**NLP Unit Evaluation**

**17-Oct-2015 3:30 pm to 6:30 pm**

**Objective:**

In this lab we would develop two language models using Bidirectional RNN and compare the perplexity values of same machine with the Language model which is developed in Lab 3 Evaluation.

**Steps:**

1. Take the Wikipedia dataset which is already developed in Lab 3 Evaluation.

2. Rank order various words as per the unigram

3. Decide some threshold value and if unigram is >= threshold, then add the word to the Vocabulary else reject the word as unknown word.

Restrict the vocabulary size upto 1000. (1000 unique words)

4. You can use 80% corpus for training and 20% corpus for testing.

5. Implement the Bidirectional RNN as per lecture slides.

5. Evaluate the new Language Model using perplexity and Compare it with the earlier model which was developed using RNN.

**Deliverables:**

1. Demo of your work by 60:20 pm, 9-Nov-2015.
2. Submit the following by Tuesday 10-Nov-2015 6 pm:

1. Source code of Language models.

2. Post your analysis on the Facebook