Time: 3 Hours **Total Marks: 80**

N.B: 1) Question **number 1** is compulsory.
2) Attempt **any three** out of the remaining questions.
3) Assume suitable data if **necessary** and justify the assumptions.
4) Figures to the **right** indicate full marks.

1. Attempt any **FOUR.**

a) Define Natural Language Processing. List down various applications of NLP? [05]

b) What are ambiguities in the context of NLP? Why are they significant in NLP design? [05]

c) List all bi-grams and compute first 5 bigram probabilities for the sentence given below: [05]
<s> the students pass the test<\s>
<s> the students wait for the pass<\s>
<s> teachers test students<\s>
Explain how spelling correction is done using N-grams

d) Using the following CFG, Parse the sentence "**Book the flight**" [05]
S -> NP | VP
S -> VP
NP -> Det | N
NP-> Det |Adj | N
VP -> V
VP -> V | NP

e) The following POS tagged Corpus is given [05]
<s> the/DT students/NN pass/V the/DT test/NN<\s>
<s> the/DT students/NN wait/V for/P the/DT pass/NN<\s>
<s> teachers/N test/V students/NN<\s>
Write down the transition and emission probabilities for a bigram HMM

2. a) Assume that we modify the cost incurred for operations in calculating Levenshtein distance such that both the insertion and deletion operations occur a cost of one each while substitution incurs a cost of 2. [10]

Calculate the distance between string *reading* and *writing*.

b) What is morphological analysis in NLP? Compare and contrast derivational morphology with inflectional morphology? [10]

3. a) Explain the n-gram language model. Given the following corpus [10]

<S> I am Henry </S>
<S> I like college </S>
<S> Do Henry like college </S>
<S> Henry I am </S>
<S> Do I like Henry </S>
<S> Do I like college </S>
<S> I do like Henry </S>

1) Find which statement is more probable using a bi-gram model? Apply **Laplace smoothing** if required.

i) I like college

ii) do I like Henry

2) Predict the next probable word for "**I like.**."

b) What are tokenization, lemmatization, and stemming in NLP? Compare and contrast lemmatization and stemming. [10]

4. a) What is POS tagging? Explain how the Viterbi algorithm is used for POS tagging. [10]

b) Define semantic analysis and lexical semantics in NLP. How are ambiguities related to lexical semantics? Explain its importance and provide examples. [10]

5. a) While working with context extraction from a text data, you encountered two different sentences: The tank is full of soldiers. The tank is full of nitrogen. Which methods can be used to remove word sense disambiguation in sentences? Discuss any one method of WSD in detail. [10]

b) What are syntactic and semantic constraints on coreference? Compare and contrast between them. [10]

6. Attempt any FOUR

a) What is sentiment analysis? Explain in detail the methods used for it. [05]

b) Illustrate preprocessing operations/steps in NLP: segmentation , Tokenization, stop word removal, script validation, filtration. [05]

c) What are collocations? What is their significance in NLP? [05]

d) What is WordNet? Explain the terms 'gloss' and 'synset' in the context of WordNet. [05]

e) What are referring expressions in NLP? Explain how pronouns act as referring expressions, with examples. [05]

****