjadi “bhe” maupun huruf “e” pada kata “balom” sehingga “baloem”.

Pemanjangan kata merupakan suatu variasi dalam suatu aktivitas *chatting* yang mana bentuk ini biasa digunakan sebagai wujud dalam mengekspresikan suatu ungkapan tertentu. Beberapa contoh penggunaan kata dalam aktivitas *chatting* bahasa Melayu Sambas dapat dilihat seperti pada tabel 3.4 berikut.

Tabel 3.4 Tabel Penggunaan Pemanjangan Kata

| **Pesan Teks** | **Pesan Asli** |
| --- | --- |
| Suroh jem bace pesan kamek itokkkkk | Suroh jem bace pesan kamek itok |
| Soalnye kapasitasnye kacik inyan di hape yeeee | Soalnye kapasitasnye kacik inyan di *handphone* ye |
| Soalnye tadek Iis agek jumpe dosennnn | Soalnye tadek Iis agek jumpe dosen |
| Oooh hrus pkai gy k die rti y | Oh harus pakai gaye ke die rati ye |
| Aok lahh wak nak gile lah ny, langsong nyarekkk | Aok lah wak nak gile lah nye, langsong nyarek |

Seperti yang dapat dilihat pada tabel 3.4 beberapa contoh pemanjangan kata dilakukan dengan menulis suatu huruf dalam kata beberapa kali dan pemanjangan kata digunakan dengan tujuan untuk menggambarkan ekspresi penulis ketika kata yang dituliskan dibaca oleh penerima pesan. Pemanjangan kata memberikan kesan penegasan terhadap kata yang dituliskan dan variasi pemanjangan kata dapat berupa pemanjangan huruf vokal ataupun huruf konsonan.

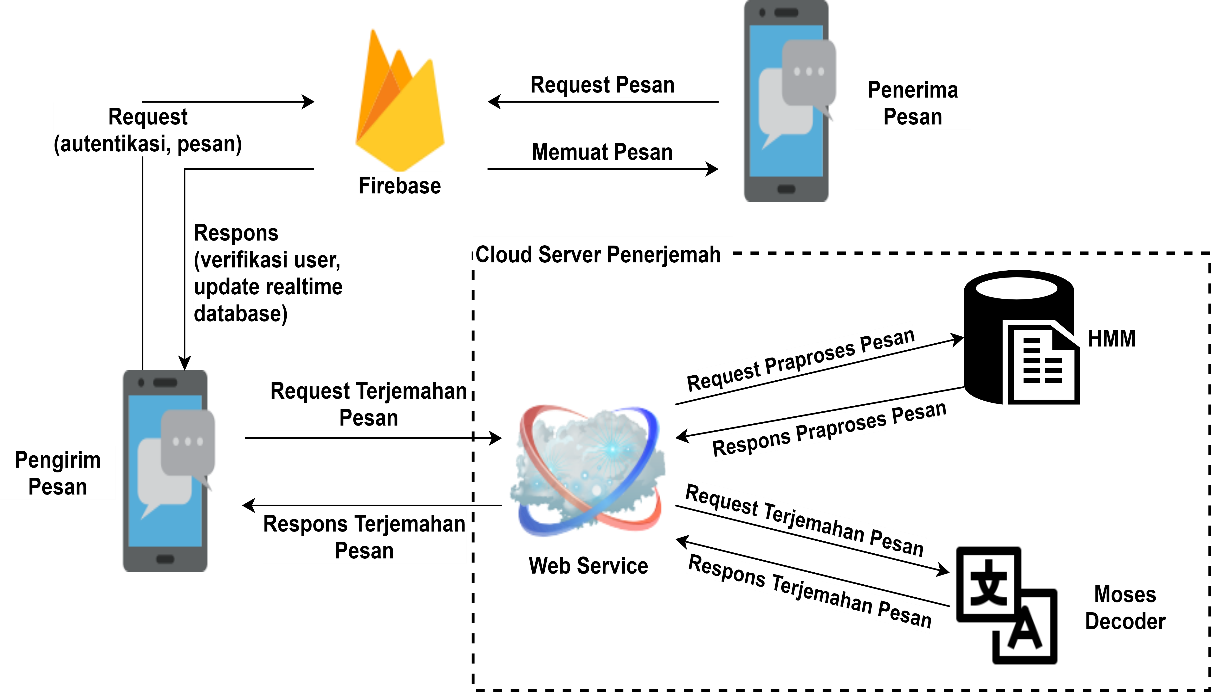
Setiap orang/pelaku aktivitas *chatting* umumnya memiliki variasi kebiasaan penulisan bahasa yang berbeda-beda dan dapat berubah-ubah berdasarkan faktor atau kondisi-kondisi tertentu. Hal ini juga terjadi pada percakapan atau *chatting* dalam bahasa Melayu Sambas, namun secara umum bentuk variasi yang digunakan oleh penutur atau pelaku *chatting* merupakan bentuk abreviasi, penggunaan slang maupun pemanjangan kata seperti yang telah diuraikan di atas.

## Perancangan Sistem

Perancangan sistem dilakukan dengan tujuan untuk memberikan gambaran tentang sistam yang akan dibuat. Tahapan dalam perancangan ini dilakukan dengan menggunakan *Unified Modeling Language* (UML) yang terdiri dari *use case diagram*, *class diagram*, *statechart diagram*, dan *sequence diagram*. Perancangan tentang bagaimana sistem ini berjalan akan digambarkan arsitektur sistem.

### Perancangan Arsitektur Sistem

Arsitektur sistem merupakan gambaran garis besar cara kerja sistem yang akan dibangun melalui model-model yang saling berhubungan. Gambaran arsitektur sistem aplikasi *chatting* penerjemah yang akan dibangun diperlihatkan pada gambar 3.3 berikut.



Gambar 3.3 Arsitektur Sistem

Pada gambar 3.3 dapat dilihat cara kerja sistem yang memiliki beberapa tahap proses, yaitu sebagai berikut:

Untuk masuk ke aplikasi, pengguna mengirim autentikasi *login* berupa nomor telepon aktif ke Firebase. Sistem Firebase kemudian mengirimkan respons berupa kode verifikasi berupa SMS ke nomor telepon pengguna. Kode verifikasi yang dikirim kemudian digunakan oleh aplikasi sehingga pengguna dapat mengakses aplikasi secara keseluruhan.

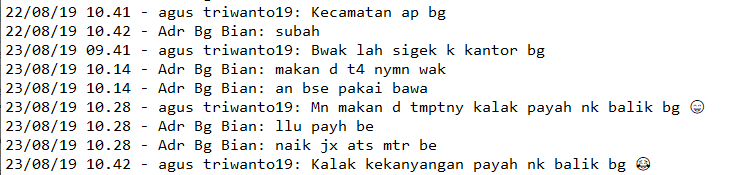
Proses penerjemahan pesan dilakukan dengan mengirimkan *request* terjemahan pesan ke *web service*. Proses penerjemahan pesan dimulai dengan melakukan *request* pra proses pesan ke mesin HMM yang kemudian mengirimkan respon berupa pesan yang sudah diperbaiki/dilakukan pra proses. Hasil pra proses pesan kemudian dilakukan penerjemahan pada Moses Decoder sehingga dihasilkan respons berupa pesan yang sudah diterjemahkan. Pesan yang sudah diterjemahkan dikembalikan ke *web service* yang kemudian akan dikembalikan ke pengirim pesan.

Proses pengiriman pesan dimulai dengan proses penerjemahan pesan, hasil dari proses penerjemahan pesan berupa pesan yang sudah diterjemahkan. Pesan yang sudah diterjemahkan kemudian dikirimkan ke Firebase untuk disimpan pada *realtime database*. Penerima pesan melakukan *load* pesan atau memuat dari *realtime* *database* pesan baru yang diterima.

#### Pembuatan Korpus

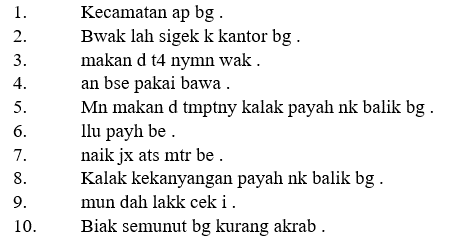
Korpus merupakan data atau dokumen utama yang diperlukan dalam membangun mesin penerjemah statistik. Pada penelitian ini korpus yang digunakan bersumber dari dokumentasi aktivitas *chatting* berbahasa Melayu Sambas. Korpus didapat melalui tiga tahapan yaitu pengumpulan dan pemilihan data teks *chatting*, *tagging* perbaikan teks *chatting*,dan penerjemahan teks *chatting* bahasa Melayu Sambas ke bahasa Indonesia secara manual.

Pengumpulan data pada penelitian ini dilakukan dengan mencari sumber data yaitu percakapan dalam bahasa Melayu Sambas sehari-hari oleh penutur bahasa Melayu Sambas. Penulis meminta dokumentasi percakapan kepada penutur Bahasa Melayu Sambas dalam bentuk dokumen teks yang diekspor dari aplikasi WhatsApp. Data dokumentasi percakapan berbahasa Melayu Sambas didapat dari sembilan orang penutur yang berbeda dengan lawan *chatting* yang berbeda-beda. Contoh dokumen teks *chatting* yang didapat dari proses ekspor percakapan pada aplikasi WhatsApp dapat dilihat pada Gambar 3.4 berikut.



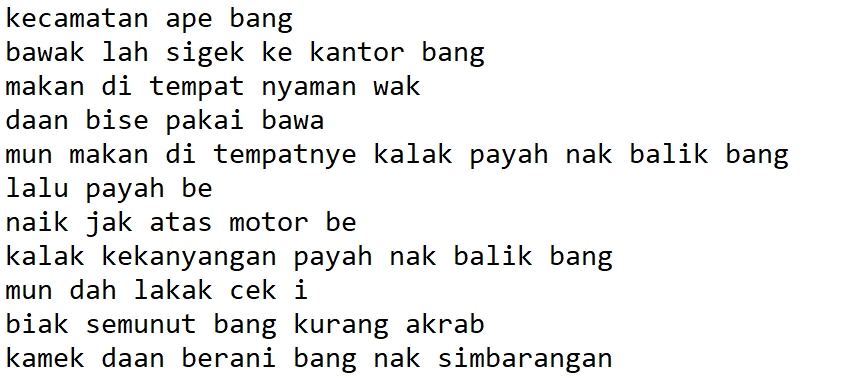
Gambar 3.4 Dokumen Teks *Chatting* WhatsApp

Setelah dokumen teks yang diperlukan terkumpul maka akan dipilih baris-baris teks *chatting* dikumpulkan dalam satu file dokumen teks. Dalam pemilihan baris-baris teks *chatting* penulis mengambil teks percakapan yang nantinya akan digunakan untuk pembuatan korpus *tagging*. Baris-baris teks *chatting* yang dipilih dikumpulkan dalam satu file dokumen teks sehingga menjadi satu korpus teks *chatting* seperti pada gambar 3.5 berikut.



Gambar 3.5 Korpus Teks *Chatting*

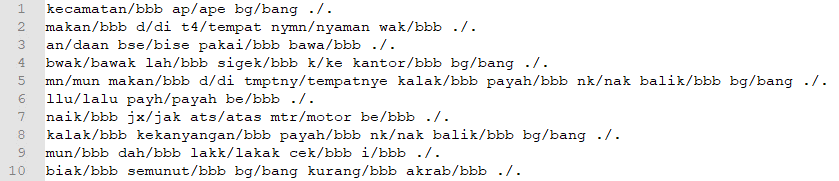
Selanjutnya penulis mencari ahli bahasa yang mampu mengolah dan menerjemahkan korpus teks *chatting* menjadi korpus paralel bahasa Melayu Sambas dan bahasa Indonesia. Korpus bahasa Melayu Sambas disini merupakan perbaikan dari korpus teks *chatting* yang sudah didapat sebelumnya sehingga menjadi bahasa Melayu Sambas yang baik menurut ahli bahasa. Ahli bahasa yang dipilih adalah seorang penutur bahasa Melayu Sambas yang berasal dari kabupaten Sambas. Korpus bahasa Melayu Sambas yang dihasilkan dapat dilihat pada Gambar 3.6 berikut.



Gambar 3.6 Korpus Bahasa Melayu Sambas

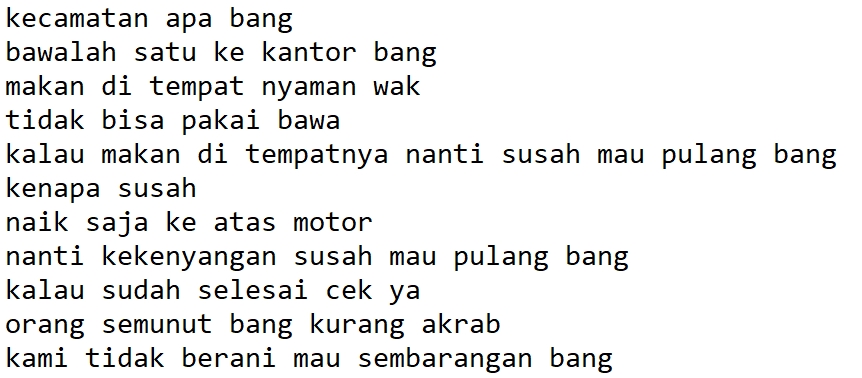
Korpus teks *chatting* dan korpus bahasa Melayu Sambas merupakan korpus yang digunakan untuk membuat korpus *tagging*, korpus *tagging* sendiri adalah korpus yang akan digunakan oleh mesin HMM dalam membuat mesin/layanan pra proses bahasa. Korpus *tagging* didapat dengan cara menandai kata per kata pada korpus teks *chatting* terhadap korpus bahasa Melayu Sambas.

Apabila kata pada dokumen teks *chatting* sama dengan kata pada korpus bahasa Melayu Sambas maka penulis menandai dengan huruf “bbb”. Adapun jika padanan kata tidak sama, maka kata pada dokumen teks *chatting* akan ditandai dengan kata pada korpus bahasa Melayu Sambas. Untuk lebih jelasnya korpus *tagging* yang dihasilkan dapat dilihat pada Gambar 3.7 berikut.



Gambar 3.7 Korpus *Tagging*

Korpus yang terakhir akan dibuat adalah korpus bahasa Indonesia. Sumber korpus bahasa Indonesia disini adalah hasil dari penerjemahan korpus bahasa Melayu Sambas ke bahasa Indonesia oleh ahli bahasa. Korpus paralel bahasa Melayu Sambas dan bahasa Indonesia yang didapat nantinya akan digunakan sebagai sumber dalam penerjemahan bahasa. Korpus bahasa Indonesia yang dihasilkan dapat dilihat pada Gambar 3.8 berikut.



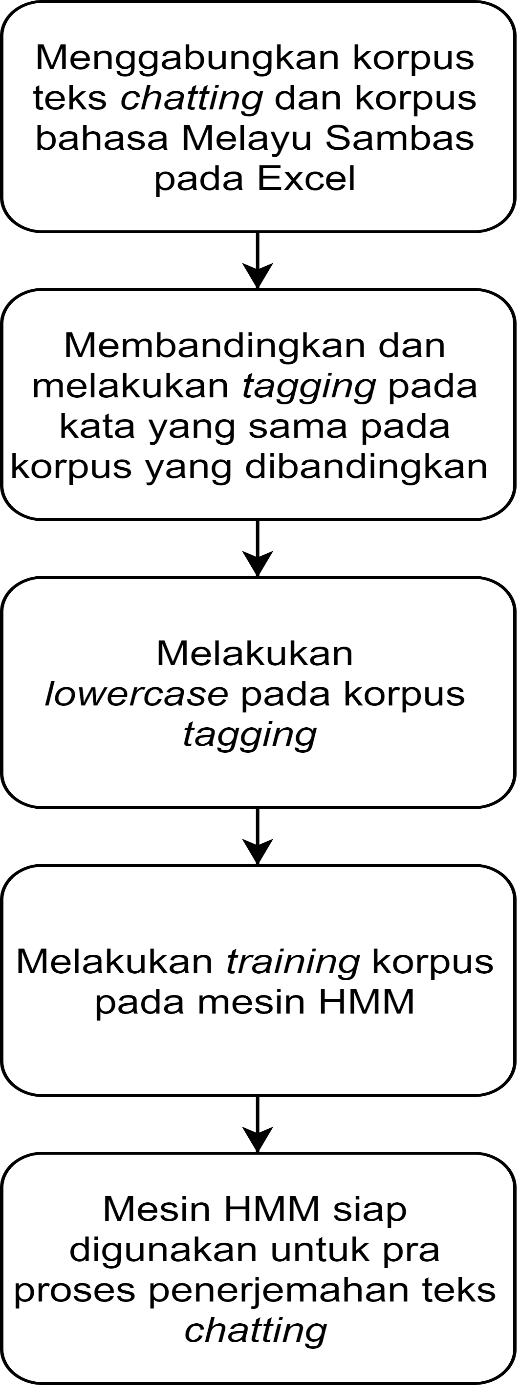
Gambar 3.8 Korpus Bahasa Indonesia

Hasil akhir korpus yang digunakan dalam penelitian ini ada dua jenis korpus yaitu korpus *tagging* dan korpus paralel bahasa Melayu Sambas dan bahasa Indonesia. Korpus *tagging* akan digunakan mesin HMM dalam melakukan pra proses terhadap pesan yang akan dikirim. Sedangkan korpus paralel akan digunakan oleh Moses Decoder dalam menerjemahkan pesan yang akan dikirim.

#### Pra Proses HMM

Pra proses HMMdilakukan untuk memperbaiki kalimat *chatting* bahasa Melayu Sambas ke bahasa Indonesia. Pada aplikasi yang dibangun, pra proses HMMbersifat opsional yang artinya pesan dapat dilakukan penerjemahan secara langsung ataupun dilakukan pra proses terlebih dahulu pada pesan baru kemudian dilakukan penerjemahan. Pra proses HMMpada penelitian ini menggunakan *framework* IPostagger.

Pembuatan mesin HMM dimulai dari pembuatan korpus *tagging* hingga dapat dilakukan *training* dan dapat digunakan untuk pra proses penerjemahan teks *chatting* bahasa Melayu Sambas. Pra proses oleh mesin HMM diperlukan untuk memperbaiki bahasa teks *chatting* yang memiliki ragam variasi abreviasi, slang ataupun pemanjangan kata sehingga menjadi bahasa Melayu Sambas yang lebih baik sebelum diterjemahkan ke bahasa Indonesia oleh Moses Decoder. Alur pembuatan mesin HMM yang akan digunakan untuk pra proses penerjemahan bahasa dapat dilihat pada gambar 3.9 berikut.



Gambar 3.9 Alur Pembuatan Mesin HMM

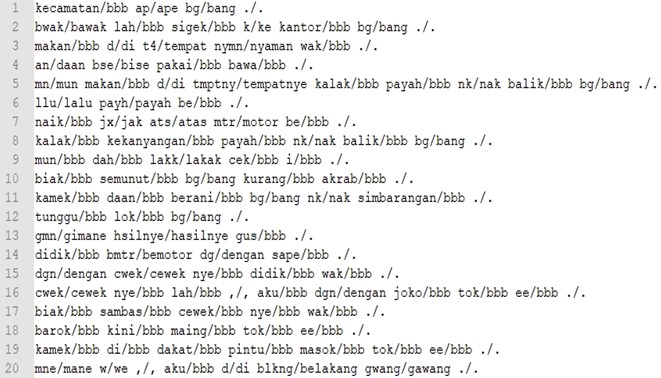
Proses pembuatan mesin HMM dimulai dengan membuat korpus *tagging* yang dilakukan dengan membandingkan korpus teks *chatting* dengan korpus bahasa Melayu Sambas. Korpus teks *chatting* dan korpus Bahasa Melayu Sambas akan dipadankan setiap kata dari kedua korpus dan digabung ke dalam satu buah *file* *excel* seperti pada gambar 3.10 berikut.



Gambar 3.10 Perbandingan Korpus

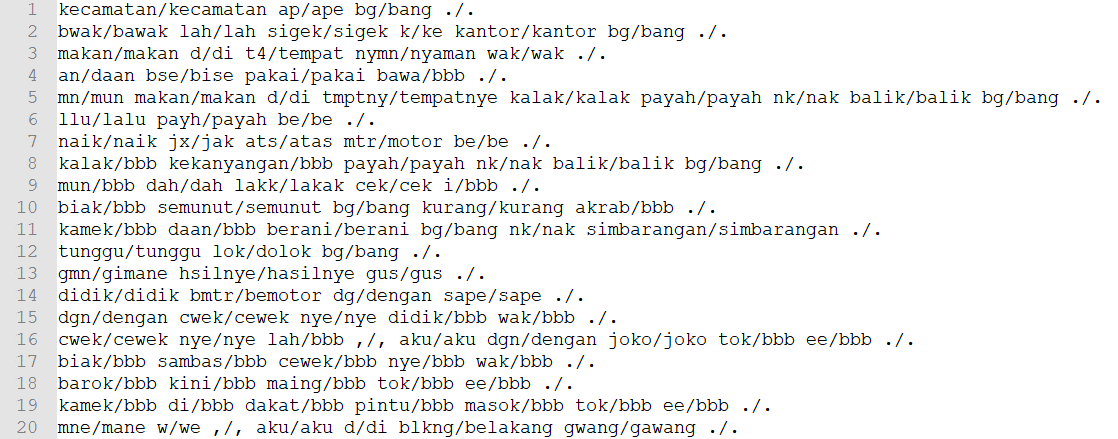
Setelah didapat *file excel* maka akan dilakukan perbandingan kata antara korpus teks *chatting* dan korpus bahasa Melayu Sambas dan penandaan/*tagging*. *Tagging* dilakukan dengan membandingkan setiap padanan kata dari korpus teks *chatting* terhadap kata dari korpus bahasa Melayu Sambas.

Dalam pra proses HMM pada penelitian ini dilakukan dengan 2 skema atau skenario pembuatan korpus *tagging* atau korpus *training*. Skenario pertama yaitu untuk setiap kata yang sama dari korpus teks *chatting* maka padanan kata akan diubah dengan huruf unik “bbb” seperti yang diperlihatkan pada bagian pembuatan korpus di atas. Hasil penandaan atau korpus *tagging* untuk pra proses dengan skenario pertama seperti yang diperlihatkan pada gambar 3.11 berikut.



