

# Perbandingan Translation Library Pada Python (Studi Kasus: Analisis Sentimen Penyakit Menular Di Indonesia)

Ni Kadek Ratna Sari<sup>a1</sup>, I Made Agus Dwi Suarjaya<sup>a2</sup>, Putu Wira Buana<sup>a3</sup>

<sup>a</sup>Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana, Bali

e-mail: [1kdratnasari@student.unud.ac.id](mailto:1kdratnasari@student.unud.ac.id), [2agussuarjaya@it.unud.ac.id](mailto:2agussuarjaya@it.unud.ac.id), [3wbhuana@gmail.com](mailto:3wbhuana@gmail.com)

## Abstrak

Era saat ini pengelolaan data dalam jumlah besar dan kompleks ini perlu adanya penguraian data dengan fitur terjemahan bahasa agar lebih efisien. Penggunaan machine translation dapat menerjemahkan data secara otomatis dari satu bahasa ke bahasa lainnya. Pendekatan machine translation yang diusulkan yaitu penggunaan library pada Python seperti deep-translator, textblob, dan goslate. Pada penelitian analisis sentimen penyakit menular terdapat tahapan pengolahan analisa yang di dalamnya terdapat proses translate data tweet. Hasilnya mengungkapkan bahwa ketiga library yang dibandingkan dapat digunakan untuk melakukan translation data, tetapi dapat dipilih dan disesuaikan dengan keperluan. Penggunaan library yang paling baik dihasilkan library deep-translator yang memiliki access free dengan unlimited translation dan support features yang lengkap serta dapat mendeteksi kata typo lalu diterjemahkan merujuk pada arti yang sesungguhnya, maka pada studi kasus analisis sentimen penyakit menular menggunakan library deep-translator untuk mengolah jumlah data tweet yang didapat saat crawling data.

**Kata kunci:** Mesin Penerjemah, Deep-Translator, Textblob, Goslate

## Abstract

In this era, when managing large and complex data, it is necessary to parse data with language translation features to make it more efficient. The use of machine translation can translate data automatically from one language to another. The proposed machine translation approach is the use of libraries in Python as follows deep-translator, textblob, and goslate. In infectious disease analysis research, there are stages of analysis in which there is a process of translating tweet data. The results reveal that the third library compared can be used to perform data translation, but can be selected and adjusted according to need. The best use of the library is the deep-translator library which has free access with unlimited translations and complete support features and can detect typos and refer to the true meaning. Thus, in this study of sentiment analysis of infectious diseases using the deep-translator library to process the amount of tweet data obtained during data crawling.

**Keywords :** Machine Translation, Deep-Translator, Textblob, Goslate

## 1. Introduction

Penyakit menular merupakan penyakit yang disebabkan oleh mikroorganisme seperti virus, bakteri, *parasite* dan jamur yang dapat menyebar secara langsung maupun tidak langsung, dari satu orang ke orang lainnya (WHO). Penyakit menular masih menjadi masalah kesehatan masyarakat yang penting di Indonesia, dimana potensi, permasalahan, pencegahan dan pengendalian penyakit menjadi input dalam menentukan arah kebijakan. Indonesia memiliki tingkat kerawanan yang tinggi terhadap potensi dari penyakit menular seperti HIV, *tuberculosis*, malaria dan demam berdarah.

