# PDF Highlighted Text Translator

This repository contains Python code that demonstrates the extraction of highlighted text from a PDF file and translates it into Hinglish (a mixture of Hindi and English). The translation is performed using two different methods - One involving a transformer model and the other using Google Translate via the “googletrans” library.

1. **Requirements**

* Python 3.x
* Packages: fitz, nltk, googletrans, transformers

1. **Installation**

* Clone this repository:

“ ”

* Install the required packages:
* pip install -r requirements.txt

1. **Usage**

**Running the Code**

* Ensure the PDF file path is specified correctly in the main block of translation.

pdf\_path = 'path/to/your/PDF/file.pdf' # Replace with your PDF path 4

* Run the code:

translatior.py

1. **Important Notes**

* This code utilizes fitz for PDF processing, nltk for natural language processing, and two translation methods: a transformer model and Google Translate.
* In case of network issues with nltk, ensure the code is run in a different network as there might be restrictions due to network configurations.
* The code performs the following actions:
* Extracts highlighted text from the provided PDF.
* Translates the extracted text into Hinglish using two methods:
* Using Google Translate via the googletrans library for translation.
* Utilizing transformer models (e.g. “findnitai/t5-hinglish-translator” and “SkAndMl/english-to-hinglish”) for translation.

1. **Further Details**

The Translation class within the code performs the following key functions:

1. **extract\_highlighted\_text():** Extracts highlighted text from the provided PDF file.
2. **hindi\_translator (text):** Translates English text to Hindi using the Google Translate API.
3. **Translator (hindi, english):** Translates English nouns to Hindi and merges them to create Hinglish text.
4. **perform\_translation (highlighted\_text):** Uses transformer models to generate translations for the extracted highlighted text.
5. **Troubleshooting**

If you encounter issues while running the code, such as nltk not functioning due to network restrictions, it's advisable to run the code in a different network environment.