Gambar 3.11 Isi Korpus *Tagging* Pra Proses Skenario Pertama

Pada pembuatan korpus *tagging* untuk pra proses bahasa teks *chatting* ke bahasa baku skenario kedua hampir sama dengan korpus *tagging* skenario pertama namun pada kata yang padanan katanya tidak sama maka akan ditandai dengan padanan kata yang benar dan kemudian pada kata sebelum dan kata sesudahnya akan ditandai dengan padanan kata yang benar. Untuk lebih jelas, hasil korpus *tagging* skenario kedua diperlihatkan pada gambar 3.12 berikut.

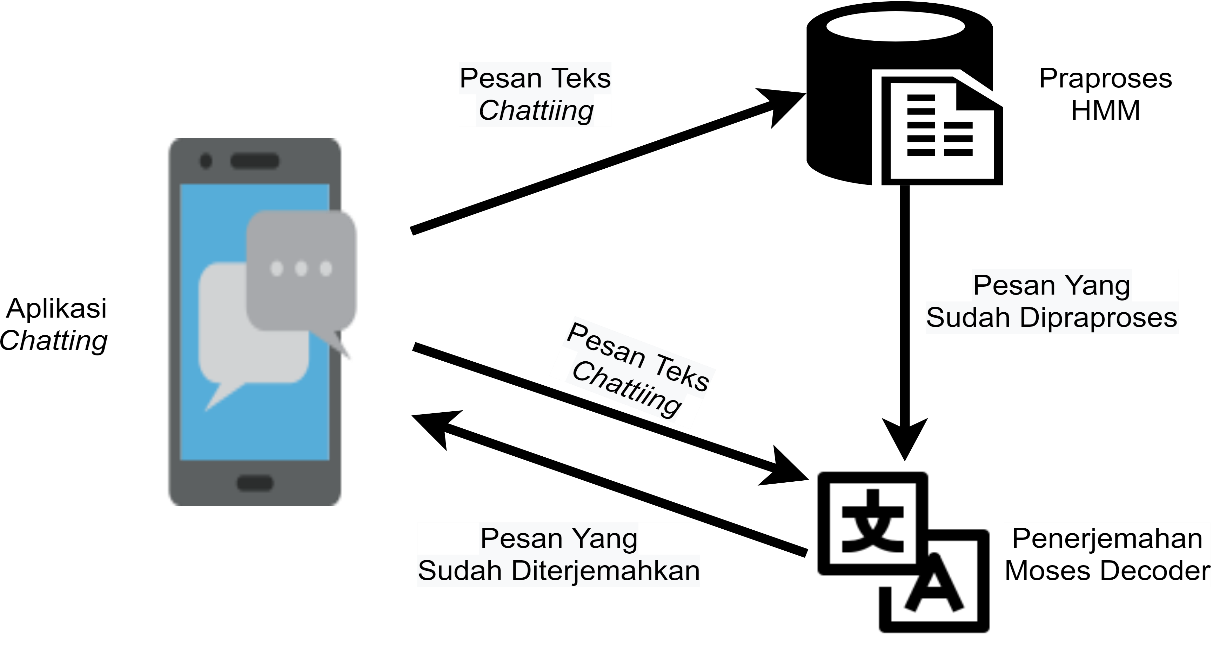


Gambar 3.12 Isi Korpus *Tagging* Pra Proses Skenario Kedua

Setelah didapat korpus *tagging* (dengan format *file* \*.crp) maka proses selanjutnya adalah melakukan *trianing* pada dua skenario korpus *tagging* dengan memasukan perintah *training* pada dua folder yang berbeda. Hasil setiap *training* korpus kemudian akan menghasilkan dua buah *file* dengan nama Lexicon.trn dan Ngram.trn. Setelah dilakukan *training* maka setiap mesin HMM siap digunakan untuk pra proses bahasa teks *chatting* menjadi bahasa Melayu Sambas sebelum dilakukan penerjemahan.

#### Proses Penerjemahan Pesan

Proses penerjemahan pesan dimulai dengan pra proses pesan teks *chatting* oleh mesin HMM dan kemudian hasil pra proses pesan berbahasa Melayu Sambas tersebut akan diterjemahkan ke bahasa Indonesia menggunakan mesin penerjemah Moses Decoder. Alur penerjemahan pesan dapat dilihat pada gambar 3.13 berikut.

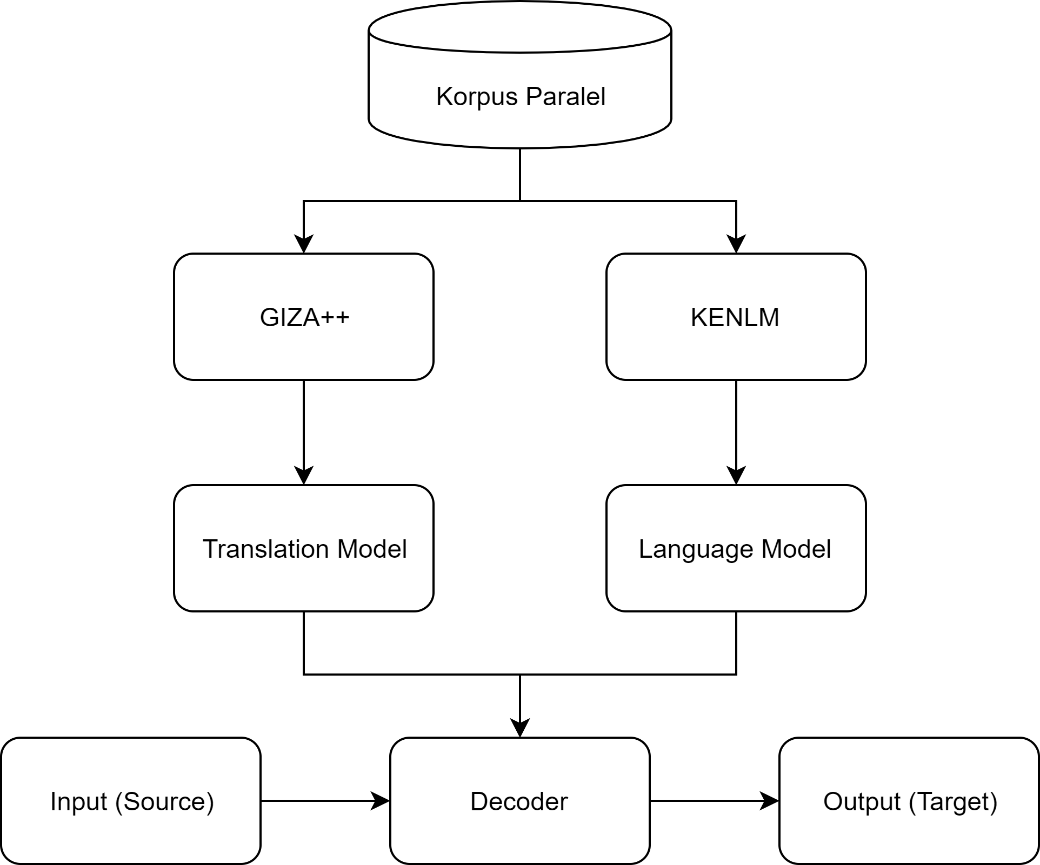


Gambar 3.13 Proses Penerjemahan Pesan

Proses penerjemahan Moses Decoder akan menerjemahkan pesan bahasa Melayu Sambas hasil pra proses HMM ke bahasa Indonesia, namun pesan juga dapat diterjemahkan langsung dengan menghiraukan pra proses pesan. Proses penerjemahan dengan Moses Decoder memerlukan korpus paralel, yaitu kumpulan dari kalimat-kalimat dalam dua bahasa yang berbeda, yang merupakan kalimat-kalimat selaras, setiap kalimat dalam satu bahasa cocok dengan kalimat yang diterjemahkan dalam bahasa lain. Korpus paralel yang digunakan pada penelitian ini adalah korpus bahasa Melayu Sambas dan korpus bahasa Indonesia.

Proses penerjemahan bahasa pada Moses Decoder dimulai dengan pra proses/*preprocessing* terhadap korpus paralel yaitu merapikan korpus dengan *script* bantuan yang berfungsi untuk *cleaning,* tokenisasi, dan *case folding*. *Cleaning* berfungsi untuk menghapus titik diakhir koma dan juga menghapus yang berlebihan. Sedangkan *case folding* berfungsi untuk mengganti huruf kapital menjadi huruf kecil disetiap kalimat.

Setelah dilakukan *preprocessing* terhadap korpus paralel maka dilakukan pemodelan translasi dan pemodel bahasa pada korpus paralel. Proses pemodelan translasi dilakukan oleh GIZA++ dan akan menghasilkan *translation model*, sedangkan pemodelan bahasa dilakukan dengan *language model toolkit* KENLM. Untuk lebih jelasnya arsitektur mesin penerjemah statistik Moses Decoder diperlihatkan pada gambar 3.14 berikut.

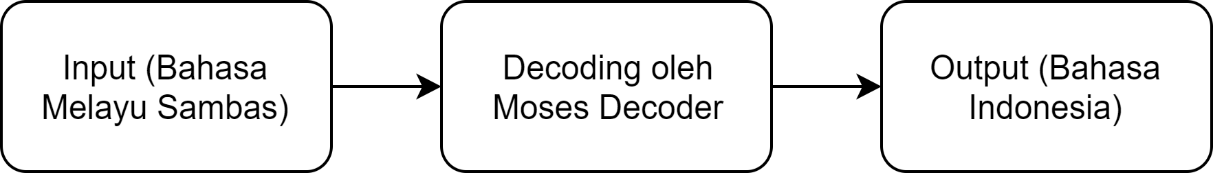


Gambar 3.14 Arsitektur Mesin Penerjemah Statistik

Pada proses pemodelan translasi dengan menggunakan GIZA++ akan menghasilkan *vocabulary corpus*, *word aligment*, dan *lexical table*. Dokumen *vocabulary corpus* berisikikan 3 atribut utama yaitu *uniq* id, atribut kata (token) dan nilai frekuensi kemunculan kata dalam dokumen korpus. Korpus yang digunakan adalah korpus paralel dengan demikian dokumen *vocabulary corpus* yang dihasilkan ialah *vocabulary corpus* bahasa Melayu Sambas dan bahasa Indonesia. Dokumen *word aligment* terdiri dari 3 baris kalimat yaitu baris pertama berisi letak kalimat target dalam korpus, panjang kalimat sumber, panjang kalimat target dan nilai *aligment*. Baris kedua berisikan kalimat bahasa sumber dan baris ketiga merupakan *aligment* kalimat bahasa target terhadap kalimat sumber. Tabel model translasi terdiri dari tabel kata yang berisi kosakata dari bahasa sumber yang memiliki makna pada bahasa sasaran ataupun sebaliknya (leksikal), setiap kosakata yang dihasilkan memiliki nilai probabilitas.

Langkah selanjutnya dilakukan pemodelan bahasa atau *language model training* dengan KENLM yang kemudian menghasilkan tabel model bahasa dengan *n-gram* data. Model Bahasa *n-gram* memiliki nilai probabilitas dalam bahasa target. Dalam proses pemodelan bahasa diantaranya ada empat proses yang harus dilalui yaitu: *Text (Input)*, *N-gram Modelling*, *N-gram Probabilities*, *Discont Methods & Frequency Smoothing* hingga menghasilkan *output language model* berupa ARPA File.

Setelah *languange model* dan *translation model* didapatkan dari proses *training* maka akan digunakan dalam proses penerjemahan oleh moses dengan menggunakan *input* bahasa Melayu Sambas dengan terjemahannya dalam bahasa Indonesia. Proses penerjemahan moses dapat diperlihatkan pada Gambar 3.15.



Gambar 3.15 Proses Penerjemahan oleh Moses Decoder

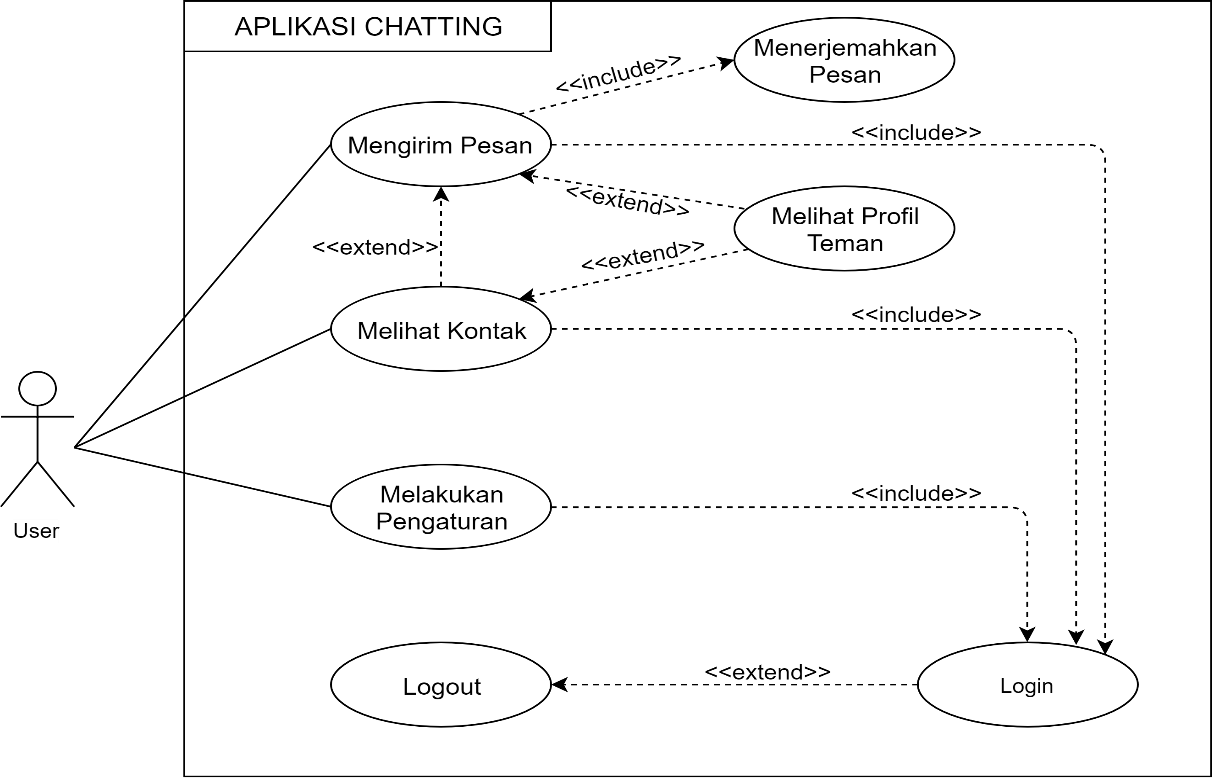
Moses akan menerjemahkan kalimat *input* berupa kalimat sumber dari korpus paralel. Kemudian kalimat masukan akan diproses oleh decoder moses dan akan menghasilkan kalimat *output* berupa kalimat hasil terjemahan ke dalam Bahasa target. Proses *decoding* dilakukan dengan mencari terjemahan frasa bahasa target untuk setiap frasa dari bahasa sumber pada tabel translasi frasa, kemudian dilanjutkan dengan mencari probabilitas maksimum untuk frasa bahasa target pada tabel model translasi hingga dihasilkan kata atau kalimat terjemahan dari masukan yang diproses oleh *decoder*.

### Unified Modelling Language (UML)

Dalam perancangan aplikasi *chat* penerjemah, jenis diagram UML yang digunakan berupa *Use Case diagram, Statechart diagram, Sequence diagram*, dan *Class diagram.*

#### Use Case Diagram

*Use Case diagram* merupakan pemodelan untuk perilaku aktor di dalam sistem aplikasi yang dibuat. *Use case diagram* menggambarkan interaksi antara aktor terhadap sebuah sistem. *Use case diagram* dari aplikasi *chat* penerjemah yang akan dibuat dapat dilihat pada gambar 3.16 berikut.



Gambar 3.16 *Use Case* Diagram Aplikasi *Chatting*

Dari gambar 3.16 dapat dilihat bahwa aktor melakukan interaksi dengan beberapa *use case*. Definisi dari aktor pada gambar diatas diperlihatkan pada tabel 3.5 berikut.

Tabel 3.5 Definisi Aktor

|  |  |  |
| --- | --- | --- |
| **No.** | **Aktor** | **Deskripsi** |
| 1 | User | *User* merupakan aktor atau seorang pengguna dari aplikasi *chat* penerjemah. *User* dapat melakukan *login* dan *logout*, mengirim pesan ke *user* lain dan sebaliknya (menerima pesan dari *user* lain), menerjemahkan pesan yang dikirim, melihat profil teman (*user* lain), serta melakukan pengaturan bahasa dan profil. |

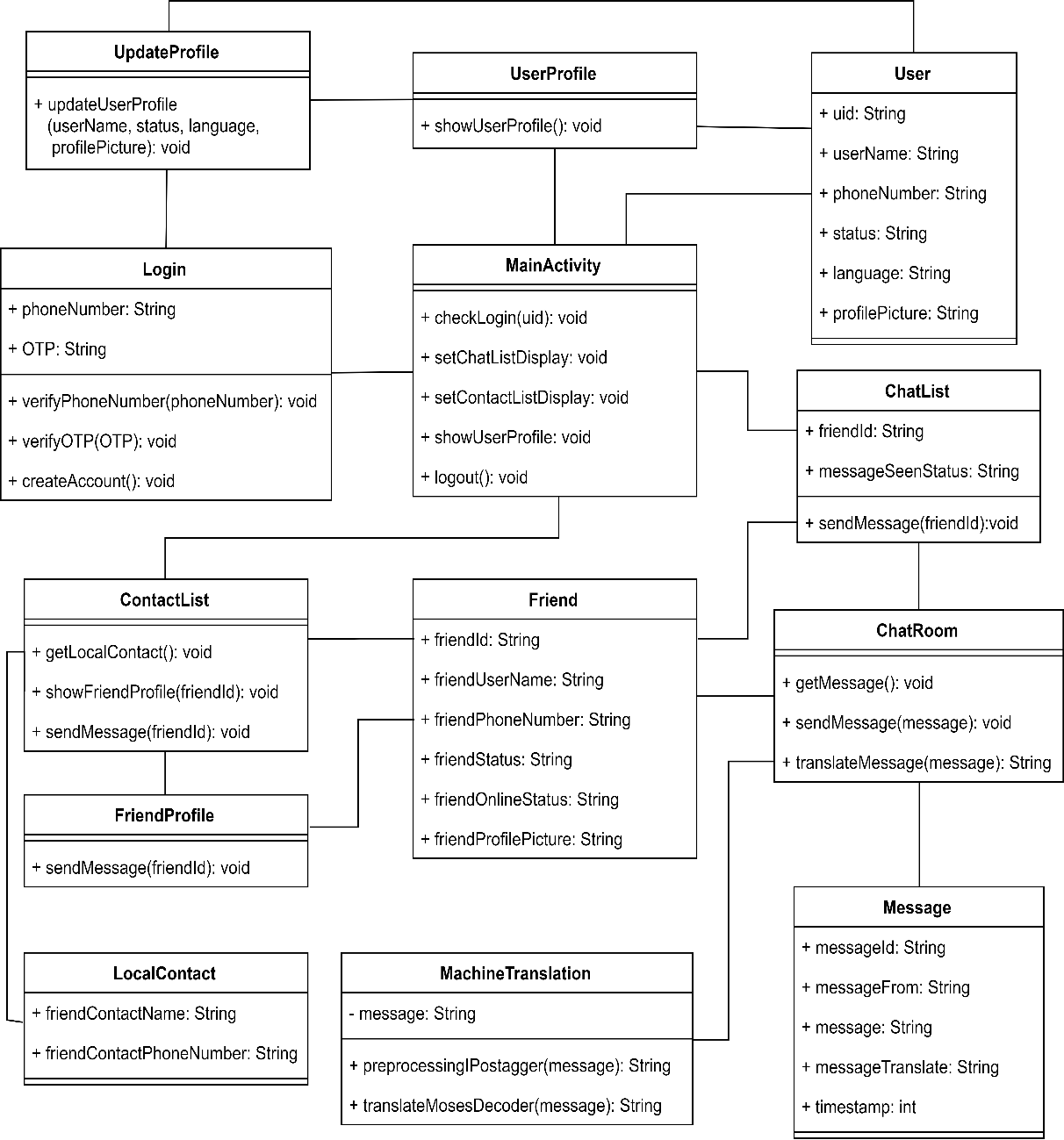
Definisi dari *use case* diagram dijelaskan berdasarkan setiap fungsinya dan dapat dilihat pada Tabel 3.6 berikut.

Tabel 3.6 Definisi *Use Case*

| **No.** | **Nama *Use Case*** | **Deskripsi** |
| --- | --- | --- |
| 1 | Mengirim Pesan | *User* dapat mengirim pesan kepada *user* lain melalui ruang obrolan. Pesan yang dikirim terlebih dahulu dilakukan penerjemahan, yaitu menerjemahkan pesan berbahasa Melayu Sambas menjadi pesan berbahasa Indonesia. |
| 2 | Melihat Kontak | *User* dapat melihat daftar *user*/pengguna lain aplikasi *chat* penerjemah, yang tersimpan pada lokal kontak perangkat/*smartphone user*. Ketika mengakses fitur kontak *user* dapat melihat profil *user* lain (teman) dan mengirim pesan ke *user* lain (teman). |
| 3 | Melakukan Pengaturan | *User* dapat melakukan pengaturan, yaitu melakukan perubahan terhadap data profil *user* dan mengubah bahasa utama yang akan digunakan *user* dalam mengirim pesan. |
| 4 | Menerjemah-kan Pesan | Menerjemahkan pesan merupakan proses yang dilakukan ketika pengiriman pesan, pesan berbahasa Melayu Sambas yang dikirim terlebih dahulu diterjemahkan menjadi pesan berbahasa Indonesia. Hasil terjemahan pesan kemudian dikirim ke penerima pesan yang dituju. |
| 5 | Melihat Profil Teman | Merupakan fitur untuk melihat profil dari teman atau *user* lain yang menjadi lawan *chat* atau yang terdaftar pada kontak. |
| 6 | *Login* | *Login* merupakan proses untuk memverifikasi *user* yang akan masuk ke aplikasi. Aplikasi akan mengingat atau menyimpan data *user* yang berhasil *login* sehingga *user* tidak perlu lagi untuk melakukan *login*. |
| 7 | *Logout* | *Logout* adalah proses menghapus data *login user* yang tersimpan. Untuk dapat masuk ke halaman utama aplikasi maka *user* diharuskan untuk melakukan *login* kembali. |

#### Class Diagram

*Class diagram* menggambarkan struktur sistem dari pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. Struktur sistem harus dapat melakukan fungsi-fungsi yang sesuai dengan kebutuhan sistem dan saling berhubungan satu dengan yang lain, *class diagram* dari aplikasi *chatting* yang akan dibangun dapat dilihat pada gambar 3.17 berikut.