Perkembangan di era saat ini mengalami pertumbuhan yang sangat pesat, menurut Asosiasi Penyelenggara Jasa Internet Indonesia (2018) penggunaan internet mencapai 64,8%, dengan persentase 18,9% dalam mengakses sosial media, salah satu penggunaan sosial media yang sangat digemari saat ini Twitter. Pengguna Twitter dapat untuk saling membagikan

---

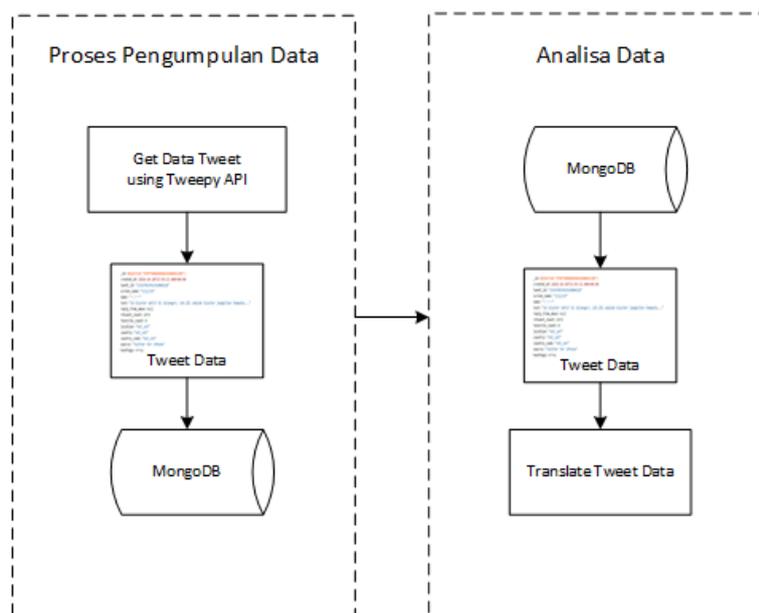
*tweet* yang dapat saling bertukar informasi, mencurahkan isi hati sehingga melahirkan sekumpulan opini. Pengelolaan data dengan jumlah besar ini tidak dapat dikerjakan secara manual karena membutuhkan waktu yang sangat lama dan tidak efisien, maka diperlukan suatu sistem yang dapat membantu dalam pengerjaannya. Sistem yang diusulkan untuk pendekatan penguraian data terdiri dari beberapa fitur seperti terjemahan bahasa [1]. Pada studi kasus Analisis Sentimen Penyakit Menular di Indonesia terdapat beberapa tahapan, dimana pada tahapan pengolahan Analisa data terdapat proses *translate* yang digunakan untuk menyamaratakan data *tweet* yang diperoleh menjadi bahasa Inggris. Terjemahan bahasa dapat dilakukan dengan pemanfaatan *machine translation* yang sudah terlatih. *Machine translation* adalah sub-bidang linguistik komputasi yang mengacu pada terjemahan secara otomatis dari satu bahasa ke bahasa lainnya melalui sistem komputerisasi tanpa campur tangan manusia [2]. Sampai saat ini *statical machine translation* dan metode *neural machine translation* menjadi *machine translation* yang paling banyak di pelajari karena masih berkembang pesat sehingga pada abad ke-21 *machine learning* menjadi layanan online di internet yang digunakan oleh perusahaan besar seperti Google dan Microsoft [2]. Google menyediakan berbagai API (*Application Programming Interface*) yang sangat berguna bagi pengembang untuk memanfaatkan berbagai berbagai fitur salah satunya ialah *translation*. Terjemahan bahasa secara otomatis yang menggunakan Google *Translate* API yaitu beberapa *library* pada Python. *Library* pada Python merupakan gabungan dari *package* dan *module* untuk memudahkan dalam membuat suatu sistem atau aplikasi. Terdapat beberapa *library translation* pada Python yaitu *deep translator*, *textblob*, *goslate*, *googletrans*, *py-translate*, dll. *Library translator* sangat membantu dalam menerjemahkan teks, kalimat, frasa bahkan wacana dengan lebih efisien. Selain itu, *library translation* juga sering digunakan untuk beberapa penelitian seperti *sentiment analysis*, *machine translation*, *classification*, dll yang menggunakan data dalam jumlah besar menggunakan suatu bahasa tertentu yang ingin diterjemahkan menjadi bahasa yang diinginkan untuk keperluan penelitian contohnya data teks dari Bahasa Indonesia yang ingin diterjemahkan menjadi Bahasa Inggris.

