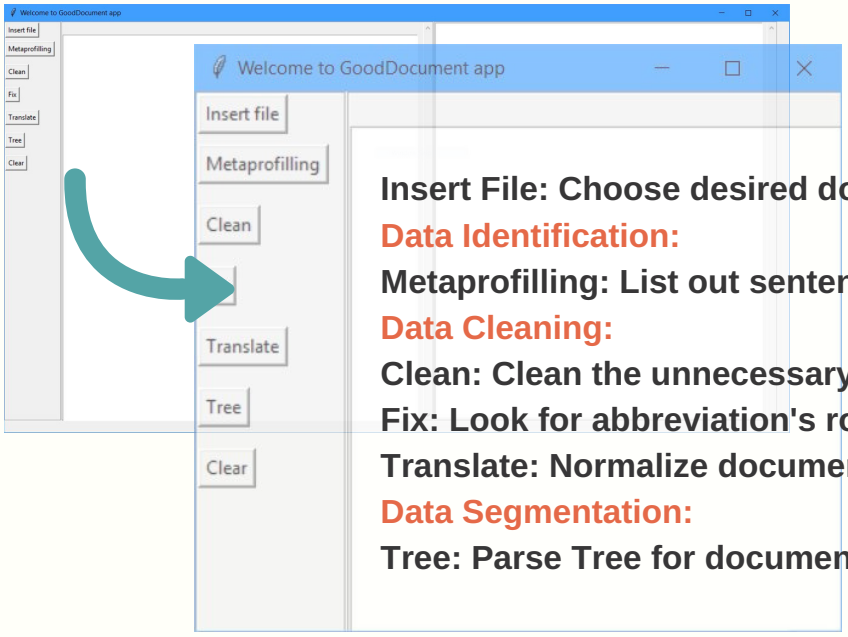
NLP ENGINE PROTOTYPE

# GOODDOCUMENT

Gooddocument is a basic NLP engine basic prototype which is able to process real-world Social Media (Socmed) data in order to further study Z generation who are always on social media:

1. Normalization (BM -> Eng)
2. Syntax Classifier (Parsing)

Tools involved:
Python 3
Open-Source Libraries:
NLTK
Numpy
Pandas
n-gram
tkinter




**Insert File: Choose desired documents to analyse**

**Data Identification:**

**Metaprofilling: List out sentences by same user**

**Data Cleaning:**

**Clean: Clean the unnecessary text like reply, emoticon**

**Fix: Look for abbreviation's root word**

**Translate: Normalize document into English language**

**Data Segmentation:**

**Tree: Parse Tree for document**

**Cleaning**

**Normalization**

1. Look for abbreviation's original words
2. Profiling to identify the user's sentences
3. Syntax Parsing (From English)

UNIVERSITY OF MALAYA
FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY


Reference:
Chekima, K., & Alfred, R. (2018). Sentiment Analysis of Malay Social Media Text. (R. Alfred, H. Iida, A. A. Ag. Ibrahim, & Y. Lim, Eds.) (Vol. 488). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-8276-4
https://hub.packtpub.com/clean-social-media-data-analysis-python/
https://www.kdnuggets.com/2018/03/text-data-preprocessing-walkthrough-python.html



YU KANG | AZWAN | ALAN