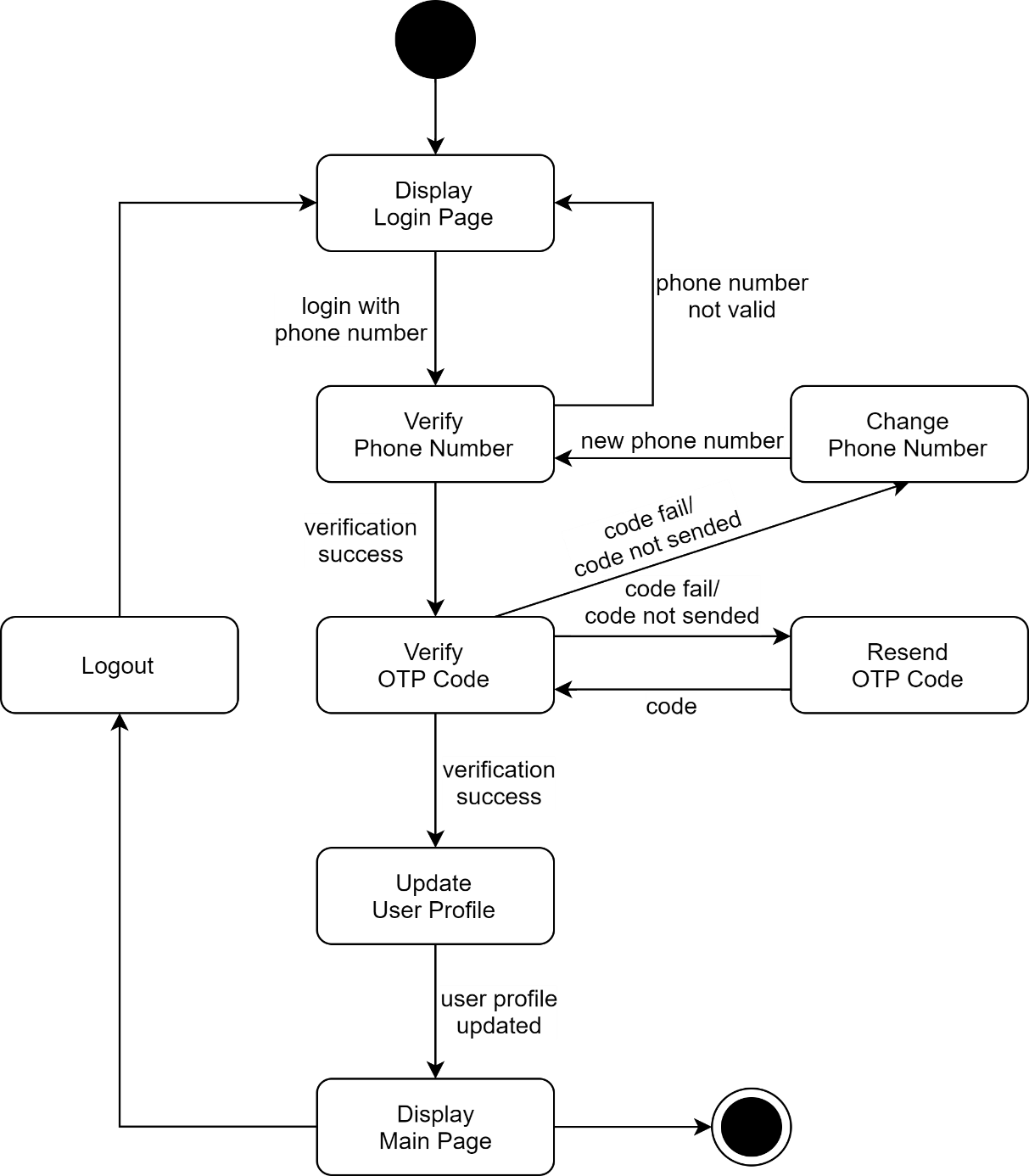
Gambar 3.17 *Class Diagram* Aplikasi *Chatting*

#### Statechart Diagram

*Statechart diagram* atau *state diagram* merupakan diagram uml yang menggambarkan perilaku (*bahavior*) dinamis sistem, transisi atau perubahan keadaan sistem dalam merespon suatu aksi yang diterima. Suatu *state*/keadaan dalam sebuah *class* dapat berubah jika terjadi suatu *event*/aksi yang memengaruhinya.

1. *State Diagram User*

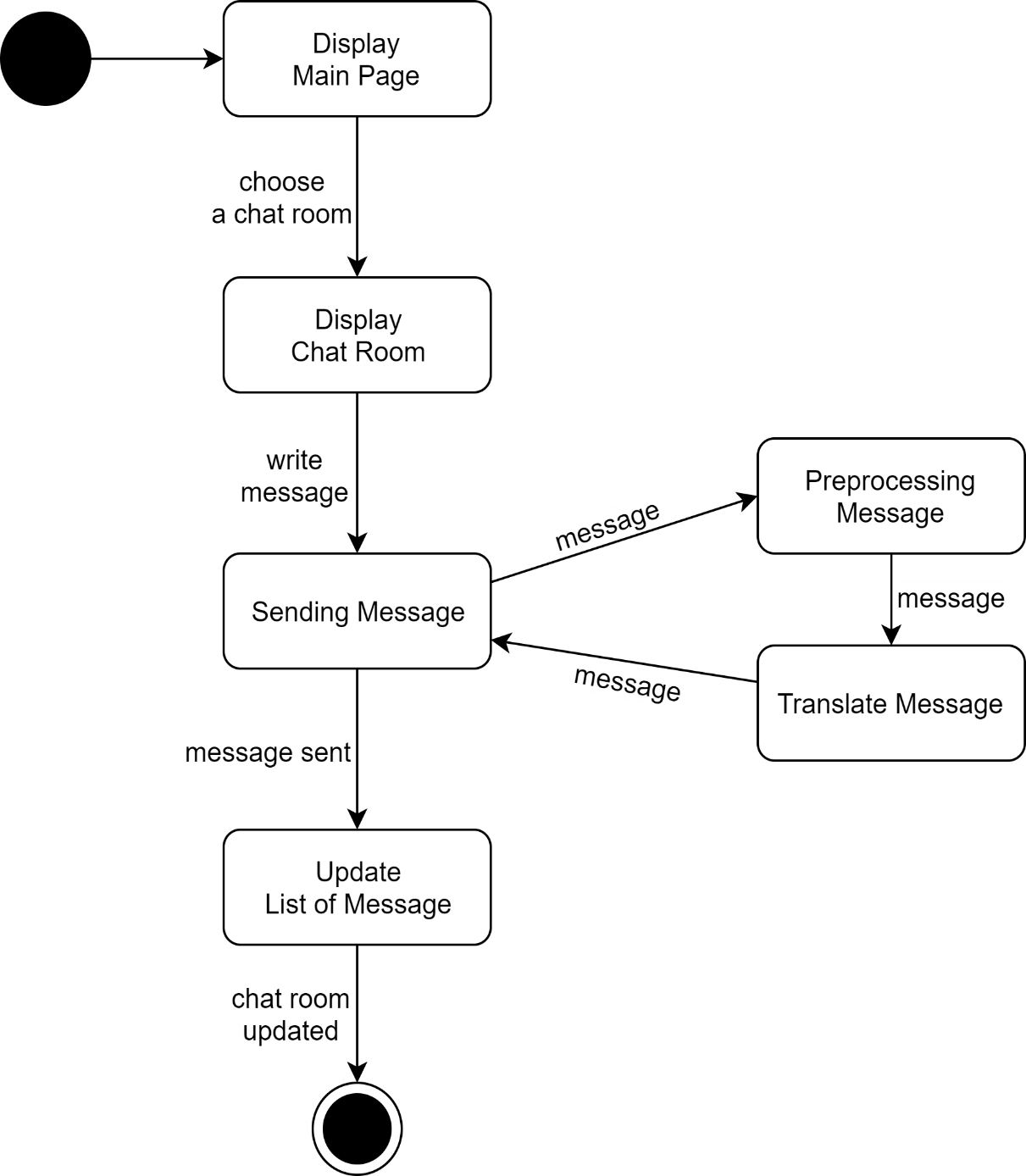
*State diagram user* pada gambar 3.18 menggambarkan perubahan perilaku pada sistem ketika *user* melakukan proses masuk atau *login* ke aplikasi.



Gambar 3.18 *State Diagram User*

1. *State Diagram Chat*

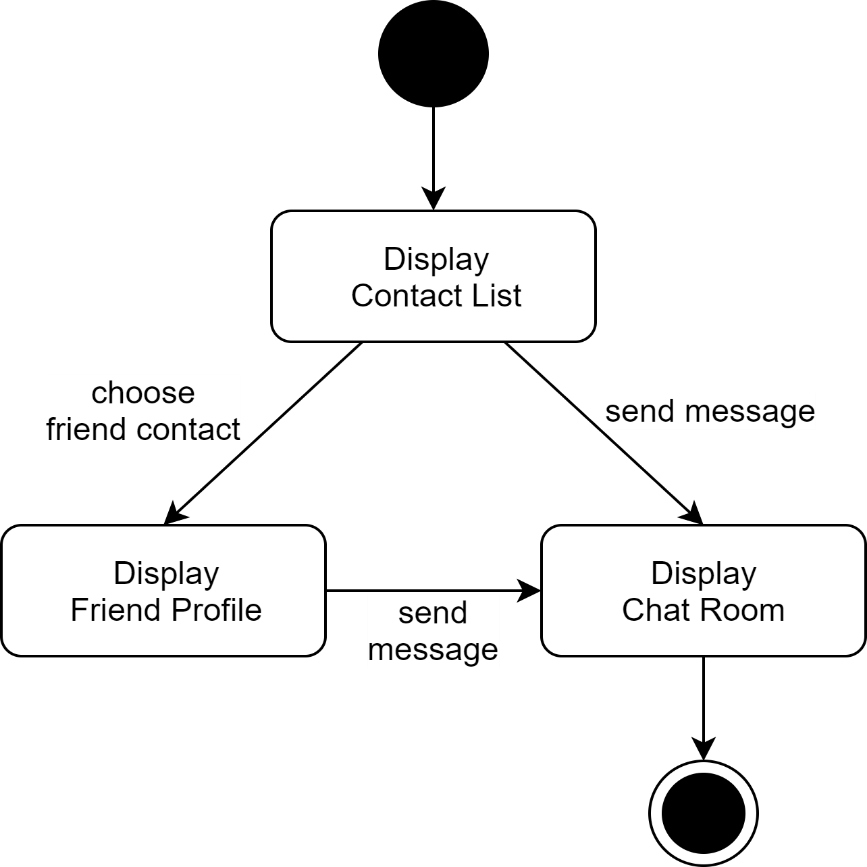
*State diagram chat* pada gambar 3.19 menggambarkan perubahan perilaku pada sistem ketika *user* mengirimkan pesan kepada *user* lain.



Gambar 3.19 *State Diagram Chat*

1. *State Diagram Contact*

*State diagram contact* pada gambar 3.20 menggambarkan perilaku sistem ketika *user* mengakses fitur kontak.



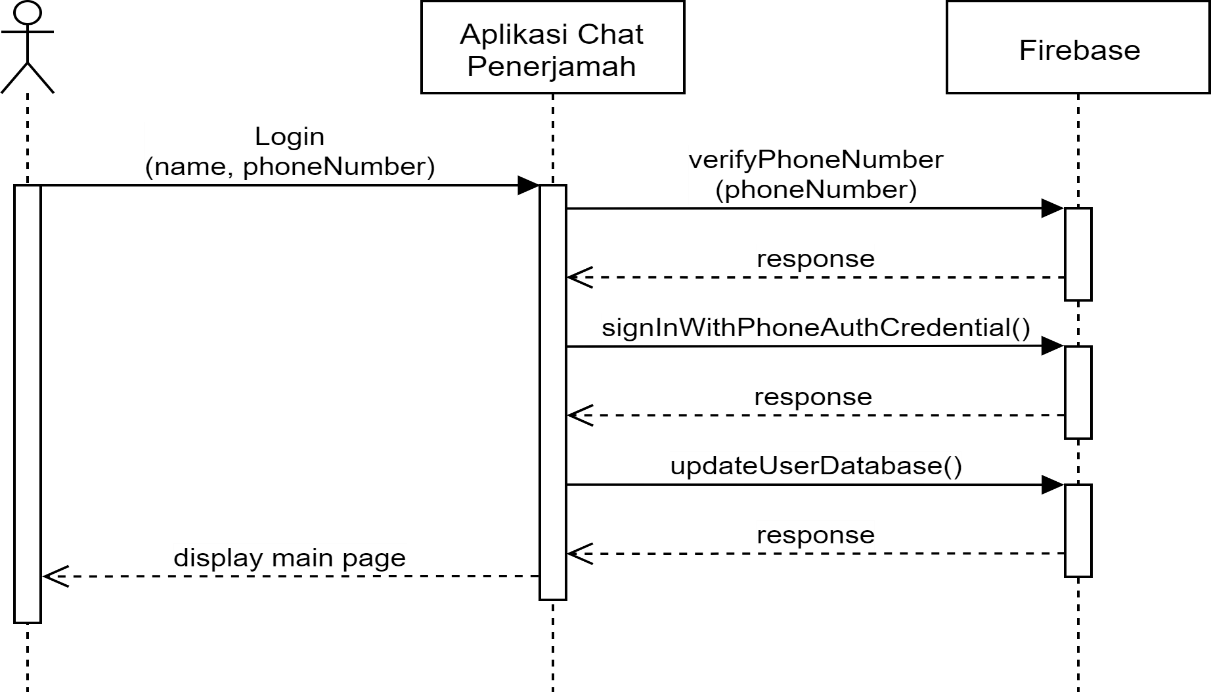
Gambar 3.20 *State Diagram Contact*

#### Sequence Diagram

*Sequence Diagram* merupakan salah satu diagram interaksi yang menjelaskan bagaimana suatu operasi itu dilakukan, pesan apa yang dikirim dan kapan pelaksanaannya. Diagram ini diatur berdasarkan waktu di mana objek-objek yang berkaitan dengan proses berjalannya operasi diurutkan dari kiri ke kanan berdasarkan waktu terjadinya dalam pesanan yang terurut.

1. *Sequence Diagram Login* Aplikasi

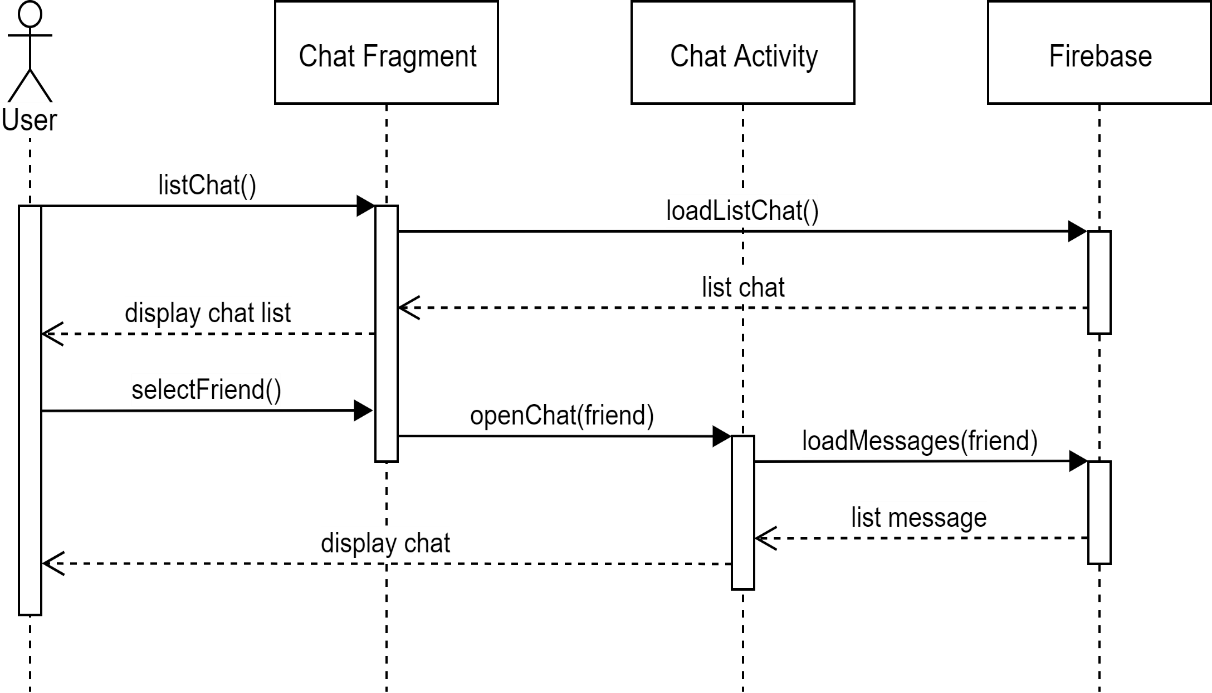
*Sequence diagram login* aplikasi pada gambar 3.21 menggambarkan proses sistem ketika melakukan autentikasi data *user* untuk dapat masuk ke halaman utama aplikasi.



Gambar 3.21 *Sequence Diagram Login*

1. *Sequence Diagram* Memuat *Chat*

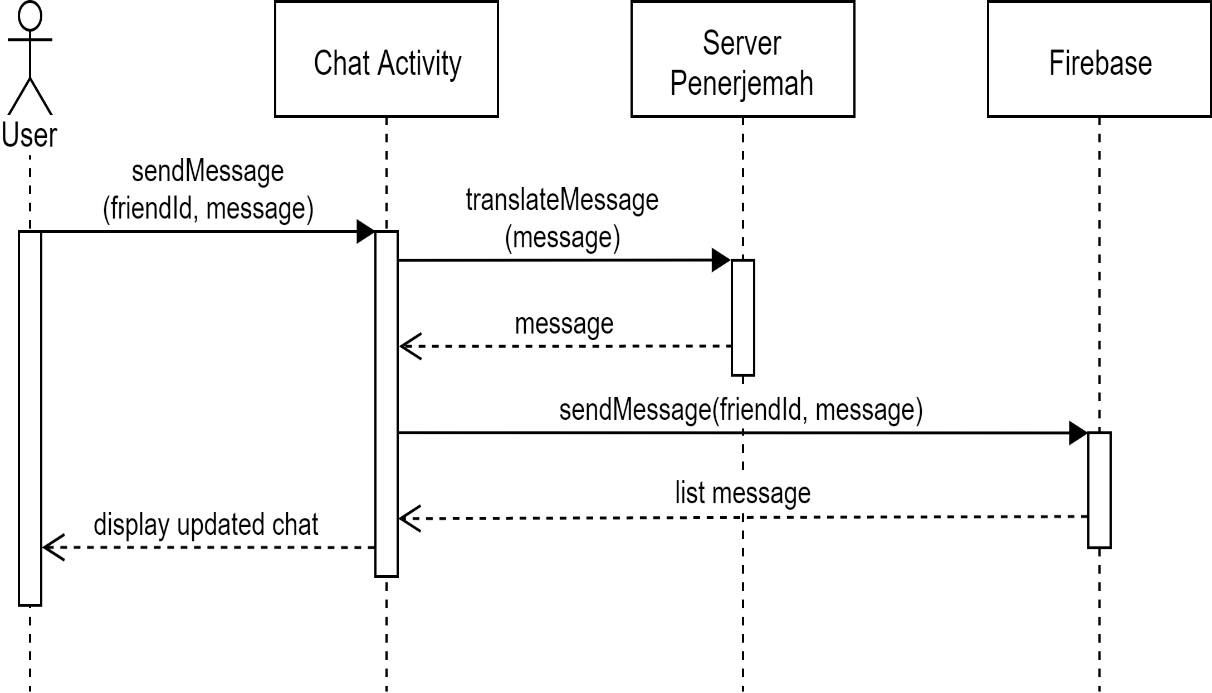
*Sequence diagram chat* pada gambar 3.22 menggambarkan proses sistem ketika memuat daftar *chat* dan kemudian memuat daftar pesan pada ruang obrolan.



Gambar 3.22 *Sequence Diagram* Memuat *Chat*

1. *Sequence Diagram* Mengirim Pesan

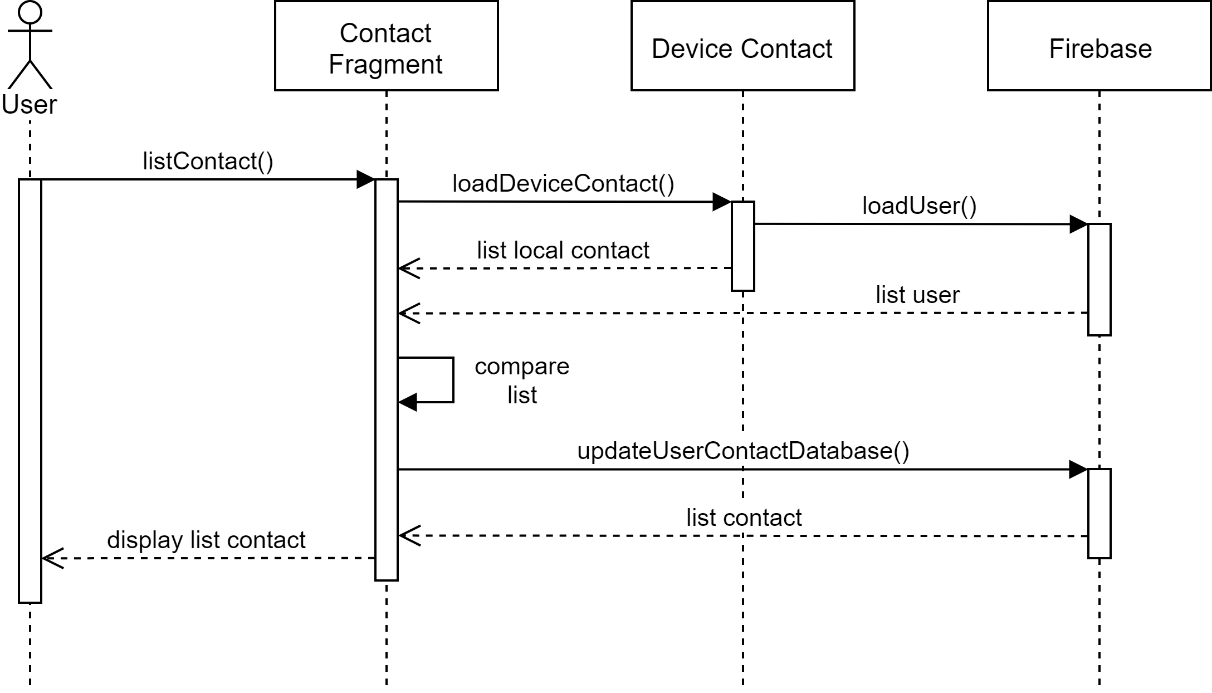
*Sequence diagram* mengirim pesan pada gambar 3.23 menggambarkan proses sistem ketika menangani pengiriman pesan oleh *user*.