Penelitian ini bertujuan untuk melihat perbandingan dari *library translation* yaitu *deep translator*, *textblob*, dan *goslate* pada Python. Penelitian ini berfokus pada terjemahan teks Bahasa Indonesia ke Bahasa Inggris. Diharapkan penelitian ini dapat membantu dalam pengambilan keputusan untuk menentukan *library* yang cocok digunakan untuk keperluan penelitian seperti pada studi kasus Analisis Sentimen Penyakit Menular di Indonesia pada tahapan pengolahan analisa data yaitu pada proses *translate*.

## 2. Research Method / Proposed Method

Metodologi penelitian menjelaskan mengenai tahapan penelitian dalam melakukan perbandingan *library translate*. Alur penelitian menjadi pedoman dasar dalam sebuah penelitian. Penelitian dilakukan dengan melalui beberapa tahapan. Tahapan-tahapan tersebut merupakan langkah detail pengerjaan penelitian agar menjadi lebih sistematis dan teratur. Berikut ini merupakan tahapan dari alur penelitian.

---



Gambar 1. Alur Penelitian

Gambar 1 merupakan tahapan dari alur penelitian perbandingan *library translate*. Tahapan awal yaitu proses pengumpulan data tweet menggunakan Tweepy API yang disimpan pada *database* MongoDB, selanjutnya data yang sudah tersimpan dipanggil kembali untuk melakukan proses *translate* data *tweet*. Terdapat tiga *library translation* yang dibandingkan yaitu *deep translator*, *textblob*, dan *goslate*.

### 3. Literature Study

Studi literatur merupakan metode pengumpulan data atau bahan referensi sebagai penunjang dalam penelitian perbandingan *library translaye* menggunakan Bahasa pemrograman python. Pustaka terkait penelitian akan dijelaskan sebagai berikut.

#### 3.1 Crawling Data

Crawling data *tweet* merupakan tahapan awal dari proses analisis data twitter penyakit menular di Indonesia. Tahapan ini dimulai dari pendaftaran untuk dapat mengakses twitter menggunakan API twitter sehingga mendapatkan kode token untuk hak akses mengunduh data pada twitter [3]. Proses *crawling data* menggunakan *keyword* penyakit menular di Indonesia seperti kata kunci dan *hashtag* yang berhubungan dengan penyakit menular pada twitter. Proses *crawling* data dilakukan dimulai dari bulan Maret 2020 sampai dengan Maret 2021.

#### 3.2 Library

Library pada Python merupakan gabungan dari package dan module untuk memudahkan dalam membuat suatu sistem atau aplikasi. Library merupakan sebutan kode program tambahan untuk keperluan tertentu, dengan penggunaan library ini dapat menghasilkan kode program secara efisien tanpa perlu menuliskan seluruh skrip. Library bersifat reusable yang dapat digunakan berkali – kali [4].

#### 3.3 Machine Translation

Machine translation adalah sub-bidang linguistik komputasi yang mengacu pada terjemahan secara otomatis dari satu bahasa ke bahasa lainnya melalui sistem komputerisasi tanpa campur tangan manusia [2].

##### 3.3.1 Rule-Based Machine Translation

*Rule-based machine translation* (RBMT) dapat melacak kesalahan *paraphrase* dan menangani permasalahan terkait urutan kata dengan pengetahuan linguistik. Sistem ini dibagi menjadi tiga metode yaitu metode *direct*, metode transfer, dan metode interlingua (IL). Direct method melakukan terjemahan kata per kata secara langsung. Metode transfer menghasilkan representasi abstrak dari kalimat dan yang dihasilkan menggunakan kamus dwi bahasa serta

aturan tata bahasa. Metode interlingua tidak perlu melewati langkah transfer karena representasi dari kalimat diasumsikan sama. Sistem ini tidak dapat menangani ambiguitas, maka setelah proses editing perlu disesuaikan dengan gaya penulisan dan target audiens [5].

