Linguistic Data 2: Collection & Modeling

Assignment 4

8 April, 2020

Each one of the course participants have been allotted the type of tagset they will be working with. We will be working with only Hindi in this assignment. If you feel that data for your tagset is not sufficient for training your model, then convert the data of the other tagset into your tagset using the scripts you wrote in Assignment 3.

The task is as follows:

- Build and train a statistical/neural model to tag data on the allocated tagset.

- You have to prepare a report for this assignment. Describe your model and also an error analysis on the output of the model on the test data.

- The model you build and train has to be saved and uploaded on your IIIT OneDrive. Furthermore, in your submission include

    1. A requirements.txt (read here if you do not know about this)
    2. A test file for your model.
    3. The URL for your saved model.
    4. A script to test your model. The input to that script will be the location of the test file and the location of the model. The output should be written to a file named *RollNo_output.txt*.

Post any and all doubts on the Moodle thread for this assignment.

DEADLINE: 21 April 2020, 11:55 PM