Gambar 3.23 *Sequence Diagram* Mengirim *Chat*

1. *Sequence Diagram* Kontak

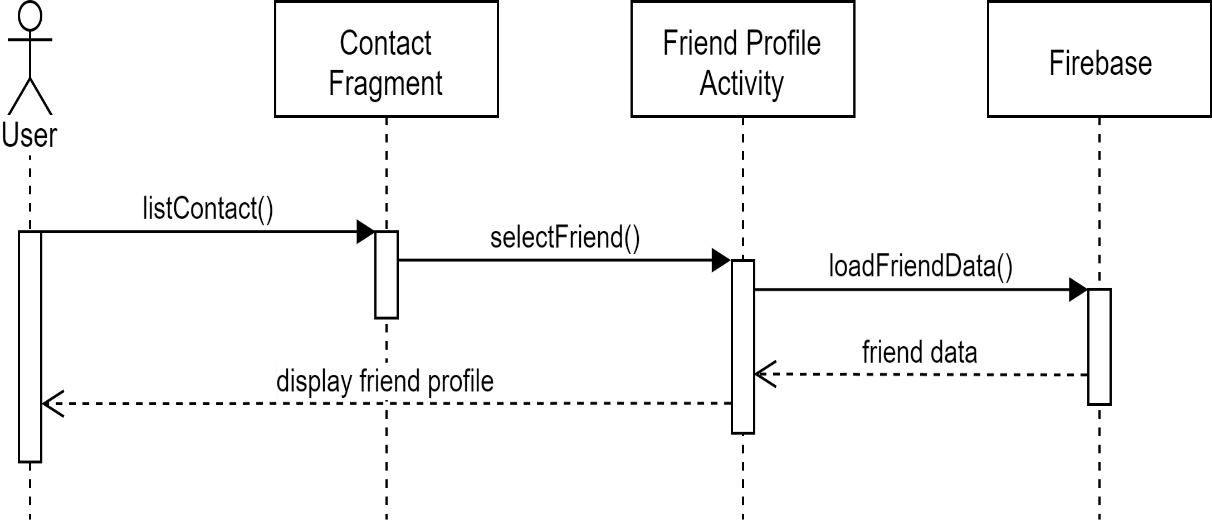
*Sequence diagram* kontak pada gambar 3.24 menggambarkan proses sistem ketika memuat dan menampilkan daftar kontak pada.



Gambar 3.24 *Sequence Diagram* Memuat Kontak

1. *Sequence Diagram* Profil Teman

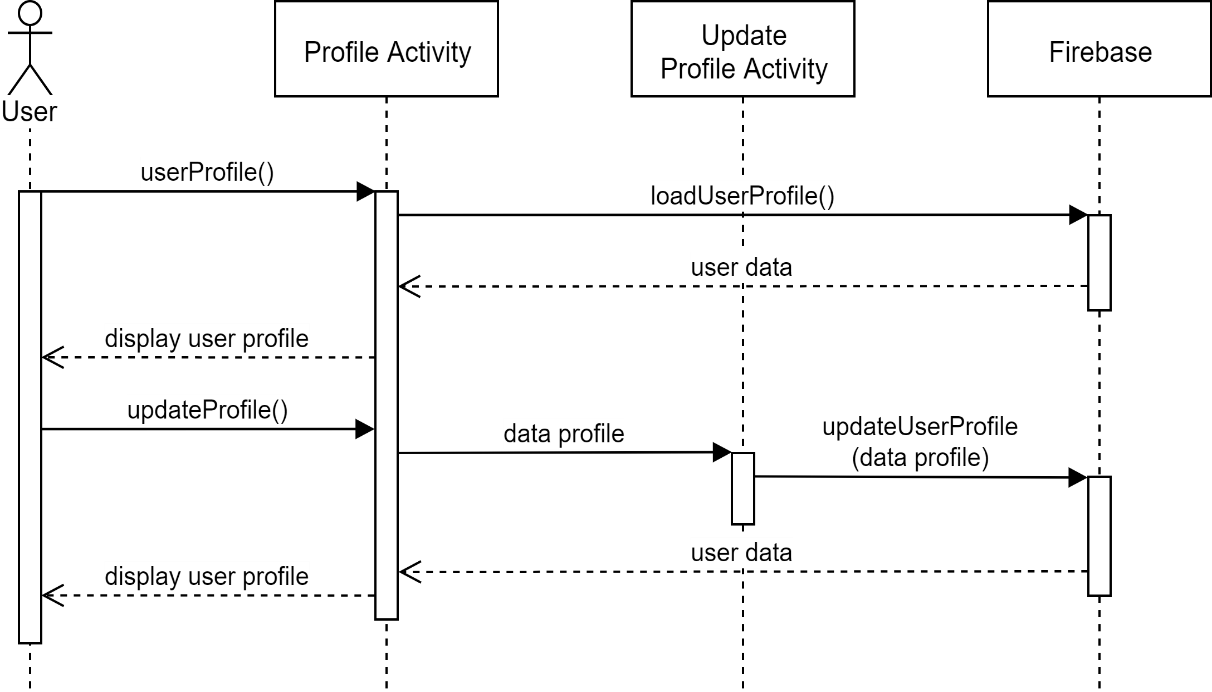
*Sequence diagram* profil teman pada gambar 3.25 menggambarkan proses sistem ketika memuat atau menampilkan profil teman.



Gambar 3.25 *Sequence Diagram* Profil Teman

1. *Sequence Diagram* Pengaturan

*Sequence diagram* pengaturan pada gambar 3.26 menggambarkan proses sistem ketika *user* memperbarui data profil maupun bahasa yang digunakan *user*.



Gambar 3.26 *Sequence Diagram* Profil

### Perancangan *Layout* Antarmuka Aplikasi

Perancangan *layout* dibuat untuk memberikan gambaran tata letak atau penataan elemen-elemen pada suatu aplikasi yang akan dibuat. Rancangan *layout* yang akan dibuat diantaranya adalah sebagai berikut:

1. Perancangan Halaman Verifikasi Nomor Telepon

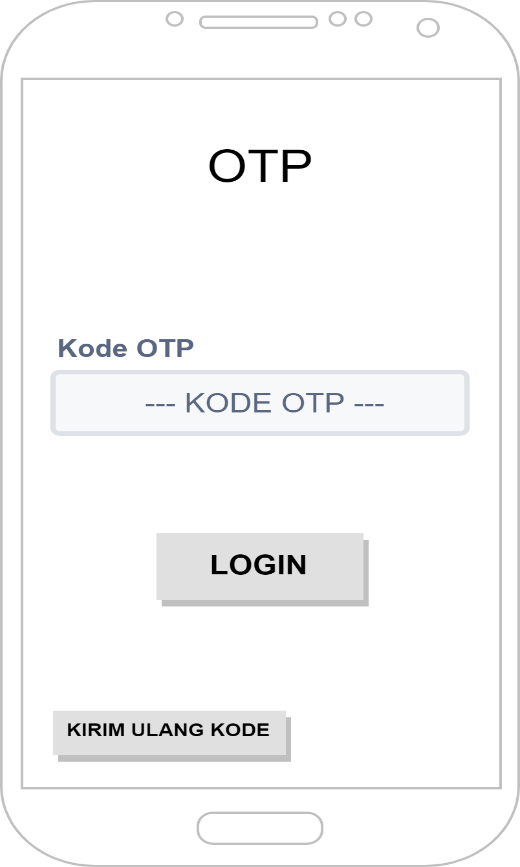
Halaman verifikasi nomor telepon merupakan halaman pertama yang ditampilkan ketika *user* pertama kali membuka aplikasi setelah aplikasi berhasil terpasang pada perangkat *user*. Pada halaman ini *user* memasukkan nomor telepon aktif yang digunakan oleh *user* untuk kemudian diverifikasi oleh Firebase. Rancangan halaman verifikasi nomor telepon diperlihatkan pada gambar 3.27 berikut.



Gambar 3.27 Perancangan Halaman Verifikasi Nomor Telepon

1. Perancangan Halaman Verifikasi OTP

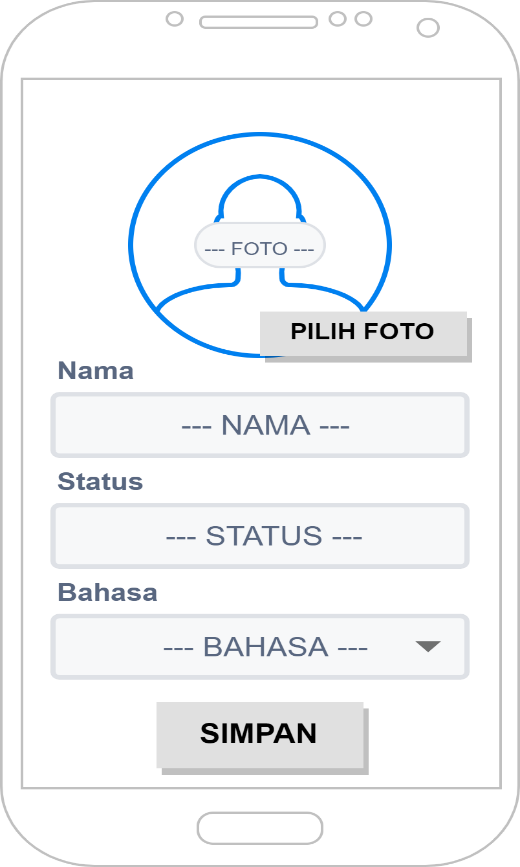
Halaman verifikasi OTP merupakan halaman untuk memverifikasi kode OTP yang dikirimkan ke nomor telepon *user* setelah *user* berhasil melakukan verifikasi nomor telepon. Halaman verifikasi OTP juga dapat disebut halaman *login* karena setelah berhasil melakukan verifikasi kode OTP maka aplikasi akan menyimpan sesi *user* sehingga untuk selanjutnya *user* tidak perlu melakukan verifikasi nomor telepon dan kode OTP lagi. Rancangan halaman verifikasi OTP diperlihatkan pada gambar 3.28 berikut.



Gambar 3.28 Perancangan Halaman Verifikasi OTP

1. Perancangan Halaman *Update* Profil

Halaman *update* profil merupakan halaman yang ditampilkan setelah *user* berhasil melakukan verifikasi kode OTP. Pada halaman ini *user* menetapkan ataupun memperbarui foto profil, status, dan bahasa yang digunakan *user*. Rancangan halaman *update* profil diperlihatkan pada gambar 3.29 berikut.



Gambar 3.29 Perancangan Halaman *Update* Profil

1. Perancangan Halaman Daftar Obrolan

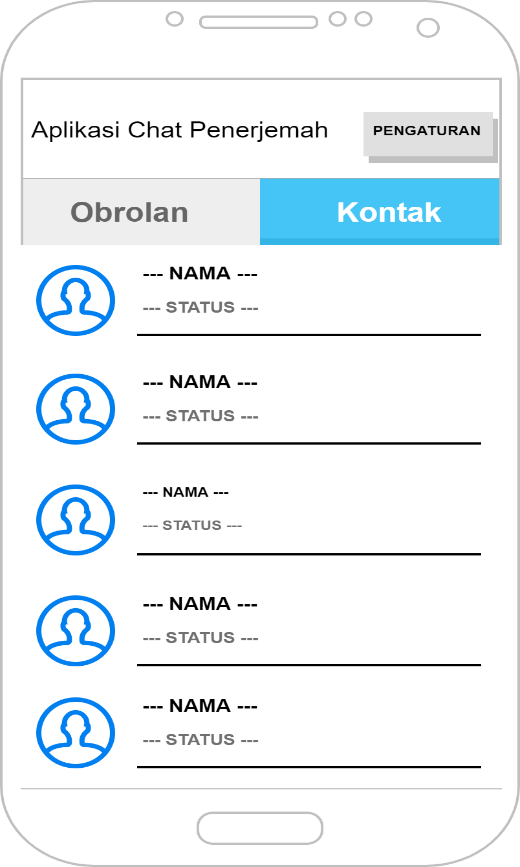
Halaman daftar obrolan merupakan halaman utama yang akan ditampilkan setiap kali *user* membuka aplikasi (setelah *user* melakukan *login* aplikasi). Halaman daftar obrolan berisi obrolan-obrolan yang dilakukan *user* dengan *user* lain dalam aplikasi. Rancangan halaman daftar obrolan diperlihatkan pada gambara 3.30 berikut.



Gambar 3.30 Perancangan Halaman Daftar Obrolan

1. Perancangan Halaman Daftar Kontak

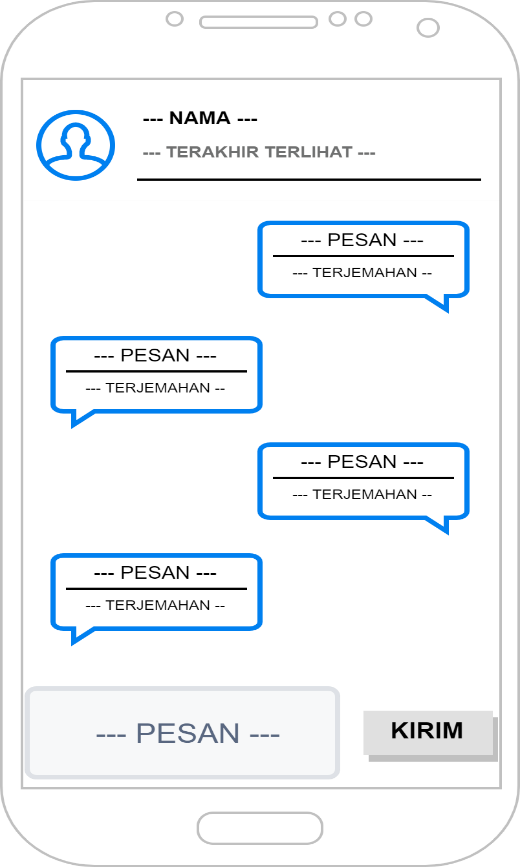
Halaman daftar kontak merupakan halaman yang menampilkan daftar teman yang menggunakan aplikasi *chat* penerjemah. Daftar kontak yang ditampilkan merupakan daftar teman pada kontak perangkat *user* yang juga menggunakan atau terdaftar pada aplikasi *chat* penerjemah. Rancangan halaman daftar kontak diperlihatkan pada gambar 3.31 berikut



Gambar 3.31 Perancangan Halaman Daftar Kontak

1. Perancangan Halaman Ruang Obrolan

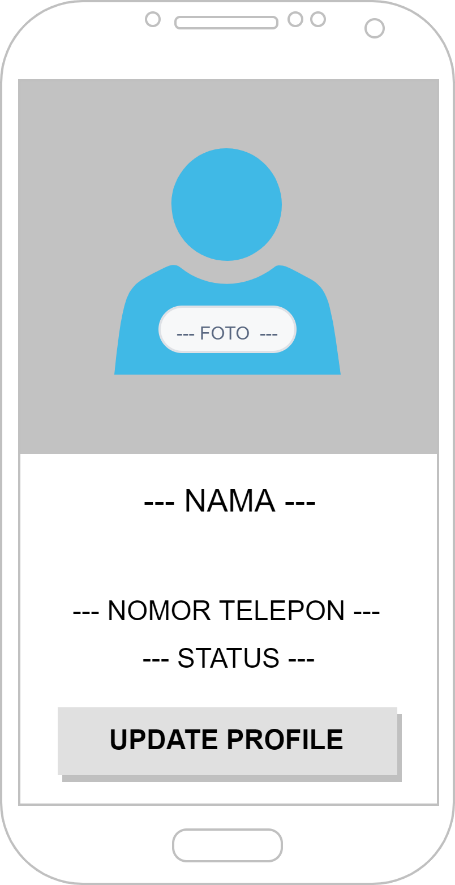
Halaman ruang obrolan merupakan halaman di mana *user* mengobrol atau berkirim pesan dengan *user* lain. Rancangan halaman ruang obrolan diperlihatkan pada gambar 3.32 berikut.



Gambar 3.32 Perancangan Halaman Ruang Obrolan

1. Perancangan Halaman Profil Saya

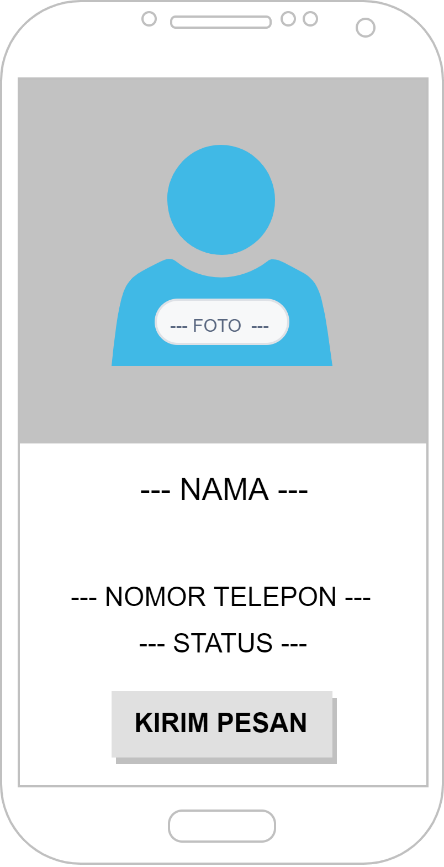
Halaman profil saya merupakan halaman yang menampilkan data profil *user*. Halaman profil saya menampilkan foto profil, nama *user*, nomor telepon dan status *user*. *User* dapat memperbarui data melalui tombol *update profile* yang nantinya akan diarahkan ke halaman *update* profil. Rancangan halaman profil saya diperlihatkan pada gambar 3.33 berikut.



Gambar 3.33 Perancangan Halaman Profil Saya

1. Perancangan Halaman Profil Teman

Halaman profil teman merupakan halaman yang menampilkan data profil teman (*user* lain). Halaman profil teman menampilkan nama teman, foto profil, nomor telepon dan status teman (*user* lain). Rancangan halaman profil teman diperlihatkan pada gambar 3.34 berikut.



Gambar 3.34 Perancangan Halaman Profil Teman

## Implementasi

Setelah melakukan analisis dan perancangan sistem maka selanjutnya peneliti masuk ke tahap implementasi di mana dilakukan proses pembuatan server penerjemah dan proses pemrograman (*coding*). Pada pembuatan server penerjemah peneliti melakukan instalasi dan konfigurasi mesin HMM dan Moses Decoder. Untuk proses *coding* maka peneliti menggunakan bahasa pemrograman PHP untuk membuat *web service* server penerjemah, bahasa pemrograman Java untuk membuat aplikasi andorid, dan layanan Firebase untuk melakukan otentikasi *user*, penyimpanan data dan notifikasi.

### Pembuatan Server Penerjemah

#### Instalasi dan Konfigurasi Mesin HMM

Mesin HMM yang dibangun pada penelitian ini menggunakan *framework* IPostagger yang dibuat oleh Alfan Farizki Wicaksono dangan rangkaian *N-gram* yang digunakan adalah *trigram*. Mesin HMM digunakan untuk melakukan pra proses pesan yang dikirim sebelum pesan tersebut diterjemahkan. Langkah-langkah dalam instalasi dan konfigurasi mesin HMM adalah sebagai berikut.

1. Mengunduh IPostagger yang dibuat oleh Alfan Farizki Wicaksono (dapat diunduh dari halaman <https://ir.cs.ui.ac.id/alfan/old/>).(Wicaksono and Ayu, 2010)
2. Membuat korpus *tagging*, yaitu dengan menandai teks *chatting* (pesan) dengan bahasa Melayu Sambas.
3. Melakukan *training* korpus dengan memasukkan perintah seperti pada gambar 3.35 berikut.



Gambar 3.35 Perintah *Training* Mesin HMM

1. Membuat sebuah program Java sehingga mesin HMM dapat melakukan *tagging* melalui terminal. Isi program Java untuk melakukan *tagging* pra proses HMM melalui terminal diperlihatkan pada kode program 3.2 berikut.

Kode Program 3.1. *Source Code* Program Java *Tagging*

1. import NLP\_ITB.POSTagger.HMM.Decoder.MainTagger;
2. import java.util.ArrayList;
3. public class IPostaggerCli {
4. public static void main(String args[]) {
5. int lm = 1;
6. int affix = 1;
7. int pass2 = 0;
8. int lex = 1;
9. MainTagger mainTagger = new MainTagger(
10. "./resource/Lexicon.trn",
11. "./resource/N-gram.trn",
12. lm,
13. 3,
14. 3,
15. 0,
16. affix,
17. false,
18. 0.2,
19. pass2,
20. 500.0,
21. Lex
22. );
23. ArrayList<String> result = mainTagger.taggingStr(args[0]);
24. int i = 0;
25. for (i = 0; i < result.size(); i++){
26. String[] paths = result.get(i).split("/");
27. if (paths[1].equals("bbb")){
28. result.set( i, paths[0]);
29. } else {
30. result.set( i, paths[1]);
31. }
32. System.out.print(result.get(i)+" ");
33. if (result.get(i).equals("./.")){
34. System.out.print("\n");
35. }
36. }
37. }
38. }
39. Menjalankan kode program Java melakukan *tagging* melalui terminal dengan memasukkan perintah seperti pada gambar 3.36 berikut.



Gambar 3.36 Perintah Menjalankan *Tagging* HMM

#### Instalasi dan Konfigurasi Moses Decoder

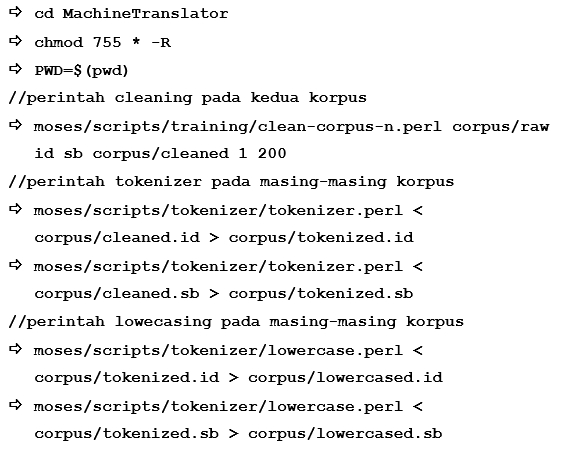
Moses Decoder merupakan implementasi mesin penerjemah dengan pendekatan statistik (*statistical machine translation*) yang menggunakan konsep probabilitas dalam melakukan penerjemahan bahasa berdasarkan hasil analisis dari korpus paralel. Moses Decoder digunakan untuk menerjemahkan pesan bahasa Melayu Sambas (hasil *tagging* mesin HMM) menjadi pesan bahasa Indonesia. Gambar 3.37 berikut menunjukkan *syntax* atau perintah-perintah untuk memasang Moses Decoder pada server.



