Computational Linguistics 1

Assignment 2

Due: 26 July 2021, 11:55 PM

Question 1

Consider 20 sentences in your native language, identify the named entities in the sentences and save it in .txt file. You will be using both BIO (https://en.wikipedia.org/wiki/Inside%E2%80%93outside%E2%80%93beginning_(tagging)) and NER (PER, LOC, etc) tagging.

Note: There should be ample examples of named entities in the sentences you choose.

Question 2

Google translate the 20 native sentences and use any online available NER tool (like Stanford NER tool) and save your observations in .txt file.

Question 3

Manually translate the 20 native sentences to English and then use online NER tool and save your observations in .txt file.

4 Submission Format

PLEASE FOLLOW THE SUBMISSION FORMAT STRICTLY.

Please do NOT handwrite this assignment.

Submit just ONE zip file, named <your-roll-number>.zip,

which has the following file structure

your-roll-number

|- IndianLanguageNER.txt

|- GoogletranslateNER.txt

|- EnglishLanguageNER.txt

|- Report.pdf

Sample output file is attached. The report should include all your observations.