

Readme

Author: Michael Peter Hoffmann, Fachbereich Mathematik und Informatik, FU Berlin.

Introduction

This brief document describes how to upload the datasets and run the code for both the transformer-based models and the SVM-based models.

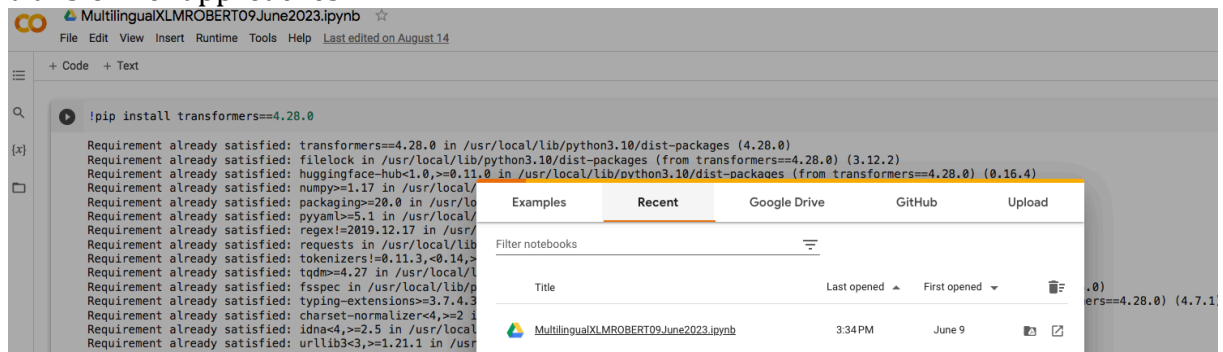
Installation

1. The datasets

- open up a google account (if you do not have it)
- copy all nine files from the folder „Datasets“ into your google drive account:
 1. the original german dataset: germanOriginal.txt
 2. the deepl translation of the german dataset: germanDeepLTrans.txt
 3. the google translation of the german dataset: germanGoogleTrans.txt
 4. the original italian dataset: italianOriginal.tsv
 5. the deepl translation of the italian dataset: italianDeepLTrans.txt
 6. the google translation of the italian dataset: italianGoogleTrans.txt
 7. the original spanish dataset: spanishOriginal.txt
 8. the deepl translation of the spanish dataset: spanishDeepLTrans.txt
 9. the google translation of the spanish dataset: spanishGoogleTrans.txt

2. The Transformer Scripts:

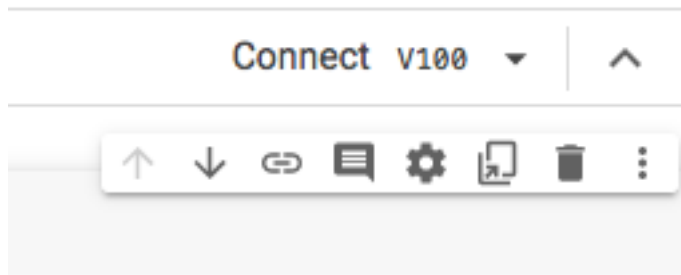
- open google colab
- open each of the seven files for monolingual, multilingual and translation based transformer approaches



- Each of the seven scripts has the BERT model + the respective SVM model in the code

Running the Scripts:

1. Before running the transformer scripts, please ensure to either have a A100 or V100 accelerator enabled (Note: You probably need a suscription to google colab)
2. Example: V100 enabled is displayed in upper right corner:



3.