Gambar 3.37 Perintah Instalasi Moses Decoder

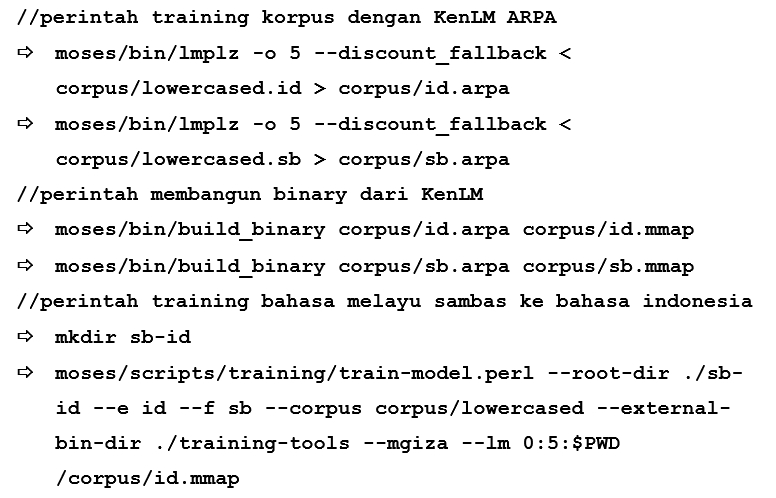
Setelah Moses Decoder berhasil terpasang pada server maka diperlukan beberapa persiapan sebelum melakukan *training* korpus paralel. Beberapa persiapan sebelum melakukan *training* korpus paralel adalah sebagai berikut.

1. Membuat sebuah folder pada server dengan nama “MachineTranslator”.
2. Membuat sebuah folder dengan nama “corpus” di dalam folder “MachineTranslator”.
3. Membuat *file* dengan nama “raw.sb” yang berisi korpus bahasa Melayu Sambas di dalam folder “corpus”.
4. Membuat *file* dengan nama “raw.id” yang berisi korpus bahasa Indonesia di dalam folder “corpus”.
5. Melakukan *preprocessing* korpus paralel, yaitu *cleaning*, *tokenizing* dan *lowercasing* pada korpus. Gambar 3.38 berikut menunjukkan perintah-perintah untuk *preprocessing* korpus.



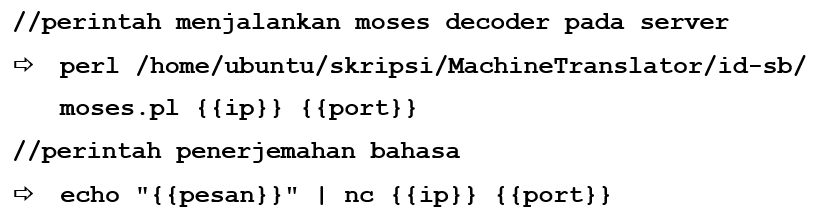
Gambar 3.38 Perintah *Preprocessing* Korpus Paralel

Setelah dilakukan beberapa persiapan untuk *training* korpus paralel maka selanjutnya adalah proses *training* korpus paralel. Gambar 3.39 berikut menunjukkan perintah-perintah untuk melakukan *training* korpus paralel sehingga Moses Decoder dapat menerjemahkan pesan bahasa Melayu Sambas menjadi bahasa Indonesia.



Gambar 3.39 Perintah *Training* Korpus Paralel Bahasa Melayu Sambas ke Bahasa Indonesia

Setelah melakukan proses *training* korpus paralel bahasa Melayu Sambas dan bahasa Indonesia hingga Moses Decoder dapat menerjemahkan masukan bahasa Melayu Sambas menjadi Bahasa Indonesia maka selanjutnya adalah menjalankan mesin penerjemah statistik (Moses Decoder) pada server. Gambar 3.40 berikut menunjukkan perintah-perintah untuk menjalankan atau *running* Moses Decoder pada server dan perintah penerjemahan setelah Moses Decoder berhasil *running* pada server.



Gambar 3.40 Perintah Menjalankan Moses Decoder pada Server

## Pengujian Sistem

Dalam penelitian ini dilakukan tiga pengujian yaitu untuk menguji nilai akurasi hasil pra proses HMM, nilai akurasi hasil penerjemahan, dan keberhasilan pembuatan aplikasi. Pengujian nilai akurasi hasil pra proses HMM dilakukan dengan mencari nilai *accuracy, precision, recall*, dan *f-measure* dari hasil *tagging* mesin HMM dengan *trigram* yang dibangun. Pengujian nilai akurasi hasil penerjemahan dilakukan dengan pengujian hasil penerjemahan mesin translasi oleh BLEU dengan metode *k-fold cross validation* dan pengujian hasil penerjemahan mesin translasi oleh ahli bahasa. Sedangkan untuk menguji keberhasilan pembuatan aplikasi dilakukan dengan pengujian *black box*.

### Pengujian *Accuracy*, *Precision*, *Recall* dan *F-Measure* Hasil Pra Proses HMM dengan Metode *K-Fold Cross Validation*

Pengujian *accuracy, precision, recall*, dan *f-measure* dilakukan untuk mengetahui nilai akurasi keberhasilan dalam melakukan *tagging* perbaikan atau pra proses bahasa Melayu Sambas *chatting* menjadi bahasa Melayu Sambas baku. Pengujian dilakukan dengan membandingkan hasil *tagging* otomatis teks *chatting* oleh mesin HMMdengan hasil *tagging* manual teks *chatting*.

Pada penelitian ini menggunakan korpus teks *chatting* sebanyak 4.500 kalimat. Pada 4.500 kalimat korpus teks *chatting* dilakukan *tagging* manual sehingga diperoleh 4.500 baris korpus *tagging* manual. Pengujian *accuracy, precision, recall*, dan *f-measure* dilakukan menggunakan metode *k-fold cross validation* dengan skema 10-*fold cross validation*.

Pada 4.500 kalimat korpus teks *chatting* akan dibagi ke dalam 10-*fold* sehingga setiap satu *fold* berisi 450 kalimat teks *chatting* (korpus uji). Pada 4.500 baris korpus *tagging* dilakukan hal yang sama sehingga pada setiap satu *fold* berisi 450 baris korpus *tagging* (korpus *tagging* manual), dan sisanya sebanyak 4.050 baris korpus *tagging* akan digunakan sebagai data pelatihan (korpus *training*).

Pengujian *accuracy, precision, recall*, dan *f-measure* dilakukan pada 10 *fold* secara bergantian, kemudian dihitung nilai rata-rata 10 *fold* sebagai nilai akhir. Pengujian dilakukan dengan NLTK *tools*.

### Pengujian Hasil Penerjemahan Mesin Translasi oleh BLEU dengan Metode *K-Fold Cross Validation*

Evaluasi otomatis adalah metode evaluasi yang dilakukan secara otomatis dalam mengukur tingkar keakuratan hasil terjemahan dari mesin penerjemah. Salah satu metode evaluasi otomatis adalah BLEU (*Bilingual Evaluation Understudy*), yaitu metode yang mengevaluasi kualitas dari hasil terjemahan yang telah diterjemahkan oleh mesin dari bahasa alami ke bahasa lain.

Pada penelitian ini akan dilakukan pengujian atau evaluasi otomatis menggunakan BLEU dengan metode *k-fold cross validation*. Adapun untuk penelitian ini akan digunakan skema 10*-fold cross validation*, sehingga dilakukan 10 kali percobaan pengujian akurasi pada *Language Model Toolkit* yang digunakan (KenLM). Jadi dalam pemilihan data untuk korpus uji dari jumlah total korpus kalimat sebanyak 4.500 dibagi menjadi 10 kali percobaan didapatkan 450 kalimat untuk setiap percobaannya. Misalnya pada percobaan 1 korpus ujinya diambil dari kalimat 1 sampai 450 sedangkan untuk korpus yang di*training* diambil dari kalimat 451 sampai 4.500, percobaan 2 korpus ujinya diambil dari kalimat 450 sampai 900 sedangkan korpus yang di*training* diambil dari kalimat 1 sampai 449 dan 901 sampai 4.500 begitu seterusnya sampai percobaan ke 10.

### Pengujian Hasil Penerjemahan Mesin Translasi oleh Ahli Bahasa

Pengujian ahli bahasa dilakukan apabila telah dilakukan pengujian hasil evaluasi mesin penerjemah statistik dengan BLEU menggunakan metode *k-fold cross validation*. Pengujian ahli bahasa digunakan untuk mengetahui apakah hasil terjemahan dari mesin penerjemah statistik mendekati atau bahkan sudah sesuai dengan terjemahan dari ahli bahasa. Proses evaluasi ahli bahasa diperlihatkan pada gambar 3.41 berikut.



Gambar 3.41 Proses Evaluasi Secara Manual oleh Ahli Bahasa

Nilai skor atau persentase akurasi hasil penerjemahan dari pengujian hasil penerjemahan mesin translasi oleh ahli bahasa didapat melalui persamaan 3.1 berikut.

(3.1)

Keterangan:

P = Persentase akurasi

C = Jumlah kata yang diterjemahkan dengan tepat menurut penilaian dari ahli bahasa

R = Jumlah kata hasil terjemahan

### Pengujian *Black Box*

Pengujian *black* *box* merupakan pengujian yang dapat dilakukan dengan mengamati hasil eksekusi beberapa data pengujian dan memeriksa fungsi-fungsi yang terdapat pada perangkat lunak. Pengujian *black box* dipilih karena tidak memerlukan akses ke kode program atau aplikasi dan dapat dilakukan ketika pengembangan telah selesai. Pada penelitian ini dilakukan pengujian *black box* dengan teknik pengujian *robustness testing*. Pada teknik ini data dipilih dari luar range yang didefinisikan dan bertujuan untuk membuktikan tidak terjadinya kesalahan jika masukan yang tidak valid.

Hasil pengujian *black box* akan dirangkum dalam sebuah tabel seperti pada tabel 3.7 berikut.

Tabel 3.7 Hasil Pengujian Black Box

|  |  |  |  |
| --- | --- | --- | --- |
| Fungsi | Kasus Uji | Hasil Eksekusi | Keterangan |
|  |  |  |  |

Tabel pengujian terdiri dari 4 kolom, yaitu :

Fungsi : menunjukkan fungsi yang akan dilakukan pengujian.

Kasus uji : merupakan skenario pengujian yang akan dijalankan terhadap suatu fungsi yang akan diuji.

Hasil eksekusi : merupakan hasil dari pengujian yang telah dilakukan berdasarkan skenario pengujian pada kolom kasus uji. Isi kolom ini berupa status berhasil atau tidaknya fungsi yang dijalankan berdasarkan skenario pengujian.

Keterangan : merupakan penjelasan dari hasil eksekusi pengujian, berupa pesan kesalahan atau respon balik dari aplikasi.

# HASIL IMPLEMENTASI DAN PENGUJIAN

## Hasil Implementasi Aplikasi

Berdasarkan perancangan yang dilakukan maka berhasil dilakukan implementasi dalam bentuk aplikasi berbasis android yaitu “Aplikasi Chat Penerjemah”. Aplikasi ini dibangun menggunakan bahasa pemrograman java dan menggunakan IDE (Integrated Development Enviroment) Android Studio. Server yang digunakan pada penelitian ini adalah Alibaba ECS (Elastic Cloud Server) yang didalamnya menggunakan sistem operasi Ubuntu Server 16.04 LTS. Sedangkan untuk *service* penerjemah ditulis dengan bahasa pemrograman PHP.

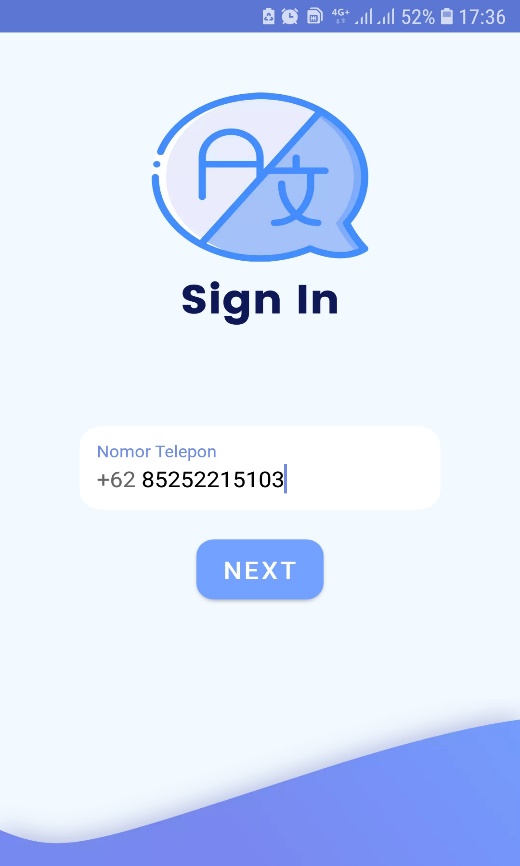
Hasil rancangan *class diagram* diimplementasikan dalam aplikasi dalam bentuk file-file yang menyusun aplikasi. Tabel 4.1 berikut menunjukkan hasil rancangan dari *class diagram* yang dibuat.

Tabel 4.1 Hasil Implementasi Rancangan *Class Diagram*

| **Nama Kelas** | **Definisi Kelas** | **Hasil Implementasi File** |
| --- | --- | --- |
| MainActivity | Kelas ini merupakan *controller* dari tampilan utama aplikasi. Kelas ini menerapkan adapter *fragment* yang merupakan hasil dari kelas ChatList dan ContactList. | * MainActivity.java |
| User | Kelas User merupakan model yang berisi atribut-atribut yang akan digunakan oleh *controller*. | * User.java |
| UserProfile | Kelas ini merupakan *controller* dari tampilan profil *user*. | * ProfileActivity.java |
| UpdateProfile | Merupakan *controller* dari tampilan untuk memperbarui data profil *user*. | * UpdateProfileActivity. java |
| Login | Merupakan *controller* dari proses *login* yaitu tampilan verifikasi nomor telepon dan verifikasi kode OTP. | * LoginActivity.java * OtpActivity.java |
| ChatList | Kelas ini merupakan *controller* yang meng*handle* tampilan daftar obrolan, merupakan bagian dari MainActivity. | * ChatsFragment.java |
| ChatRoom | Kelas ini merupakan *controller* dari tampilan di mana *user* mengobrol dengan *user* lain. | * ChatsActivity.java |
| Message | Kelas Message merupakan model yang berisi atribut-atribut yang akan digunakan pada *controller.* | * Messages.java |
| Machine Translation | Merupakan kelas yang berisi fungsi untuk melakukan penerjemahan. | * TranslateMessage.java |
| Friend | Kelas Friend merupakan model yang berisi atribut-atribut yang akan digunakan pada *controller.* | * Friend.java |
| LocalContact | Kelas LocalContact merupakan model yang berisi atribut-atribut yang akan digunakan pada *controller* ContactList*.* | * LocalContact.java |
| ContactList | Kelas ini merupakan *controller* yang meng*handle* tampilan daftar teman, merupakan bagian dari MainActivity. | * ContactsFragment.java |
| FriendProfile | Kelas ini merupakan *controller* yang menghandle tampilan profil teman. | * ProfileFriendActivity. java |

### Halaman Verifikasi Nomor Telepon

Halaman verifikasi nomor telepon merupakan rangkaian dari proses *login* pada aplikasi yang mana pada halaman ini *user* diharuskan untuk memasukkan nomor telepon aktif/valid yang digunakan oleh *user* pada perangkatnya. Tampilan dari halaman verifikasi nomor telepon diperlihatkan pada gambar 4.1 berikut.



Gambar 4.1 Halaman Verifikasi Nomor Telepon

### Halaman Verifikasi OTP

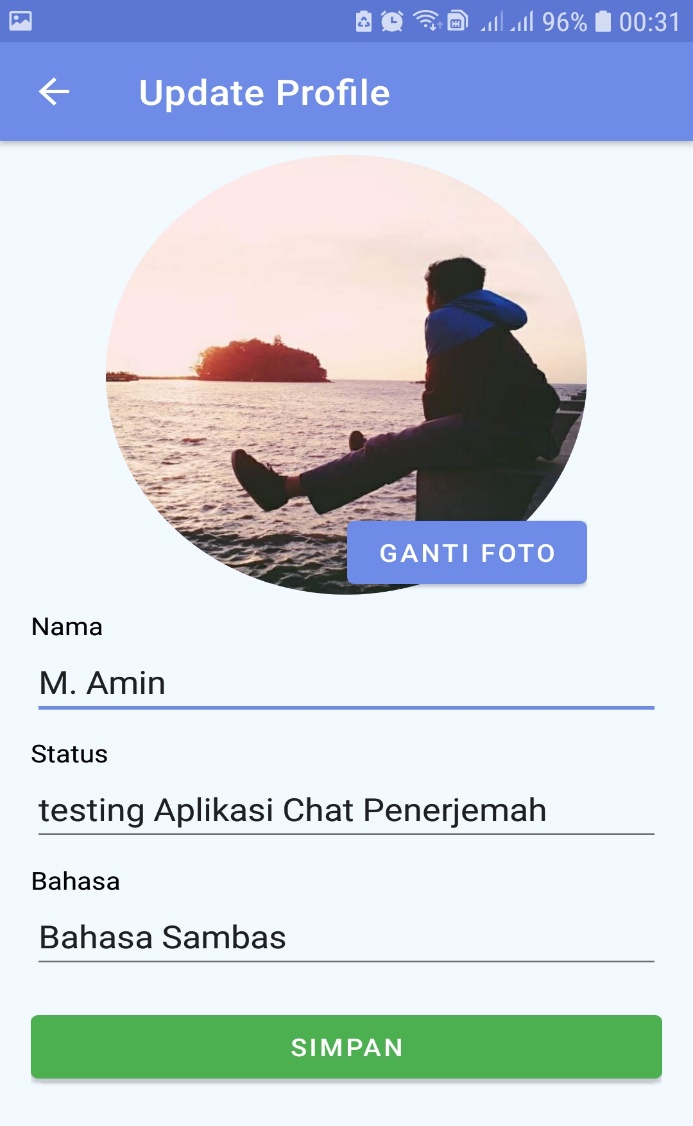
Halaman verifikasi otp merupakan langkah terakhir dalam melakukan *login* pada aplikasi *chat* penerjemah yang mana halaman ini hanya dapat diakses ketika verifikasi nomor telepon *user* berhasil dilakukan. Setelah verifikasi nomor telepon berhasil maka Firebase akan mengirimkan sms berupa kode OTP, kode OTP tersebut akan secara otomatis ditangkap oleh aplikasi dan kemudian *user* dapat lakukan verifikasi untuk menyelesaikan proses *login*. Tampilan halaman verifikasi OTP diperlihatkan pada gambar 4.2 berikut



Gambar 4.2 Halaman Verifikasi OTP

### Halaman *Update* Profil

Halaman *update* profil dapat diakses dengan dua cara yaitu setelah proses *login* berhasil dilakukan dan juga melalui halaman profil saya. Pada halaman *update* profil *user* dapat mengganti foto profil, nama, status, dan bahasa yang akan digunakan *user* dalam mengirim pesan. Jika *user* memilih bahasa Indonesia maka pesan tidak akan diterjemahkan, melainkan akan langsung dikirim ke penerima pesan. Sedangkan jika *user* memilih bahasa Sambas maka pesan yang dikirim *user* akan diterjemahkan ke bahasa Indonesia baru kemudian dikirim ke penerima pesan. Tampilan dari halaman *update* profil diperlihatkan pada gambar 4.3 berikut.



Gambar 4.3 Halaman *Update* Profil

### Halaman Daftar Obrolan

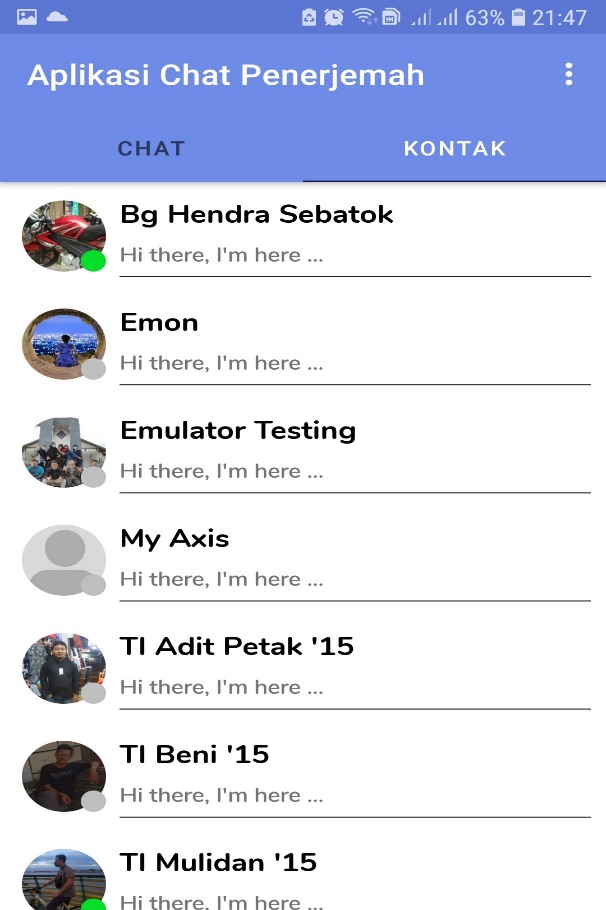
Halaman daftar obrolan merupakan halaman yang menampilkan daftar teman atau *user* lain yang melakukan obrolan dengan *user*. Informasi yang ditampilkan pada daftar obrolan ini antara lain nama teman, foto profil teman, status aktif (apakah *user* lain sedang menggunakan aplikasi atau tidak), pesan terakhir yang dikirim, dan tanggal pesan terakhir dikirim. Teman atau *user* lain yang sedang aktif dapat dilihat dari indikator berupa lingkaran kecil berwarna hijau muda pada pojok kanan bawah dari foto profil teman, sedangkan untuk teman yang tidak aktif maka indikator akan menunjukkan warna abu-abu. Untuk lebih jelasnya tampilan dari halaman daftar obrolan diperlihatkan pada gambar 4.4 berikut.



