

Enrolment No: E22CSEJ0827 Name of Student: MADHAV GUPTADepartment/ School: SCSET**MID-TERM EXAMINATION, ODD SEMESTER 2024****COURSE CODE:** CSET346**MAX. DURATION:** 1 HRS**COURSE NAME:** Natural Language Processing**PROGRAM:** B.Tech**TOTAL MARKS:** 20

Mapping of Questions to Course Outcomes				
Q.No.	A1	A2	A3	A4
CO	1,2	2,3	1,2,3	1,2
BTL	L3	L3	L2	L1

GENERAL INSTRUCTIONS: -

1. Do not write anything on the question paper except **name, enrolment number** and **department/school**.
2. Carrying mobile phones, smartwatches and any other non-permissible materials in the examination hall is an act of **UFM**.

COURSE INSTRUCTIONS:

- a) All the answers should be written with the proper index number.
- b) If any diagram is incorporated, it should be neat and clear.
- c) All questions are compulsory to answer.

SECTION A**Max Marks: 20**

A1) Explain N grams language modelling.

(2+3= 5 Marks)

Consider the following corpus

<S> I love to play</S>

<S> I love to read</S>

<S> I like to play</S>

<S> I like to read</S>

<S> I I like love to play </S>

Assume that we use a bigram language model based on above data. What does the most probable next word predict by the model of the following data?

1. <S>I li/lo
2. <S>I like to
3. <S>I like to _____
4. <S>I love to _____

A2) Describe the Term Frequency (TF) and Inverse Document Frequency (IDF) and write the formula to calculate TF and IDF. (2+3= 5 Marks)

Suppose we have the following three documents:

Document 1: "The cat in the hat"

Document 2: "The cat sat on the mat"

Document 3: "The dog in the fog"

Query Words: The cat

1. Calculate Term Frequency (TF) of the Query Words.
2. Calculate Inverse Document Frequency (IDF) of the Query Words.
3. Calculate TF-IDF score of the Query words
4. Rank the documents based on the average TF-IDF scores.

A3) Describe different phases of NLP? Write the different ambiguities encountered in NLP with proper examples? (3+2= 5 Marks)

A4) a) Explain the Coreference Resolution in NLP with proper examples? (2.5+2.5= 5 Marks)

b) Write the difference between stemming and lemmatization with proper example?