

Department of Computer Science and Engineering

Course: Natural Language Processing [Elective - III]

Assignment - 1

Max. Marks: 02

Course Outcome: CO2

VIII Semester - Shift I and Shift II

Design Language Model for English Language.

1. Select Corpus Size : K
 2. Select size of Vocabulary : L
 3. Design any 10 sentences of your choice.
 4. Draw count/confusion matrix for one of the sentence newly constructed using the vocabulary of the corpus.
 5. Draw Probability matrix using confusion matrix.
 6. Perform Add-1 Smoothing and comment on high, low, marginal reduction in probability values. Draw the probability matrix
 7. Perform Witten - Bell smoothing on selected words of corpus.
 8. Compute Perplexity of Language model.
 9. Compute Entropy of Language model.
 10. Your conclusion on language model.
-

Group Size : 02 students per group

Every group should design unique 10 sentence related to fixed domain.

Last date of submission: 03 Feb 2017

M.B.Chandak

HoD CSE