Gambar 4.4 Halaman Daftar Obrolan

### Halaman Daftar Kontak

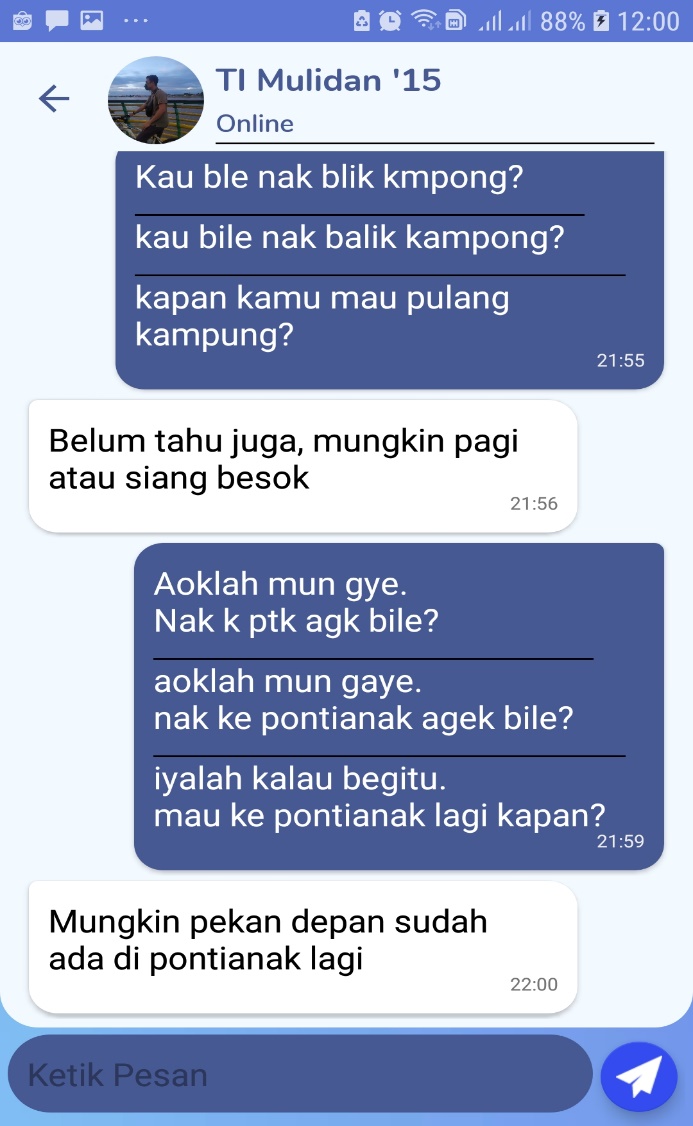
Halaman daftar kontak merupakan halaman yang menampilkan daftar teman atau *user* lain yang menggunakan aplikasi *chat* penerjemah. Daftar kontak yang ditampilkan merupakan daftar teman pada kontak perangkat *user* yang juga menggunakan atau terdaftar pada aplikasi *chat* penerjemah. Pada *item* dari daftar kontak maka *user* dapat memilih untuk membuka profil ataupun mengirim pesan kepada teman (*item*)yang dipilih. Tampilan halaman daftar kontak diperlihatkan pada gambar 4.5 berikut.



Gambar 4.5 Halaman Daftar Kontak

### Halaman Ruang Obrolan

Halaman ruang obrolan merupakan halaman yang menampilkan percakapan atau obrolan antara *user* dengan *user* lain. Pada halaman ini berisi nama *user* yang menjadi lawan komunikasi, foto profil, dan status aktif *user*. Pada *bubble chat* yang dikirim *user* yang menggunakan bahasa Melayu Sambas, isi *bubble chat* dibagi menjadi tiga bagian yaitu pesan asli yang berada pada bagian atas, pesan hasil pra proses HMM pada bagian tengah, dan pesan yang sudah diterjemahkan oleh mesin penerjemah yang berada pada bagian bawah. Sedangkan *bubble* *chat* yang dikirim *user* yang menggunakan bahasa Indonesia hanya berisi pesan asli yang ditulis oleh *user* tersebut. Untuk lebih jelasnya tampilan halaman ruang obrolan ditunjukkan pada gambar 4.6 berikut.



Gambar 4.6 Halaman Ruang Obrolan

### Halaman Profil Saya

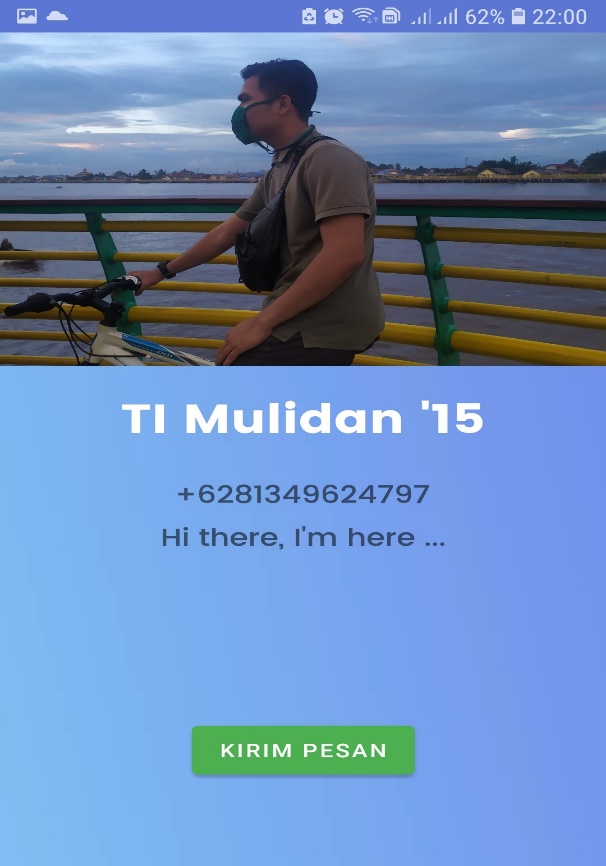
Halaman profil saya merupakan halaman yang menampilkan foto profil *user*, nama, status, dan bahasa yang digunakan. Pada halaman profil saya juga terdapat sebuah tombol yang berfungsi untuk mengarahkan *user* ke halaman *update* profil. Tampilan halaman profil saya ditunjukkan pada gambar 4.7 berikut.



Gambar 4.7 Halaman Profil Saya

### Halaman Profil Teman

Halaman profil teman merupakan halaman yang menampilkan data profil teman (*user* lain). Halaman profil teman menampilkan foto profil, nama, nomor telepon dan status teman (*user* lain). Pada halaman profil teman terdapat sebuah tombol yang berfungsi untuk mengarahkan *user* ke halaman ruang obrolan sehingga *user* dapat mengirim pesan ke teman yang dipilih. Tampilan halaman profil teman diperlihatkan pada gambar 4.8 berikut.



Gambar 4.8 Halaman Profil Teman

## Hasil Pengujian

Pada penelitian ini dilakukan tiga jenis pengujian berupa pengujian nilai akurasi hasil pra proses HMM, pengujian akurasi hasil terjemahan mesin translasi Moses Decoder dan pengujian fungsionalitas aplikasi. Pengujian nilai akurasi hasil pra proses HMM dilakukan dengan mencari nilai *accuracy, precision, recall*, dan *f-measure* dari hasil pra proses. Pengujian akurasi hasil terjemahan mesin translasi dilakukan dengan cara evaluasi otomatis menggunakan BLEU dengan metode *k-fold cross validation* dan evaluasi secara menual oleh ahli bahasa. Sedangkan pada pengujian pada aplikasi dilakukan dengan metode pengujian *black box* untuk memeriksa fungsional aplikasi dalam mengimplementasikan mesin penerjemah statistik (Moses Decoder).

### Pengujian *Accuracy*, *Precision*, *Recall* dan *F-measure* Pra Proses HMM

Pengujian *accuracy, precision, recall*, dan *f-measure* merupakan pengujian yang bersifat otomatis yang dilakukan dengan modul NLTK dan dilakukan dengan skema 10-*fold cross validation*. Kode program untuk menjalankan pengujian *accuracy, precision, recall*, dan *f-measure* diperlihatkan pada kode program 4.1 berikut.

Kode Program 4.1. Kode Pengujian *Accuracy*, *Precision*, *Recall*, dan *F-Measure*

1. import time
2. import nltk
3. from nltk.metrics import \*
4. startTime = time.time()#mengukur waktu proses --> mulai
5. file1 = open("zManualAsli")
6. corp\_manualasli = file1.read()
7. file1.close()
8. file2 = open("zOtomatisAsli")
9. corp\_otomatisasli = file2.read()
10. file2.close()
11. print("confusion\_matrix")
12. print(nltk.ConfusionMatrix(corp\_manualasli.split(), corp\_otomatisasli.split()))
13. # perintah mengitung nilai accuracy
14. a = accuracy(corp\_manualasli.split(), corp\_otomatisasli.split())
15. print(f"accuracy\_score = {a}")
16. reference\_set = set(corp\_manualasli.split())
17. test\_set = set(corp\_otomatisasli.split())
18. # perintah mengitung nilai precision
19. p = precision(reference\_set, test\_set)
20. print(f"precision\_score = {p}")
21. # perintah mengitung nilai recall
22. r = recall(reference\_set, test\_set)
23. print(f"recall\_score = {r}")
24. # perintah mengitung nilai f-measure
25. f = f\_measure(reference\_set, test\_set)
26. print(f"f1\_score = {f}")
27. # ukur waktu proses --> akhir
28. print(f"\nwaktu proses = {time.time()-startTime}")

Nilai tertinggi untuk setiap hasil pengujian adalah 1 (satu), semakin nilai yang didapat mendekati angka 1 maka hasil pengujian semakin baik. Nilai 1 pada *accuracy* berarti hasil *tagging* otomatis sangat akurat terhadap *tagging* manual. Nilai 1 pada *precision* berarti seluruh *tagging* otomatis yang ditemukan adalah relevan. Nilai 1 pada *recall* berarti seluruh korpus teks *chatting* yang diuji berhasil dilakukan *tagging* otomatis dengan benar. Sedangkan nilai *f-measure* yang merupakan gabungan dari *precision* dan *recall* menggambarkan tingkat kinerja sistem atau pra proses HMM yang dibangun.

#### Pengujian Pra Proses HMM Skenario Pertama

Hasil pengujian *accuracy, precision, recall*, dan *f-measure* dari pra proses HMMdengan korpus *tagging* skenario pertama diperlihatkan pada tabel 4.2 berikut.

Tabel 4.2 Hasil PengujianPra Proses HMM Skenario Pertama

| **Fold** | ***Accuracy*** | ***Precision*** | ***Recall*** | ***F-Measure*** |
| --- | --- | --- | --- | --- |
| 1 | 0,646 | 0,840 | 0,839 | 0,840 |
| 2 | 0,913 | 0,814 | 0,794 | 0,804 |
| 3 | 0,912 | 0,836 | 0,801 | 0,818 |
| 4 | 0,939 | 0,850 | 0,839 | 0,844 |
| 5 | 0,951 | 0,875 | 0,872 | 0,874 |
| 6 | 0,910 | 0,766 | 0,775 | 0,771 |
| 7 | 0,927 | 0,826 | 0,834 | 0,830 |
| 8 | 0,928 | 0,827 | 0,834 | 0,831 |
| 9 | 0,923 | 0,845 | 0,832 | 0,838 |
| 10 | 0,948 | 0,864 | 0,868 | 0,866 |
| **Rata-Rata** | **0,9** | **0,834** | **0,829** | **0,832** |

#### Pengujian Pra Proses HMM Skenario Kedua

Hasil dari pengujian *accuracy, precision, recall*, dan *f-measure* dari pra proses HMMdengan korpus *tagging* pra proses HMM skenario kedua diperlihatkan pada tabel 4.3 berikut.

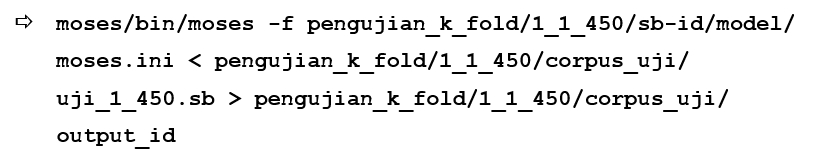
Tabel 4.3 Hasil PengujianPra Proses HMM Skenario Kedua

| **Fold** | ***Accuracy*** | ***Precision*** | ***Recall*** | ***F-Measure*** |
| --- | --- | --- | --- | --- |
| 1 | 0,763 | 0,712 | 0,675 | 0,693 |
| 2 | 0,747 | 0,696 | 0,646 | 0,670 |
| 3 | 0,704 | 0,689 | 0,636 | 0,662 |
| 4 | 0,779 | 0,737 | 0,694 | 0,715 |
| 5 | 0,784 | 0,740 | 0,711 | 0,725 |
| 6 | 0,740 | 0,621 | 0,600 | 0,610 |
| 7 | 0,743 | 0,653 | 0,623 | 0,637 |
| 8 | 0,733 | 0,653 | 0,618 | 0,635 |
| 9 | 0,717 | 0,689 | 0,637 | 0,662 |
| 10 | 0,724 | 0,660 | 0,623 | 0,641 |
| **Rata-Rata** | **0,743** | **0,685** | **0,646** | **0,665** |

Dari hasil pengujian *accuracy, precision, recall*, dan *f-measure* dari dua skenario pra proses HMM yang dilakukan maka dapat dilihat bahwa hasil pra proses dengan korpus skenario pertama memiliki nilai *accuracy, precision, recall*, dan *f-measure* yang lebih tinggi daripada hasil pra proses dengan korpus skenario kedua.

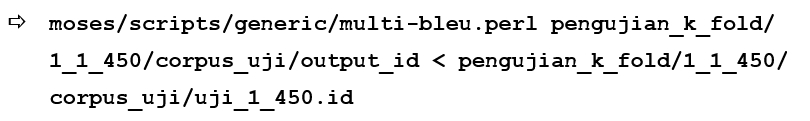
### Pengujian Hasil Penerjemahan Mesin Translasi oleh BLEU

Pengujian hasil terjemahan mesin translasi secara otomatis dilakukan menggunakan BLEU (*Billingual Evaluation Understudy*) untuk menghasilkan keluaran berupa nilai akurasi hasil terjemahan. Akurasi yang diuji pada penelitian ini adalah akurasi hasil terjemahan mesin traslasi dari bahasa Melayu Sambas ke bahasa Indonesia. Sebelum dilakukan perhitungan nilai BLEU maka dilakukan perintah untuk menghasilkan *output* atau terjemahan korpus uji dalam bahasa target (bahasa Indonesia). Contoh perintah untuk membuat hasil terjemahan ditunjukkan pada gambar 4.9 berikut.



Gambar 4.9 Contoh Perintah Membuat Hasil Terjemahan

Setelah membuat *output* atau hasil penerjemahan otomatis dari mesin penerjemah, selanjutnya adalah melakukan pengujian dengan uji evaluasi otomatis yaitu BLEU. Contoh perintah untuk menggunakan BLEU sebagai *tools* evaluasi otomatis ditunjukkan pada gambar 4.10 berikut.



Gambar 4.10 Contoh Perintah Menggunakan BLEU

Pengujian BLEU pada penelitian menggunakan metode *k-fold* *cross validation* yaitu membagi data menjadi dua subset yaitu data proses pembelajaran (korpus *training*) dan data validasi/evaluasi (korpus uji), dengan skema 10*-fold cross validation* di mana dilakukan sebanyak 10 kali iterasi/perulangan dengan korpus *training* dan korpus uji berbeda-beda. Korpus uji dan korpus *training* dari total 4500 baris kalimat dipecah ke 10 bagian di mana masing-masing bagian berisi 4050 korpus *training* dan 450 korpus uji.

Pada pengujian mesin penerjemah statistik Moses Decoder dilakukan dua pengujian BLEU yaitu pengujian pada mesin penerjemah bahasa *chatting* Melayu Sambas ke bahasa Indonesia dan mesin penerjemahan bahasa Melayu Sambas (yang sudah diperbaiki) ke bahasa Indonesia.

#### Pengujian Hasil Penerjemahan Mesin Translasi Bahasa Melayu Sambas (Chatting) ke Bahasa Indonesia

Hasil dari pengujian BLEU mesin penerjemah bahasa *chatting* Melayu Sambas ke bahasa Indonesia diperlihatkan pada tabel 4.4 berikut.

Tabel 4.4 Hasil Pengujian BLEU Penerjemahan Bahasa Melayu Sambas (*chatting*) ke Bahasa Indonesia

| **Fold** | **Data Uji** | **Hasil Pengujian** |
| --- | --- | --- |
| 1 | 1 – 450 | 34,24 |
| 2 | 451 – 900 | 46,97 |
| 3 | 901 – 1.350 | 53,2 |
| **4** | **1.351 – 1.800** | **57,85** |
| 5 | 1.801 – 2.250 | 51,09 |
| 6 | 2.251 – 2.700 | 46,56 |
| 7 | 2.701 – 3.150 | 45,27 |
| 8 | 3.151 – 3.600 | 45,57 |
| 9 | 3.601 – 4.050 | 36,9 |
| 10 | 4.051 – 4.500 | 39,55 |
| **Total** | **4500** | **457,2** |
| **Rata-Rata** | | **45,72** |

Setelah dilakukan pengujian BLEU dengan 10*-fold cross validation* diperoleh rata-rata akurasi mesin penerjemah langsung dari bahasa Melayu Sambas (*chatting*) ke bahasa Indonesia sebesar 45,72% dan nilai akurasi paling tinggi berada pada *fold* 4 dengan akurasi sebesar 57,85%.

#### Pengujian Hasil Penerjemahan Mesin Translasi Bahasa Melayu Sambas (Sudah Diperbaiki) ke Bahasa Indonesia

Hasil dari pengujian BLEU dengan 10*-fold cross validation* untuk mesin penerjemah bahasa Melayu Sambas (yang sudah diperbaiki) ke bahasa Indonesia diperlihatkan pada tabel 4.5 berikut.

Tabel 4.5 Hasil Pengujian BLEU Penerjemahan Bahasa Melayu Sambas (Sudah Diperbaiki) ke Bahasa Indonesia

| **Fold** | **Data Uji** | **Hasil Pengujian** |
| --- | --- | --- |
| 1 | 1 – 450 | 40,22 |
| 2 | 451 – 900 | 57,33 |
| **3** | **901 – 1.350** | **63,82** |
| 4 | 1.351 – 1.800 | 60,87 |
| 5 | 1.801 – 2.250 | 56,97 |
| 6 | 2.251 – 2.700 | 54,55 |
| 7 | 2.701 – 3.150 | 52,56 |
| 8 | 3.151 – 3.600 | 52,76 |
| 9 | 3.601 – 4.050 | 45,78 |
| 10 | 4.051 – 4.500 | 46,5 |
| **Total** | **4500** | **531,4** |
| **Rata-Rata** | | **53,14** |

Dari hasil pengujian dua mesin penerjemah yaitu mesin penerjemah langsung bahasa Melayu Sambas (*chatting*) ke bahasa Indonesia dan mesin penerjemah bahasa Melayu Sambas (yang sudah diperbaiki) ke bahasa Indonesia, dapat dilihat bahwa mesin dengan penerjemahan bahasa Melayu Sambas yang sudah diperbaiki memiliki nilai akurasi yang lebih tinggi daripada penerjemahan langsung bahasa *chatting* Melayu Sambas ke bahasa Indonesia.

### Pengujian Hasil Penerjemahan Mesin Translasi oleh Ahli Bahasa

Pengujian manual dilakukan oleh ahli bahasa Melayu Sambas dan dua orang yang menjadi ahli bahasa Melayu Sambas pada penelitian ini adalah Dendodi, S.Pd dan M. Sutrisno, ST. Penulis memilih mereka karena merupakan asli keturunan Melayu Sambas, dan memahami mengenai bahasa Melayu Sambas, baik penulisan, pengucapan, dan arti dalam bahasa Indonesia. Mereka juga bersedia dilibatkan dalam penelitian ini.

Pengujian manual diambil berdasarkan hasil pengujian otomatis menggunakan metode *k-fold cross validation* di mana dari hasil pengujian sebanyak 10 kali diambil nilai dengan hasil tertinggi. Adapun nilai tertinggi didapat pada *fold* ketiga mesin penerjemah bahasa Melayu Sambas (yang sudah diperbaiki) ke bahasa Indonesia, dengan korpus uji kalimat 901 sampai 1.350 dan nilai akurasi 63,82%. Setelah didapatkan nilai tertinggi maka pada *fold* atau data uji dengan nilai tertinggi dilakukan lagi pengujian sebanyak 10 kali uji menggunakan metode *k-fold cross validation* dengan pembagian data uji dan hasil pengujian yang didapat seperti terlihat pada tabel 4.7 berikut.

Tabel 4.7 Hasil Pengujian BLEU Dengan Metode *K-Fold Cross Validation* Pada Fold 3 (450 Kalimat)

| **Fold** | **Data Uji** | **Hasil Pengujian** |
| --- | --- | --- |
| 1 | 901 – 945 | 51,85 |
| 2 | 946 – 990 | 50,11 |
| 3 | 991 – 1.035 | 54,86 |
| 4 | 1.036 – 1.080 | 58,24 |
| 5 | 1.081 – 1.125 | 64,39 |
| 6 | 1.126 – 1.170 | 59,09 |
| 7 | 1.1171 – 1.215 | 59,15 |
| 8 | 1.216 – 1.260 | 58,19 |
| **9** | **1.261 – 1.305** | **64,44** |
| 10 | 1.306 – 1.350 | 61,09 |

Setelah dilakukan pengujian lanjutan yaitu pengujian BLEU dengan 10-*fold cross validation* pada 450 kalimat atau data uji yang mendapat nilai tertinggi sebelumnya maka didapat hasil seperti tabel 4.2 di atas. Hasil pengujian lanjutan di atas menunjukkan *fold* 9 dengan korpus uji kalimat 1.261 sampai 1.305 memperoleh nilai akurasi tertinggi dengan nilai akurasi sebesar 64,44%. Data uji *fold* 9 yang memperoleh nilai tertinggi akan digunakan sebagai data untuk dilakukan pengujian manual oleh ahli bahasa. Hasil dari pengujian manual oleh ahli bahasa yang menguji akurasi hasil terjemahan mesin penerjemah untuk bahasa Melayu Sambas ke bahasa Indonesia diperlihatkan pada tabel 4.8 berikut.