### 3.3.2 Statistical Machine Translation

*Statistical machine translation* (SMT) merupakan mesin penerjemah yang penggunaan dengan machine learning [6]. Sistem ini merupakan penerjemah yang menggunakan pendekatan statistik dengan konsep probabilitas. Model pendekatan pada sistem ini dibagi menjadi tiga yaitu berbasis frase, berbasis sintaks, dan berbasis hirarki. SMT dapat menangani ambiguitas, dan hasil terjemahan kata yang dihasilkan lebih natural dibandingkan dengan RBMT [5]. Kelemahan SMT saat menerjemahkan suatu kata yang tidak sesuai dengan *corpus* pelatihan sehingga menghasilkan akurasi yang buruk. *Corpus* perlu disesuaikan dengan gaya tertentu agar dapat menghasilkan akurasi yang baik [7].

### 3.3.3 Hybrid Machine Translation

*Hybrid machine translation* (HMT) merupakan kombinasi dari *multiple machine translation*. Pengembangan hybrid ini berasal dari kegagalan teknik tunggal yang tidak mencapai akurasi yang baik. HMT ini juga sering kali dikaitkan dengan pendekatan *statistic* dan *rule-based* [5].

### 3.3.4 Neural Machine Translation

*Neural machine translation* merupakan mesin yang menggunakan teknologi jaringan buatan yang besar untuk memprediksi urutan sebuah kata dalam model yang sudah terintegrasi. Struktur model sistem ini lebih sederhana dibandingkan dengan sistem model berbasis frase [5]. NMT dapat memahami kesamaan antar kata, mempertimbangkan seluruh kalimat yang ada, dan dapat mempelajari hubungan kompleks antar bahasa [7].

## 3.4 Deep\_translator

Deep\_translator merupakan *library* python gratis dan tidak terbatas dalam penggunaannya yang fleksibel menerjemahkan antar bahasa dengan sederhana serta memberikan dukungan untuk semua bahasa. *Library* ini dapat menerjemahkan kata dari file, hasil terjemahan dari banyak sumber, dapat mendeteksi bahasa secara otomatis. *Library* ini memiliki abstraksi tingkat tinggi yang mendapat dukungan multi bahasa sehingga mudah digunakan serta APInya sangat mudah digunakan karena dipelihara secara teratur dan stabil. Fitur yang mendukung dalam penggunaan *library* ini meliputi *google translate*, *Microsoft translator*, *pons translator*, *linguee translator*, *Yandex translator*, *deepl translator*, *proxy usage*, *get multiple translation*, dll [8].

## 3.5 Textblob

Textblob merupakan *library* Python untuk memproses data teks. *Library* ini menyediakan API sederhana untuk Natural Language Processing (NLP) seperti *sentiment analysis*, *classification*, *translation*, dll. Fitur yang mendukung dalam penggunaan *library* ini meliputi *parsing*, *tokenization*, *n-grams*, *wordnet integration*, dll [9].

## 3.6 Goslate

Goslate merupakan *library* Python yang memberikan layanan terjemahan API ke google secara gratis ke situs web terjemahan google [10]. Goslate menggunakan *batch* dan pengambilan bersamaan secara agresif untuk mencapai kecepatan terjemahan yang dimaksimalkan secara internal [11].

## 4 Hasil dan Pembahasan

Hasil dan pembahasan merupakan laporan dokumentasi dari hasil dan pembahasan penelitian dilakukan. Berikut ini merupakan pemaparan dari hasil dan pembahasan.

### 4.3 Project Description

Penelitian ini membandingkan tiga *library translation* pada Python. Setiap *library* memiliki deskripsi yang menjelaskan dan menggambarkan detail dari penggunaan *library*. Deskripsi dari masing – masing *library* dipaparkan pada Tabel 1.

