

Instructions to the students

1. The students have to write the following in the answer sheet

**SASTRA DEEMED TO BE UNIVERSITY
SCHOOL OF COMPUTING
END SEMESTER EXAMINATION - May 2025**

B. Tech - VI-SEMESTER

Name:

Register number:

Subject code: CSE405R02

Subject name: NATURAL LANGUAGE PROCESSING

Date: 07-05-2025

Time: 11:30 AM-2:30 PM

2. Two questions are provided, each question carries 25 marks

S.No.	Q. No.	Description	Max Marks	Marks Given
1	Question 1	Aim & Procedure	10	
2		Program	10	
3		Output	5	
4	Question 2	Aim & Procedure	10	
5		Program	10	
6		Output	5	
7		Total	50	

Examiner 1

Examiner 2

3. Students have to use the dataset attached
4. Students have to verify the required packages within the first 15 minutes
5. Finally, upload the answers without fail
6. No additional sheets will be provided

SASTRA Deemed University
Department of Computer Science and Engineering
Natural Language Processing (CSE405R02)
End Semester Semi Theory Semi Lab Examination

07.07.2025

11:30AM-2:30PM

1.

- a. Read the file set2.txt and tokenize the text into individual words and remove all English stop words.**
- b. Apply stemming to the filtered tokens, print the final vocabulary (unique stemmed words) and report the total number of terms (after preprocessing).**
- c. Generate bigrams for all consecutive words from the stemmed tokens.**
- d. Use T-test and Chi-square test to check whether any of the bigrams form interesting collocations. Use a critical value of 1.962 for significance. List the top bigrams that qualify as collocations.**
- e. Result interpretation**
 - How many collocations were found?**
 - Mention two examples that passed both tests.**

2.

You are building an AI-powered email assistant that helps users complete their email responses. The assistant uses an LSTM model trained on a corpus of emails. Given a partial email (such as an email that requires a reply), complete the email with a contextually appropriate response.

Task:

Given a partial email, complete the email that fits the tone, context, and purpose.

Dataset: email_response_dataset.csv