Tabel 4.8 Hasil Pengujian Terjemahan Mesin Penerjemah Oleh Ahli Bahasa

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No.** | **Nama Ahli Bahasa** | **C** | **R** |  |
| 1. | Dendodi, S.Pd | 304 | 350 | 86,86% |
| 2. | M. Sutrisno, ST | 307 | 350 | 87,71% |

### Pengujian *Black Box*

Metode pengujian *black box* dipilih karena metode pengujian tersebut tidak memperhatikan struktur logika internal *(coding)* dalam perangkat lunak. Pengujian ini perlu dilakukan untuk melihat respon yang diberikan oleh sistem saat melakukan proses *input* data dan apakah hasil eksekusi berhasil sesuai dengan yang diharapkan ataupun tidak berhasil. Pengujian yang dilakukan antara lain pengujian terhadap proses *login*, pengujian *update* profil, dan pengujian mengirim pesan.

1. Proses *Login*

Pada proses *login* terdapat dua bagian yaitu proses verifikasi nomor telepon pada halaman verifikasi nomor telepon dan proses verifikasi kode OTP pada halaman verifikasi OTP. Hasil pengujian proses *login* diperlihatkan pada tabel 4.9 berikut.

Tabel 4.9 Hasil Pengujian Proses *Login*

| **Fungsi** | **Kasus Uji** | **Hasil Eksekusi** | **Keterangan** |
| --- | --- | --- | --- |
| Verifikasi Nomor Telepon | Nomor telepon kosong | Tidak berhasil | Menampilkan pesan peringatan kesalahan: “Masukkan nomor telepon yang benar” |
| Nomor telepon yang dimasukkan kurang dari 10 angka | Tidak berhasil | Menampilkan pesan peringatan kesalahan: “Masukkan nomor telepon yang benar” |
| Nomor telepon yang dimasukkan lebih dari 12 angka | Tidak berhasil | Menampilkan pesan peringatan kesalahan: “Nomor anda tidak valid” |
| Nomor telepon yang dimasukkan palsu, salah, tidak aktif, atau berada pada perangkat yang berbeda | Tidak berhasil | Menampilkan pesan peringatan kesalahan: “Waktu habis !  Nomor anda tidak valid” |
| Nomor telepon yang dimasukkan benar, aktif, dan digunakan pada perangkat | Berhasil | Masuk ke halaman verifikasi OTP  Mengisi kolom OTP secara otomatis.  Kode OTP dikirim melalui SMS |
| Verifikasi Kode OTP | Mengganti kode OTP otomatis | Tidak berhasil | Menampilkan pesan peringatan kesalahan: “OTP tidak valid !  Masukkan kode yang benar” |
| Memverifikasi kode lebih dari 10 menit setelah kode diterima | Tidak berhasil | Menampilkan pesan peringatan kesalahan: “OTP tidak valid !  Masukkan kode yang benar” |
| Memverifikasi kode OTP otomatis yang dikirim dalam waktu 10 menit setelah kode diterima | Berhasil | * Masuk ke halaman *update* profil * Selanjutnya tampilan utama aplikasi merupakan halaman daftar obrolan |

1. Proses Memperbarui Profil

Memperbarui profil *user* dilakukan pada halaman *update* profil, pada halaman ini *user* dapat mengganti foto profil, nama, status, dan bahasa yang digunakan. Hasil pengujian dari proses memperbarui profil *user* ditampilkan pada tabel 4.10 berikut.

Tabel 4.10 Hasil Pengujian Proses Memperbarui Profil

| **Fungsi** | **Kasus Uji** | **Hasil Eksekusi** | **Keterangan** |
| --- | --- | --- | --- |
| Memperbarui Profil *User* | Mengosongkan kolom *input* | Tidak berhasil | Menampilkan pesan peringatan kesalahan: “Lengkapi Field Yang Kosong !!!” |
| Mengosongkan salah satu kolom *input* | Tidak berhasil | Menampilkan pesan peringatan kesalahan: “Lengkapi Field Yang Kosong !!!” |
| Mengisi semua kolom *input* | Berhasil | * Masuk ke halaman utama (jika halaman *update* profil diakses setelah *login*) * Kembali ke halaman profil saya (jika diakses melalui halaman profil saya) |

1. Proses Mengirim Pesan

Proses mengirim pesan terjadi pada halaman ruang obrolan, dan terdiri dari dua kondisi yaitu mengirim pesan oleh *user* berbahasa Indonesia dan mengirim pesan oleh *user* berbahasa Melayu Sambas. Hasil pengujian mengirim pesan ditunjukkan pada tabel 4.11 berikut.

Tabel 4.11 Hasil Pengujian Pengiriman Pesan

| **Fungsi** | **Kasus Uji** | **Hasil Eksekusi** | **Keterangan** |
| --- | --- | --- | --- |
| Mengirim Pesan  (Bahasa Indonesia) | Mengirim pesan kosong | Tidak berhasil | Pesan tidak terkirim |
| Mengirim pesan “Saya sedang makan di rumah” | Berhasil | Pesan “Saya sedang makan di rumah” terkirim |
| Mengirim Pesan  (Bahasa Melayu Sambas) | Mengirim pesan kosong | Tidak berhasil | Pesan tidak terkirim |
| Mengirim pesan “Saye gek mkan drumah” | Berhasil | Pesan diterjemahkan dan dikirim:   * “Saye gek mkan drumah” (pesan asli) * “saye agek makan dirumah” (hasil pra proses) * “saya lagi makan di rumah” (terjemahan pesan) |
| Mengirim pesan “Mmng tdok diluar tok, ngade2 juag tdok di dalm” | Berhasil | Pesan diterjemahkan dan dikirim:   * “Mmng tdok diluar tok, ngade2 juag tdok di dalm” (pesan asli) * “memang tidok diluar tok , ngade-ngade juak tidok di dalam” (hasil pra proses) * “memang tidur diluar nih, sembarangan juga tidur di dalam” (terjemahan pesan) |

## Analisis Hasil Perancangan dan Pengujian

Berikut merupakan analisis terhadap hasil pengujian yang telah dilakukan:

1. Pengujian nilai *accuracy, precision, recall*, dan *f-measure* pada hasil pra proses HMM dengan *trigram* dilakukan dengan dua skenario korpus *tagging*. Pra proses HMMdengan skenario pertama memperloleh hasil sebesar 0,9 untuk nilai *accuracy*, 0,834 untuk nilai *precision*, 0,829 untuk nilai *recall*, dan 0,832 untuk nilai *f-measure*. Sedangkan pra proses dengan skenario kedua memperoleh nilai 0,743 untuk *accuracy*, 0,685 untuk nilai *precision*, 0,646 untuk nilai *recall*, dan 0,665 untuk nilai *f-measure*. Nilai ini menunjukkan bahwa pra proses HMM dengan skenario pertama memiliki nilai *accuracy, precision, recall*, dan *f-measure* yang lebih besar daripada pra proses HMMdengan skenario kedua.
2. Penilaian otomatis menggunakan BLEU dan metode *k-fold cross validation* pada mesin pertama menghasilkan nilai rata-rata akurasi penerjemahan bahasa Melayu Sambas (*chatting*) ke bahasa Indonesia sebesar 45,72%, sedangkan mesin kedua menghasilkan nilai rata-rata akurasi penerjemahan bahasa Melayu Sambas (baku) ke bahasa Indonesia sebesar 53,14%. Nilai ini menunjukkan bahwa mesin penerjemah bahasa Melayu Sambas (baku/sudah diperbaiki) ke bahasa Indonesia memiliki rata-rata nilai akurasi yang lebih tinggi daripada mesin penerjemah langsung bahasa Melayu Sambas (*chatting*) ke bahasa Indonesia.
3. Tinggi rendahnya nilai akurasi didapat berdasarkan kemiripan kata atau kalimat antara kalimat uji terhadap kalimat referensinya pada setiap *fold* pengujian. Untuk meningkatkan nilai akurasi ini dapat dilakukan dengan cara diantaranya menambah kuantitas korpus dan melakukan perbaikan kualitas korpus (konsistensi terjemahan korpus paralel).
4. Berdasarkan skema 10-*fold* *cross validation* diperoleh nilai akurasi terjemahan tertinggi pada mesin kedua pada *fold* 3 yang menggunakan data uji kalimat 901 sampai 1.350 dengan nilai 63,82% dan nilai akurasi terjemahan paling rendah adalah pada *fold* 1 yang menggunakan data uji kalimat 1 sampai 450 dengan nilai akurasi 40,22%. Berdasarkan penelitian Apriani, Novi, dan Sujaini (2016)menyatakan bahwa “kalimat referensi sangat memengaruhi penilaian, semakin dekat jumlah kata hasil terjemahan dengan kalimat referensinya maka akan menghasikan nilai yang semakin baik”. Hasil ini menunjukkan bahwa korpus uji (hasil terjemahan) pada *fold* 3 memiliki kemiripan terhadap korpus referensinya lebih besar daripada 9 *fold* lainnya.
5. Ahli bahasa melakukan pengujian manual terhadap hasil uji terbaik pengujian otomatis (BLEU dengan 10-*fold* *cross validation* ) pada *fold* 3 dengan nilai akurasi terbaik 64,44%. Hasil pengujian oleh ahli bahasa masing-masing pengujian oleh Dendodi, S.Pd menunjukkan nilai akurasi sebesar 86,86% dan pengujian oleh M. Sutrisno, ST menunjukkan nilai akurasi sebesar 87,71%. Hasil pengujian ini menunjukkan terjadinya peningkatan nilai akurasi yang dari ahli bahasa terhadap nilai akurasi dari pengujian otomatis. Terjadinya peningkatan nilai akurasi ini dikarenakan pengujian oleh ahli bahasa (manusia) memiliki nilai toleransi pemahaman terhadap hasil penerjemahan, berbeda dengan nilai akurasi dari pengujian mesin yang sudah terdekte.
6. Secara keseluruhan pengujian *black box* menunjukkan bahwa aplikasi dapat meng*handle* proses *input* dan melakukan respon dengan baik pada proses *login*, *update* profil, dan mengirim pesan. Aplikasi dapat memberikan respon peringatan dengan baik ketika *user* memberikan *input* kosong atapun *input* yang tidak valid. Dari pengujian ini juga dapat dilihat jika aplikasi dapat meng*handle* pengiriman pesan, pesan berbahasa Melayu Sambas direspon dengan penerjemahan dan pengiriman pesan dalam bahasa Indonesia.

# PENUTUP

## Kesimpulan

Berdasarkan hasil analisis dan pengujian yang dilakukan terhadap mesin penerjemah statistik dan aplikasi yang dibangun, maka dapat ditarik kesimpulan bahwa pembuatan aplikasi *chatting* berbasis android yang mengimplementasikan mesin penerjemah statistik Moses Decoder dan aplikasi dapat digunakan sebagai media komunikasi teks sudah berhasil dilakukan.

Kesimpulan ini berdasarkan keberhasilan dalam pembuatan aplikasi *chatting* yang mampu menerjemahkan pesan bahasa Melayu Sambas ke bahasa Indonesia. Hasil akurasi penerjemahan langsung mesin penerjemah statistik bahasa Melayu Sambas (*chatting*) ke bahasa Indonesia sebesar 45,72%, dan hasil akurasi penerjemahan dengan perbaikan kata mesin penerjemah statistik bahasa Melayu Sambas (baku) ke bahasa Indonesia sebesar 54,14%. Sedangkan berdasarkan pengujian ahli bahasa untuk mesin kedua (penerjemahan dengan pra proses) memperoleh nilai persentase akurasi sebesar 86,86% dan 87,71%. Hasil pra proses HMM tertinggi (skenario pertama) memperoleh rincian hasil sebesar 0,9 untuk nilai *accuracy*, 0,834 untuk nilai *precision*, 0,829 untuk nilai *recall*, dan 0,832 untuk nilai *f-measure*.

## Saran

Adapun beberapa hal yang disarankan dalam pengembangan dari aplikasi atau penelitian ini kedepannya adalah sebagai berikut:

1. Peningkatan kualitas dan kuantitas dari korpus *tagging* sehingga dapat melakukan *preprocessing* dengan lebih baik, yaitu mampu memperbaiki lebih banyak variasi teks *chatting*.
2. Peningkatan kualitas dan kuantitas korpus *training* sehingga dapat meningkatkan kualitas akurasi terjemahan mesin penerjemah statistik.
3. Menambah jenis bahasa yang dapat dilakukan pra proses dan penerjemahan sehingga dapat digunakan oleh pengguna bahasa yang lebih beragam.
4. Pengembangan aplikasi sehingga menjadi aplikasi *chatting* yang lebih lengkap seperti adanya fitur menarik/menghapus pesan, fitur melihat status pesan, fitur mengirim pesan gambar, fitur mengirim pesan suara, dan lain sebagainya.
5. Inovasi seperti penerapan kombinasi *text-to-speech* dan mesin penerjemah statistik sehingga dapat mengirim pesan suara yang kemudian diterjemahkan dan mengasilkan pesan suara dengan bahasa yang digunakan penerima pesan.

DAFTAR PUSTAKA

Aditya, S. T. (2018) *Rancang Bangun Aplikasi Kampanye Berbasis Crowdsourcing*. Universitas Tanjungpura.

Andi Juansyah (2015) ‘Pembangunan Aplikasi Child Tracker Berbasis Assisted – Global Positioning System ( A-GPS ) Dengan Platform Android’, *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, 1(1). Available at: elib.unikom.ac.id/download.php?id=300375.

Android Developer (no date) *Mengenal Android Studio*, *Android Developer*. Available at: https://developer.android.com/studio/intro/?hl=id (Accessed: 1 October 2019).

Antinasari, P., Perdana, R. S. and Fauzi, M. A. (2017) ‘Analisis Sentimen Tentang Opini Film Pada Dokumen Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), pp. 1718–1724.

Baker, M. (1995) ‘Corpora in Translation Studies’, *Target. International Journal of Translation Studies*, 7(2), pp. 223–243. doi: 10.1075/target.7.2.03bak.

Cenderamata, R. C. and Sofyan, A. N. (2019) ‘Abreviasi dalam Percakapan Sehari-Hari di Media Sosial: Suatu Kajian Morfologi’, *Metahumaniora*, 4(1). doi: 10.24198/mh.v8i2.20699.

Dharmawan, E., Sujaini, H. and Muhardi, H. (2020) ‘Perbandingan Nilai Akurasi Terhadap Penggunaan Part of Speech Set pada Mesin Penerjemah Statistik Comparison of The Accuracy Value Toward Using Part of Speech Sets on Statistical Machine Translation’, 08(3), pp. 250–256. doi: 10.26418/justin.v8i3.39810.

Dharwiyanti, S. and Wahono, R. S. (2003) ‘Pengantar Unified Modeling Language (UML)’, *IlmuKomputer.com*. Available at: http://www.unej.ac.id/pdf/yanti-uml.pdf.

Ginting, A. and AZ, N. (2012) ‘PENERJEMAH DUA ARAH BAHASA INDONESIA KE BAHASA DAERAH ( KARO ) MENGGUNAKAN TEKNIK STATISTICAL MACHINE TRANSLATION ( SMT ) SEBAGAI FITUR PADA SITUS WEB UNTUK MENINGKATKAN WEB TRAFFIC’, *TELEMATIKA MKOM Universitas Budi Luhur*, 4(1).

Hadi, I. (2014) ‘UJI AKURASI MESIN PENERJEMAH STATISTIK (MPS) BAHASA INDONESIA KE BAHASA MELAYU SAMBAS DAN MESIN PENERJEMAH STATISTIK (MPS) BAHASA MELAYU SAMBAS KE BAHASA INDONESIA Ibnu Hadi’, *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 2(3).

Hardianto, F. and Handaga, B. (2015) *APLIKASI GRUPCHAT DI ANDROID MENGGUNAKAN WEBSOCKET*.

Hariyanto, B. (2010) ‘ISTILAH-ISTILAH KHUSUS DALAM CHATTING ( Sebuah Analisis Sosiopragmatik )’, *Adabiyyāt: Jurnal Bahasa dan Sastra*, 9(3).

Hidayat, A., Sujaini, H. and Dwinyoto, R. (2015) ‘APLIKASI PENERJEMAH DUA ARAH BAHASA INDONESIA – BAHASA MELAYU SAMBAS BERBASIS WEB DENGAN MENGGUNAKAN DECODER MOSES’, *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 3(3).

Loper, E. and Bird, S. (2002) ‘NLTK: The Natural Language Toolkit’, *Proceedings of the COLING/ACL on Interactive presentation sessions -*. Morristown, NJ, USA: Association for Computational Linguistics, pp. 69–72. doi: 10.3115/1225403.1225421.

McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-based language studies: An advanced resource book*, *Language and Literature: International Journal of Stylistics*. Routledge.

Nugraha, I. G. B. B. and Rizqullah, R. D. (2019) ‘Normalisasi Kata Tidak Baku yang Tidak Disingkat dengan Jarak Perubahan’, *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 8(3), p. 218. doi: 10.22146/jnteti.v8i3.516.

Prasetyo, M. E. B. (2010) ‘Teori Dasar Hidden Markov Model’, *Makalah Sekolah Teknik Elektro dan Informatika ITB*.

Pratiwi, N. F., Sujaini, H. and Nyoto, R. D. (2017) ‘Pengembangan Antarmuka Mesin Penerjemah Statistik Multibahasa Berbasis Web’, *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 5(1).

Priansya, S. (2017) ‘Normalisasi Teks Media Sosial Menggunakan Word2vec, Levenshtein Distance, dan Jaro-Winkler Distance’, *Institut Teknologi Sepuluh Nopember*, p. Tugas Akhir-KS141501.

Putri, L. N. K. rnia (2011) *Analisis Perancangan dan Implementasi Aplikasi Chatting Berbasis Objek*. Uin Sunan Kalijaga.

Ropianto, M. (2016) ‘Pemahaman Penggunaan Unified Modelling Language’, *Jt-Ibsi*, 1(1).

Sahara, H. (2018) ‘Implementasi Pengamanan Pesan Chatting menggunakan Metode Vigenere Cipher dan Cipher Block Chainning’, 3(2).

Syaukani, M. (2010) ‘Sistem Penerjemah Inggris-Indonesia Pada Aplikasi Chatting Berbasis Web Menggunakan Pendekatan Aturan’, *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 8(2). doi: 10.12928/telkomnika.v8i2.616.

Tanuwijaya, H. and Manurung, H. M. (2009) ‘PENERJEMAHAN DOKUMEN INGGRIS-INDONESIA MENGGUNAKAN MESIN PENERJEMAH STATISTIK DENGAN WORD REORDERING DAN PHRASE REORDERING’, *Jurnal Ilmu Komputer dan Informasi*, 2(1).

Utami, D. (2010) ‘Karakteristik penggunaan bahasa pada status Facebook’.

Wahyudinata, T., Sujaini, H. and Nyoto, R. D. (2016) ‘IMPLEMENTASI MESIN PENERJEMAH STATISTIK BERBASIS ANDROID DENGAN MOSES DECODER’, *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 3(1).

Wicaksono, A. F. and Ayu, P. (2010) ‘HMM Based Part-of-Speech Tagger for Bahasa Indonesia’, in *Prooceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop*.

Wijayanto, B. (2012) *Merancang dan Membangun Aplikasi Chat Messenger Untuk Android*, *Jurusan Teknik Informatika STMIK AMIKOM YOGYAKARTA*.

Xiong, J. *et al.* (2015) ‘Extended HMM and Ranking Models for Chinese Spelling Correction’, (October), pp. 133–138. doi: 10.3115/v1/w14-6821.

LAMPIRAN A  
Contoh Korpus *Tagging* HMM Skenario 1

Lampiran A 1 Contoh Korpus *Tagging*

1. kecamatan/bbb ap/ape bg/bang ./.
2. bwak/bawak lah/bbb sigek/bbb k/ke kantor/bbb bg/bang ./.
3. makan/bbb d/di t4/tempat nymn/nyaman wak/bbb ./.
4. an/daan bse/bise pakai/bbb bawa/bbb ./.
5. mn/mun makan/bbb d/di tmptny/tempatnye kalak/bbb payah/bbb nk/nak balik/bbb bg/bang ./.
6. llu/lalu payh/payah be/bbb ./.
7. naik/bbb jx/jak ats/atas mtr/motor be/bbb ./.
8. kalak/bbb kekanyangan/bbb payah/bbb nk/nak balik/bbb bg/bang ./.
9. mun/bbb dah/bbb lakk/lakak cek/bbb i/bbb ./.
10. biak/bbb semunut/bbb bg/bang kurang/bbb akrab/bbb ./.
11. kamek/bbb daan/bbb berani/bbb bg/bang nk/nak simbarangan/bbb ./.
12. tunggu/bbb lok/dolok bg/bang ./.
13. gmn/gimane hsilnye/hasilnye gus/bbb ./.
14. didik/bbb bmtr/bemotor dg/dengan sape/bbb ./.
15. dgn/dengan cwek/cewek nye/bbb didik/bbb wak/bbb ./.
16. cwek/cewek nye/bbb lah/bbb ,/, aku/bbb dgn/dengan joko/bbb tok/bbb ee/bbb ./.
17. biak/bbb sambas/bbb cewek/bbb nye/bbb wak/bbb ./.
18. barok/bbb kini/bbb maing/bbb tok/bbb ee/bbb ./.
19. kamek/bbb di/bbb dakat/bbb pintu/bbb masok/bbb tok/bbb ee/bbb ./.
20. mne/mane w/we ,/, aku/bbb d/di blkng/belakang gwang/gawang ./.
21. intah/bbb lah/bbb tok/bbb ee/bbb ari/bbb yo/bbb galap/bbb nak/bbb ujan/bbb wak/bbb ./.
22. daan/bbb ,/, ade/bbb di/bbb luar/bbb kalau/bbb ndak/bbb salah/bbb ye/bbb wak/bbb ./.
23. ha/bbb isk/isok aku/bbb mnjam/minjam ./.
24. untkan/untokan maing/bbb dg/dengan urng/urang kantor/bbb ./.
25. daan/bbb k/ke tapi/bbb sepatu/bbb mu/bbb kenak/bbb pakai/bbb lah/bbb biak/bbb ./.