---

Tabel 1. Project Description

Library	Access	Translation Limits	Features	Programming Language
Deep-translator	Free	Unlimited	Google translate, Microsoft translator, pons translator, linguee translator, Yandex translator, deepl translator, mymemory translator, Papago translator proxy usage, automatic single language detection, batch detection, translate from text file, multiple translate, translation different paragraphs in different languages, translate from terminal.	Python >=3.7, <4.0
Textblob	Free	Limited	Noun phrase extraction, part-of-speech tagging, sentiment analysis, classification, tokenization, word and phrase frequencies, parsing, <i>n</i> -grams, word inflection and lemmatization, spelling correction, add new models or languages through extensions, wordnet integration.	Python >= 2.7 or >= 3.5
Goslate	Free	Limited	<a href="#">Simple usage</a> , <a href="#">proxy support</a> , <a href="#">romanization</a> , <a href="#">language detection</a> , <a href="#">concurrent querying</a> , <a href="#">batch translation</a> .	Python 2 and Python 3

Tabel 1 menjelaskan bahwa *library* deep-translator memiliki keunggulan dari dua *library* lainnya, terlihat pada *access* yang *free*, *translation* yang *unlimited* dan memiliki *features* yang banyak dan lengkap untuk *support* dalam dilakukannya terjemahan suatu teks menggunakan *library* ini. Tetapi selain deep-translator, *library* textblob memiliki keunggulan tersendiri dimana *library* ini juga menyelami *Natural Language Processing* (NLP) dan bagian dari NLTK yang dapat melakukan tugas selain *translation*. Versi bahasa pemrograman Python yang dibutuhkan dari ketiga *library* ini memiliki variasi yang berbeda – beda, dapat disesuaikan dengan kebutuhan pengguna.

#### 4.4 Hasil Translate

Proses translate menggunakan data *tweet* yang didapat pada saat crawling data. Perbandingan hasil *translate* data *tweet* Bahasa Indonesia ke Bahasa Inggris dari tiga *library* yang digunakan dapat dilihat pada Tabel dibawah ini.

Tabel 2. Hasil Translate Menggunakan *Library* Deep-Translator

No	Data Tweet	Output
1	twips lapor teliti baru terbit tanggal 4 maret 2020 kemarin lapor asal singapura kelabu hasil tes sebab diagnosa pasien positif anggap sakit dbd demam darah dengue	twips meticulous report just published on march 4 2020 yesterday reported from singapore gray test results because the patients diagnosis is positive he thinks dengue fever is dengue
2	alhamdulillah doi sembuh corona	thank god he recovered from corona its

senank lihat happy

Tabel 2 menunjukkan hasil *translate* menggunakan *library* deep-translator dimana pada data *tweet* no 1 penulisan teks input benar menghasilkan terjemahan yang benar juga. Pada data *tweet* no 2 terdapat beberapa kata dalam Bahasa Indonesia yang penulisannya tidak sesuai atau *typo* seperti kata "senank" tetapi setelah dilakukan proses *translate* kata tersebut merujuk pada arti yang sebenarnya.

Tabel 3. Hasil Translate Menggunakan *Library* Textblob

No	Data Tweet	Output
1	twips lapor teliti baru terbit tanggal 4 maret 2020 kemarin lapor asal singapura kelabu hasil tes sebab diagnosa pasien positif anggap sakit dbd demam darah dengue	twips meticulous report just published on march 4 2020 yesterday reported from singapore gray test results because the patients diagnosis is positive he thinks dengue fever is dengue
2	alhamdulillah doi sembuh corona senank lihat	thank god he recovered from corona its happy

Tabel 3 menunjukkan hasil *translate* menggunakan *library* textblob dimana menunjukkan hasil yang sama dengan hasil *translate* menggunakan *library* deep-translator pada data *tweet* no 1 penulisan teks input benar menghasilkan terjemahan yang benar juga. Pada data *tweet* no 2 terdapat beberapa kata dalam Bahasa Indonesia yang penulisannya tidak sesuai atau *typo* seperti kata "senank" tetapi setelah dilakukan proses *translate* kata tersebut merujuk pada arti yang sebenarnya.

