# CS689A COMPUTATIONAL LINGUISTICS FOR INDIAN LANGUAGES - Assignment 02

**Student Name : Saqeeb**
**Roll Number : 22111053**

---

## Installation guide

This assignment code requires python library like **conllu, collections, itertools**. Install these libraries and run the code. To install conllu library see the command below.

```
!pip install conllu
```

We have the conllu file which has many sentences. In a particular sentence S consisting of n words w1 . . . wn. Corresponding to that are the POS tags t1 . . . tn. For each word wi, assume the gender, case and number to be gi, ci and bi respectively. There are also n dependency relations in the sentence (including the one with the root). Assume a dependency relation to be ⟨wi,R,wj⟩ which indicates that the word wi is connected to its head word wj by the relation R. Correspondingly, the POS tag relation is ⟨ti,R,tj⟩.

**Solution 1a**

- Find the frequencies of POS tags of words.
- I have iterated over each word of each sentence to get the POS tag of words

**Solution 1b**

- Listed the 50 most frequent words corresponding to each POS tag
- For each POS tag I have found the 50 most frequent words for that POS tag

**Solution 1c**

- Find the frequencies of gender, case and number of words separately.
- Here I have iterated over each word of each sentence and found the gender, case and number.

**Solution 1d**

- Listed the 50 most frequent combinations of gender, case and number as a 3-tuple.
- Here I have found all the possible combination of the 3-tuple.

**Solution 1e**

- Found the frequencies of POS tags corresponding to only head words.
- Here I have found the POS tags of all the head words in the sentence.

**Solution 1f**

- Found the directed POS tag tuples, i.e., ⟨ti , tj ⟩. For each such 2-tuple, listed the frequencies separately for each relation R as well as total.
- Here I have created a dictionary of all the directed POS tag tuple and it's relation R.

**Solution 1g**

- For each dependency relation R, list the frequencies separately for each 2-tuple ⟨ti,tj⟩ as well as total.
- Here I have found list of all POS tag tuples ⟨ti,tj⟩ for each realtionship R

**Solution 2**

- Fine-tune the pre-trained BERT model for the UPOS prediction task.
- Parameters used for training :

    - batch_size = 20
    - learning_rate = 10e-5,
    - number of epochs = 100
    - weight_decay = 0.01
- Final Results :
    - 'eval_loss': 0.5430908799171448
    - 'eval_precision': 0.9448654563138272
    - 'eval_recall': 0.945095257397649
    - 'eval_f1': 0.9449803428849349
    - 'eval_accuracy': 0.9482500680025387