LAMPIRAN B  
Contoh Korpus *Tagging* HMM Skenario 2

Lampiran A 1 Contoh Korpus *Tagging*

1. kecamatan/kecamatan ap/ape bg/bang ./.
2. bwak/bawak lah/lah sigek/sigek k/ke kantor/kantor bg/bang ./.
3. makan/makan d/di t4/tempat nymn/nyaman wak/wak ./.
4. an/daan bse/bise pakai/pakai bawa/bbb ./.
5. mn/mun makan/makan d/di tmptny/tempatnye kalak/kalak payah/payah nk/nak balik/balik bg/bang ./.
6. llu/lalu payh/payah be/be ./.
7. naik/naik jx/jak ats/atas mtr/motor be/be ./.
8. kalak/bbb kekanyangan/bbb payah/payah nk/nak balik/balik bg/bang ./.
9. mun/bbb dah/dah lakk/lakak cek/cek i/bbb ./.
10. biak/bbb semunut/semunut bg/bang kurang/kurang akrab/bbb ./.
11. kamek/bbb daan/bbb berani/berani bg/bang nk/nak simbarangan/simbarangan ./.
12. tunggu/tunggu lok/dolok bg/bang ./.
13. gmn/gimane hsilnye/hasilnye gus/gus ./.
14. didik/didik bmtr/bemotor dg/dengan sape/sape ./.
15. dgn/dengan cwek/cewek nye/nye didik/bbb wak/bbb ./.
16. cwek/cewek nye/nye lah/bbb ,/, aku/aku dgn/dengan joko/joko tok/bbb ee/bbb ./.
17. biak/bbb sambas/bbb cewek/bbb nye/bbb wak/bbb ./.
18. barok/bbb kini/bbb maing/bbb tok/bbb ee/bbb ./.
19. kamek/bbb di/bbb dakat/bbb pintu/bbb masok/bbb tok/bbb ee/bbb ./.
20. mne/mane w/we ,/, aku/aku d/di blkng/belakang gwang/gawang ./.
21. intah/bbb lah/bbb tok/bbb ee/bbb ari/bbb yo/bbb galap/bbb nak/bbb ujan/bbb wak/bbb ./.
22. daan/bbb ,/, ade/bbb di/bbb luar/bbb kalau/bbb ndak/bbb salah/bbb ye/bbb wak/bbb ./.
23. ha/ha isk/isok aku/aku mnjam/minjam ./.
24. untkan/untokan maing/maing dg/dengan urng/urang kantor/kantor ./.
25. daan/daan k/ke tapi/tapi sepatu/bbb mu/bbb kenak/bbb pakai/bbb lah/bbb biak/bbb ./.

LAMPIRAN C  
Contoh Korpus Paralel Bahasa Indonesia − Bahasa Melayu Sambas

LAMPIRAN B 1 Contoh Korpus Paralel Bahasa Indonesia − Bahasa Melayu Sambas

|  |  |
| --- | --- |
| **Bahasa Indonesia** | **Bahasa Melayu Sambas** |
| bawak lah sigek ke kantor bang | bawalah satu ke kantor bang |
| mun makan di tempatnye kalak payah nak balik bang | kalau makan di tempatnya nanti susah mau pulang bang |
| kalak kekanyangan payah nak balik bang | nanti kekenyangan susah mau pulang bang |
| kamek daan berani bang nak simbarangan | kami tidak berani mau sembarangan bang |
| kau balom nak ke pontianak ke | kamu belum mau ke pontianak |
| eh indok aku madahkan ye , gerogi aku mun kau nonton | eh tidak mau aku mengatakan itu , grogi aku kalau kamu tonton |
| tunggu lok aku nanyak biak agek kamek tadek nak ngantar surat | tunggu dulu aku tanya mereka lagi kami tadi mau mengantar surat |
| nak nanyak direk hujan ke digaling boss | ingin bertanya dengan kamu apakah hujan di galing boss |
| ape kabar kau tok e wak , lamak daan jumpe kite | apa kabar kamu sekarang ini wak , lama tidak bertemu kita |
| kite jumpe di simpang sepakat ajak kalak malam | kita bertemu di simpang sepakat saja nanti malam |
| bukan gie yang di maksud ibu be ade kayak kuesioner nye gie | bukan begitu yang dimaksud ibu tuh ada seperti kuesionernya begitu |
| daan sih jak barok tadek datang nye tapi takut layok sih | tidak sih hanya baru tadi datang nya tapi takut layu sih |
| daan ngape ngape sih , atorannye aku nak minjam laptop mu bentar | tidak kenapa-kenapa , sebenarnya saya mau pinjam laptop kamu sebentar |
| mane nunggu kau lamak gilak wik barok nak bejalan | manalah menunggu kamu terlalu lama wik baru mau jalan-jalan |
| untong rumah di sitok dah daan agek di masokek tikus | beruntung rumah di sini sudah tidak lagi dimasuki tikus |
| tadek aku dah nak pagi kau lamak balasnya ke asrama agek aku | tadi aku sudah mau pergi kamu lama balasnya ke asrama lagi aku |
| mun memang maok malam itok yang bise | Kalau memang mau malam ini yang bisa |

LAMPIRAN D  
*Service* Penerjemah PHP

LAMPIRAN C 1 *Service* Penerjemah PHP

Pada *service* penerjemah yang dibangun melakukan *request* ke mesin HMM dan mesin penerjemah statistik Moses Decoder. Beberapa bagian dari *service* php yang dibangun antara lain :

* + - 1. File **token\_tanda\_baca.php** yang berfungsi untuk menangani pesan yang memiliki karakter spasi, titik, koma, tanda tanya, maupun *enter* yang berlebihan pada pesan

<?php

**function** pisahTB($string) {

$string = **str\_replace**(".", " .", $string);

$string = **str\_replace**(",", " ,", $string);

$string = **str\_replace**("?", " ?", $string);

$string = **str\_replace**("!", " !", $string);

$string = **str\_replace**("\\n", " \\n ", $string);

$string = **str\_replace**("\\n", "\\n &&ent&& \\n", $string);

$string = bersihkanTB($string);

$string = tambahEnter($string);

**return** $string;

}

**function** gabungTB($string) {

$string = **str\_replace**(" .", ".", $string);

$string = **str\_replace**(" ,", ",", $string);

$string = **str\_replace**(" ?", "?", $string);

$string = **str\_replace**(" !", "!", $string);

**return** $string;

}

**function** bersihkanTB($string) {

**while** (**strpos**($string, ' ') || **strpos**($string, '..') || **strpos**($string, ',,') || **strpos**($string, '??') || **strpos**($string, '!!')){

$string = **str\_replace**(" ", " ", $string);

$string = **str\_replace**("..", ".", $string);

$string = **str\_replace**(",,", ",", $string);

$string = **str\_replace**("??", "?", $string);

$string = **str\_replace**("!!", "!", $string);

}

**while** (**strpos**($string, '\\n \\n')){

$string = **str\_replace**("\\n \\n", "\\n", $string);

}

**return** $string;

}

**function** tambahEnter($string) {

$string = **str\_replace**(". ", ". \\n ", $string);

$string = **str\_replace**("? ", "? \\n ", $string);

$string = **str\_replace**("! ", "! \\n ", $string);

$string = **str\_replace**("\\n \\n", "\\n &&ent&& \\n", $string);

**return** $string;

}

**function** bersihkanEnter($string) {

**while** (**strpos**($string, '\\n \\n')){

$string = **str\_replace**("\\n \\n", "\\n", $string);

}

**return** $string;

}

?>

* + - 1. Perintah untuk melakukan *preprocessing* HMM

**exec**("cd '/home/ubuntu/skripsi/IPOSTAgger\_v1.1/'; java IPostaggerCli \"{$masukan}\" ", $result);

* + - 1. Perintah untuk menerjemahkan pesan menggunakan Moses Decoder

**exec**("echo \"{$result[$i]}\" | nc {$ip} {$port\_sb\_id}", $result2[$i]);

* + - 1. Isi lengka dari file **terjemah.php**

<?php

**include** 'token\_tanda\_baca.php';

$masukan = "";

$keluaran = "";

$tmpMasukan = **array**();

$masukan1 = $\_POST['textBahasaMasukan'];

$ip = "127.0.0.1";

$port\_sbc\_sb = "2501";

$port\_sb\_id = "2502";

$masukan = **strtolower**($masukan1);

$masukan = **trim**($masukan);

$masukan = bersihkanTB($masukan);

$masukan = pisahTB($masukan);

**if** (**empty**($masukan)) {

**echo** "kosong";

} **else** {

$result = **null**;

$result2 = **null**;

**exec**("cd '/home/ubuntu/skripsi/IPOSTAgger\_v1.1/'; java IPostaggerCli \"{$masukan}\" ", $result);

**for** ($i = 0; $i < **count**($result); $i++) {

**exec**("echo \"{$result[$i]}\" | nc {$ip} {$port\_sb\_id}", $result2[$i]);

$result2[$i] = gabungTB($result2[$i]);

}

$separated = **implode**(" ", **array\_filter**(**array\_map**(**function**($v){

**return** **implode**(" ", **array\_filter**($v));

}, $result2)));

**while** (**strpos**($separated, ' &&ent&& &&ent&& ')){

$separated = **str\_replace**(" &&ent&& &&ent&& ", " &&ent&& ", $separated);

}

$result2 = **explode**(" &&ent&& ",$separated);

$post\_data = **array**(

'masukan' => $masukan,

'output' => $separated,

'pra-proses' => **array**(

'sbc->sb' => $result

),

'hasil' => **array**(

'sb->id' => $result2

)

);

**echo** **json\_encode**($post\_data);

}

?>

LAMPIRAN E  
Contoh Hasil Terjemahan Langsung Moses Decoder   
(Bahasa Sambas *Chatting* ke Bahasa Indonesia)

|  |  |  |
| --- | --- | --- |
| **Masukan** | **Hasil Terjemahan** | **Terjemahan Manual** |
| Pun klk acare d rumah kkde pun kkde ajak malm maok k | nanti kalau acara di rumah kakak kalau kakak saja malam mau enggak | nanti kalau acara di rumah kakak kalau kakak saja malam mau enggak |
| Assalamualaikum bngde pagi lah ari minggu nikahan along ee | assalamualaikum abang pergi lah hari minggu nikahan along tuh | assalamualaikum bangde pergi lah hari minggu nikahan along tuh |
| Urusan kerje k ape ratinye | urusan kerja apa memangnya ya | urusan kerja atau apa memangnya |
| Daan bakalan turun nampaknye maksu, Soalnye daan ad nak keluar tok e | tidak bakalan turun kelihatannya bibi, soalnya tidak ada mau keluar nih | tidak bakalan turun kelihatannya bibi, soalnya tidak ada mau keluar nih |
| Aaaa iyelah maksu blm tau tok e, Sean mseh kanak telepon tok e | ha itulah bibi belum tahu nih, tidak ada masih kena telepon nih | ha itulah bibi belum tahu nih, tidak ada masih kena telepon nih |
| tdek nk ngjakek makan sate yo d pasar | tadi mau mengajak makan sate tuh di pasar | tadi mau mengajak makan sate tuh di pasar |
| Dah kupadahkan k die dolok wak,, untok birisan nurjannah,, katenye direk sekalian nak minta talangkan ee?? | sudah kupadahkan ke dia dulu wak, untuk birisan nurjannah, katanya kamu sekalian mau minta talangkan tuh? | sudah aku bilang ke dia dulu wak, untuk birisan nurjannah, katanya kamu sekalian mau minta talangkan tuh? |
| dn smpt smri nk k sambas msh ade yg nak d kerjekn | tidak sempat kemarin mau ke sambas masih ada yang mau dikerjakan | tidak sempat kemarin mau ke sambas masih ada yang mau dikerjakan |
| Posisi dmne y, Taman d rumah sakit Antonius wak | posisi dimana tuh, saya di rumah sakit antonius wak | posisi dimana itu, saya ada di rumah sakit antonius wak |
| Klau bise d pasti kn ajk x Wak, spye sanang urng Laing Gie ae nk ngntar ny | kalau bisa di pasti kan saja x wak, supaya mudah orang lain begitu tuh mau mengantar nya | kalau bisa di pasti kan saja barangkali wak, biar mudah orang lain mau mengantar nya begitu |
| Bayarny wak, tdek dah masuk soalny d stok agk eror ATM BRI | bayarnya wak, tadi sudah masuk soalnya di sini lagi eror atm bri | bayarnya wak, tadi sudah masuk soalnya di sini lagi eror atm bri |
| Wak maaf semalam aq ad tamu d asrama jdi ndak bise ke tempat drek | maaf wak semalam aku ada tamu di asrama jadi tidak bisa ke tempat kamu | maaf wak semalam aku ada tamu di asrama jadi tidak bisa ke tempat kamu |
| Okelah, tpi aq Daan bise keluar tok y Wak nak ngmbek ny, Maaf inyn jdi dah merepotkan e drek | okelah, tapi aku tidak bisa keluar nih wak mau mengambilnya, maaf benar jadi sudah merepotkan kamu tuh | okelah, tapi aku tidak bisa keluar nih wak mau mengambilnya, maaf sekali jadi sudah merepotkan kamu tuh |
| ag di mane? Kmek k sie ie. Tp itok mseh di jln nk balik juak | lagi di mana? aku ke situ. ya tapi ini masih di jalan mau pulang juga | lagi di mana? aku ke situ ya. tapi ini masih di jalan mau pulang juga |
| Iye lah kk pun gye, Isok nk ngambekny nk jm berape kk | iyalah kalau begitu kak, besok mau mengambilnya mau pukul berapa kak | iya lah kalau begitu kak, besok mau mengambilnya mau pukul berapa kak |
| Ngntarny jk, Soalny nk longsong ke sambas, Rumahny di mne kk | ngntarny saja, soalnya mau longsong ke sambas, rumahny dimana kak | mengantarnya saja, soalnya mau sekalian ke sambas, rumahnya di mana kak |
| dn smpt smri nk k sambas msh ade yg nak d kerjekn | tidak sempat kemarin mau ke sambas masih ada yang mau dikerjakan | tidak sempat kemarin mau ke sambas masih ada yang mau dikerjakan |
| ayah msh d ICU, tpi kte perawat udh bagus juak,, cuman jantung jak agk | ayah masih di icu, tapi kata perawat sudah bagus juga, cuman jantung saja lagi | ayah masih di icu, tapi kata perawat sudah bagus juga, cuman jantung saja lagi |
| Dh lmk gasaknye, Kalak lh mun sempat k sie bng | sudah lmk gasaknye, nanti lah kalau sempat ke situ bang | sudah lama ternyata, nanti lah kalau sempat ke sana bang |
| Sian no hp nye bng, Isok lh madahakn k bpk, Nantek bpk yg nyampaikannye | tidak ada nomor hp nya bang, besoklah madahakn ke bapak, nanti bapak yang nyampaikannye | tidak ada nomor hp nya bang, besoklah kasih tahu ke bapak, nanti bapak yang menyampaikannya |
| Isok msh ade kegiatan bng sampai sore, 3 hari kegiatan tok e, isok terakher | besok masih ada kegiatan bang sampai sore, 3 hari kegiatan nih, besok terakhir | besok masih ada kegiatan bang sampai sore, 3 hari kegiatan nih, besok terakhir |
| Mun bise abng buatkn, Ade kwan2 juak maok mintak buatkan | kalau bisa abang buatkan, ada kawan-kawan juga mau minta buatkan | kalau bisa abang buatkan, ada kawan-kawan juga mau minta buatkan |
| Abng dh nk balik k, Mun balik bawak jk dolok | abang sudah mau pulang ke, kalau pulang bawa saja dulu | abang sudah mau pulang ke, kalau pulang bawa saja dulu |
| Sian sih die madahkan tabel atau ndak bng, Cuman di suruhnye buat ringkasan jak | tidak ada sih dia bilang tabel atau tidak bang, cuman di suruhnye buat ringkasan saja | tidak ada sih dia bilang tabel atau tidak bang, cuman di perintahnya buat ringkasan saja |
| Ok wak mseh d Mega mall, Drek dah nak makan e? | ok wak masih di mega mall, kamu sudah mau makan tuh? | ok wak masih di mega mall, kamu sudah mau makan ya? |

LAMPIRAN F  
Contoh Hasil Terjemahan Moses Decoder Dengan Pra Proses HMM  
(Bahasa Sambas ke Bahasa Indonesia)

| **Masukan** | **Hasil Terjemahan** | **Terjemahan Manual** |
| --- | --- | --- |
| Pun klk acare d rumah kkde pun kkde ajak malm maok k | nanti kalau acara di rumah kakak kalau kakak saja malam mau enggak | nanti kalau acara di rumah kakak kalau kakak saja malam mau enggak |
| Assalamualaikum bngde pagi lah ari minggu nikahan along ee | assalamualaikum bangde pergi lah hari minggu nikahan along tuh | assalamualaikum bangde pergi lah hari minggu nikahan along tuh |
| Urusan kerje k ape ratinye | urusan kerja apa memangnya ya | urusan kerja atau apa memangnya |
| Daan bakalan turun nampaknye maksu, Soalnye daan ad nak keluar tok e | tidak bakalan turun kelihatannya bibi, soalnya tidak ada mau keluar nih | tidak bakalan turun kelihatannya bibi, soalnya tidak ada mau keluar nih |
| Aaaa iyelah maksu blm tau tok e, Sean mseh kanak telepon tok e | ha itulah bibi belum tahu nih, tidak ada masih kena telepon nih | ha itulah bibi belum tahu nih, tidak ada masih kena telepon nih |
| tdek nk ngjakek makan sate yo d pasar | tadi mau mengajak makan sate tuh di pasar | tadi mau mengajak makan sate tuh di pasar |
| Dah kupadahkan k die dolok wak,, untok birisan nurjannah,, katenye direk sekalian nak minta talangkan ee?? | sudah aku bilang ke dia dulu wak, untuk birisan nurjannah, katanya kamu sekalian mau minta talangkan tuh? | sudah aku bilang ke dia dulu wak, untuk birisan nurjannah, katanya kamu sekalian mau minta talangkan tuh? |
| dn smpt smri nk k sambas msh ade yg nak d kerjekn | tidak sempat kemarin mau ke sambas masih ada yang mau dikerjakan | tidak sempat kemarin mau ke sambas masih ada yang mau dikerjakan |
| Posisi dmne y, Taman d rumah sakit Antonius wak | posisi dimana itu, saya ada di rumah sakit antonius wak | posisi dimana itu, saya ada di rumah sakit antonius wak |
| Klau bise d pasti kn ajk x Wak, spye sanang urng Laing Gie ae nk ngntar ny | kalau bisa di pasti kan saja x wak, biar mudah orang lain begitu tuh mau mengantar nya | kalau bisa di pasti kan saja barangkali wak, biar mudah orang lain mau mengantar nya begitu |
| Bayarny wak, tdek dah masuk soalny d stok agk eror ATM BRI | bayarnya wak, tadi sudah masuk soalnya di sini lagi eror atm bri | bayarnya wak, tadi sudah masuk soalnya di sini lagi eror atm bri |
| Wak maaf semalam aq ad tamu d asrama jdi ndak bise ke tempat drek | maaf wak semalam aku ada tamu di asrama jadi tidak bisa ke tempat kamu | maaf wak semalam aku ada tamu di asrama jadi tidak bisa ke tempat kamu |
| Okelah, tpi aq Daan bise keluar tok y Wak nak ngmbek ny, Maaf inyn jdi dah merepotkan e drek | okelah, tapi aku tidak bisa keluar nih wak mau mengambilnya, maaf sekali sudah jadi merepotkan kamu tuh | okelah, tapi aku tidak bisa keluar nih wak mau mengambilnya, maaf sekali jadi sudah merepotkan kamu tuh |
| ag di mane? Kmek k sie ie. Tp itok mseh di jln nk balik juak | lagi di mana? aku ke situ ya. tapi ini masih di jalan mau pulang juga | lagi di mana? aku ke situ ya. tapi ini masih di jalan mau pulang juga |
| Iye lah kk pun gye, Isok nk ngambekny nk jm berape kk | iya lah kalau begitu kak, besok mau mengambilnya mau pukul berapa kak | iya lah kalau begitu kak, besok mau mengambilnya mau pukul berapa kak |
| Ngntarny jk, Soalny nk longsong ke sambas, Rumahny di mne kk | katanya saja, soalnya mau longsong ke sambas, hanya di mana kak | mengantarnya saja, soalnya mau sekalian ke sambas, rumahnya di mana kak |
| dn smpt smri nk k sambas msh ade yg nak d kerjekn | tidak sempat kemarin mau ke sambas masih ada yang mau dikerjakan | tidak sempat kemarin mau ke sambas masih ada yang mau dikerjakan |
| ayah msh d ICU, tpi kte perawat udh bagus juak,, cuman jantung jak agk | ayah masih di icu, tapi kata perawat sudah bagus juga, cuman jantung saja lagi | ayah masih di icu, tapi kata perawat sudah bagus juga, cuman jantung saja lagi |
| Dh lmk gasaknye, Kalak lh mun sempat k sie bng | sudah menerima gasaknye, nanti lah kalau sempat ke situ bang | sudah lama ternyata, nanti lah kalau sempat ke sana bang |
| Sian no hp nye bng, Isok lh madahakn k bpk, Nantek bpk yg nyampaikannye | tidak ada nomor hp nya bang, besoklah kasih tahu ke bapak, nanti bapak yang nyampaikannye | tidak ada nomor hp nya bang, besoklah kasih tahu ke bapak, nanti bapak yang menyampaikannya |
| Isok msh ade kegiatan bng sampai sore, 3 hari kegiatan tok e, isok terakher | besok masih ada kegiatan bang sampai sore, 3 hari kegiatan nih, besok terakhir | besok masih ada kegiatan bang sampai sore, 3 hari kegiatan nih, besok terakhir |
| Mun bise abng buatkn, Ade kwan2 juak maok mintak buatkan | kalau bisa abang buatkan, ada kawan-kawan juga mau minta buatkan | kalau bisa abang buatkan, ada kawan-kawan juga mau minta buatkan |
| Abng dh nk balik k, Mun balik bawak jk dolok | abang sudah mau pulang ke, kalau pulang bawa saja dulu | abang sudah mau pulang ke, kalau pulang bawa saja dulu |
| Sian sih die madahkan tabel atau ndak bng, Cuman di suruhnye buat ringkasan jak | tidak ada sih dia bilang tabel atau tidak bang, cuman di disuruhnya buat ringkasan saja | tidak ada sih dia bilang tabel atau tidak bang, cuman di perintahnya buat ringkasan saja |
| Ok wak mseh d Mega mall, Drek dah nak makan e? | ok wak masih di mega mall, kamu sudah mau makan ya? | ok wak masih di mega mall, kamu sudah mau makan ya? |