Tabel 4. Hasil Translate Menggunakan *Library* Goslate

No	Data Tweet	Output
1	twips lapor teliti baru terbit tanggal 4 maret 2020 kemarin lapor asal singapura kelabu hasil tes sebab diagnosa pasien positif anggap sakit dbd demam darah dengue	twips report themat just published on march 4 2020 yesterday reported singapore gray test results because the diagnosis of the patient was positively considered the pain of dbd dengue blood fever
2	alhamdulillah doi sembuh corona senank lihat	alhamdulillah doi recovered corona senank look

Tabel 4 menunjukkan hasil *translate* menggunakan *library* goslate dimana pada data *tweet* no 1 penulisan teks input benar menghasilkan terjemahan yang benar juga, sedangkan pada data *tweet* no 2 terdapat beberapa kata dalam Bahasa Indonesia yang penulisannya tidak sesuai atau *typo* seperti kata "senank" tetap menjadi kata tersebut dengan artian penggunaan *library* ini tidak bisa mendeteksi kata yang kurang tepat.

## 5 Conclusion

Pada jurnal penelitian ini dinyatakan bahwa penggunaan *machine translation* sangat membantu dalam tugas pengelolaan *translation* dalam jumlah besar. Penggunaan *library* deep-translator pada Python memiliki keunggulan yang lebih dibandingkan dengan *library* textblob dan *library* goslate. Kelebihan dari penggunaan *library* deep-translator yaitu *access* penggunaan yang *free*, *unlimited translation*, *support features* yang lengkap, terfokus pada fungsi *translation* dan dapat mendeteksi kata yang *typo* merujuk pada arti yang sesungguhnya. Hasil ini dapat menjadikan *library* deep-translator adalah solusi terbaik untuk *translation* pada studi kasus Analisis Sentimen Penyakit Menular di Indonesia pada tahapan pengolahan analisa data.

## References

- [1] K. Alnavar, R. U. Kumar, and C. N. Babu, "Document Parsing Tool for Language Translation and Web Crawling using Django REST Framework," *Journal of Physics: Conference Series*, 2021.

- [2] X. Zhang, "An Improved English to Chinese Translation Search Engine of Technical Text," 2017.
  - [3] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," Sep. 2016, pp. 11–16. doi: 10.21108/indosc.2016.111.
  - [4] S. Miftah, "Library Python Kenali Perbedaan Module, Package, dan Library Pada Python," *DQLab*, 2021.
  - [5] A. A. Septarina, F. Rahutomo, and M. Sarosa, "COMMUNICATIONS IN SCIENCE AND TECHNOLOGY Machine translation of Indonesian: A review," 2019.
  - [6] A. Esan *et al.*, "Development of a recurrent neural network model for english to yoruba machine translation," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 602–609, 2020, doi: 10.14569/IJACSA.2020.0110574.
  - [7] G. Dino, "3 Reasons Why Neural Machine Translation is a Breakthrough," <https://slator.com/3-reasons-why-neural-machine-translation-is-a-breakthrough/>, 2017.
  - [8] N. Baccouri, "Deep-translator," <https://pypi.org/project/deep-translator>, 2020.
  - [9] S. Loria, "Textblob," <https://pypi.org/project/textblob>, 2021.
  - [10] A. R. Calingo, A. M. Sison, and B. T. Tanguilig, "Prediction Model of the Stock Market Index Using Twitter Sentiment Analysis," *I.J. Information Technology and Computer Science*, vol. 10, pp. 11–21, 2016.
  - [11] Z. Qiang, "Goslate 1.5.2," <https://pypi.org/project/goslate/>.
-