



PRESIDENCY UNIVERSITY

BENGALURU

Roll No.														
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--

End - Term Examinations – May / June 2026

Date: 20- 05-2026

Time:01:00pm – 04:00pm

School: SOCSE	Program: IST, CSI		
Course Code: IST2001	Course Name: Fundamentals of Natural Language Processing		
Semester: VI	Max Marks: 100	Weightage: 50%	

CO - Levels	C01	C02	C03	C04
Marks	26	26	24	24

Instructions:

- (i) Read all questions carefully and answer accordingly.
- (ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks.

10Q x 2M=20M

1.	List the main stages involved in Natural Language Processing?	2 Marks	L1	C01
2.	Define TF, IDF.	2 Marks	L1	C01
3.	Illustrate Edit distance.	2 Marks	L2	C01
4.	List the activation functions used in a neural network.	2 Marks	L1	C02
5.	Compare Binary and Multinomial Logistic Regression.	2 Marks	L2	C02
6.	Define cross-lingual embeddings.	2 Marks	L1	C02
7.	Define Named Entity Recognition (NER).	2 Marks	L1	C03
8.	Define Emission Probability.	2 Marks	L1	C03
9.	Define Machine Translation.	2 Marks	L1	C04
10.	Define lexical resource in NLP?	2 Marks	L1	C04

Part B

Answer the Questions.

Total Marks 80M

11.	a.	Using the Histogram Intersection String Kernel (HISK), demonstrate the similarity between the two strings: String 1 = "banana" and String 2 = "bandana" Consider bigrams (k=2) as features for the kernel computation. Show all the steps of the calculation.	10 Marks	L2	C01
	b.	Explain the various stages of NLP.	10 Marks	L2	C01
Or					
12.	a.	Demonstrate the given problem using EDIT Distance for the two strings Intention and Extension with substitution cost of insert and delete is 1 and Replace is 2.	10 Marks	L2	C01
	b.	Define Text classification. Explain the text classification methods.	10 Marks	L2	C01
13.	a.	Consider a text classification problem where mails must be classified as Spam or Not Spam. The training dataset contains the following: - Spam - Send us your Password - Review Us - Not Spam - Password Review - Send us the Review Choose Naïve Bayes classifier and classify the new review: "Review us Now"	10 Marks	L3	C02
	b.	Explain the Neural Language Models with a neat diagram.	10 Marks	L2	C02
Or					
14.	a.	Consider the following four documents: - D1: "information retrieval is a core topic in data science" - D2: "data science includes machine learning and information retrieval" - D3: "deep learning advances machine intelligence" - D4: "information theory and data mining are important" - Construct TF-IDF score of the term "information" in D2. - Construct the TF-IDF score of the term "machine" in D2. - Compare the two scores and state which term is more important in D2 based on TF-IDF.	10 Marks	L3	C02
	b.	Explain the architecture and working principle of a Feed Forward Neural Network (FFNN). How does information flow from the input layer to the output layer?	10 Marks	L2	C02
15.	a.	Solve the intermediate probability and Final probability of the sequence <i>RGB</i> from the starting node to the previous node for all	10 Marks	L3	C03

previous states. Select the node with the best probability using Hidden markov model (using Table 1 and 2).

Table 1

	Urn1	Urn2	Urn3
Urn1	0.1	0.4	0.5
Urn2	0.6	0.2	0.2
Urn3	0.3	0.4	0.3

Table 2

	Urn1	Urn2	Urn3
Red	30	10	60
Green	50	40	10
Blue	20	50	30

- b.** Explain the phrase structure (constituency) rules for the sentence and construct parse tree: **"The boy eats an apple."**

10 Marks

L2

C03

Or

16.

- a.** Consider a doctor diagnoses fever by asking patients how they feel. The patients may only answer that they feel normal (n), dizzy (d), or cold (c)
- There are 2 states, "Healthy" (H) and "Fever" (F) but the doctor cannot observe them directly. They are hidden states.
- Every day, the patient can report to the doctor whether they are normal, or cold, or dizzy. Based on this (and the previous day's statement), the patient should be diagnosed as either H or F.
- Identify the diagnosis of a patient who reports **"Normal Cold Dizzy" across 3 days** with the Initial Probabilities: $H = 0.6$, $F = 0.4$ using Viterbi method and using Table 1 and 2 data.

10 Marks

L3

C03

Table 1

Transition	Healthy	Fever
Healthy	0.7	0.3

		<table><tr><td>Fever</td><td>0.4</td><td>0.6</td></tr></table> Table 2 <table><tr><th>Emissions</th><th>Healthy</th><th>Fever</th></tr><tr><td>Normal</td><td>0.5</td><td>0.1</td></tr><tr><td>Cold</td><td>0.4</td><td>0.3</td></tr><tr><td>Dizzy</td><td>0.1</td><td>0.6</td></tr></table>	Fever	0.4	0.6	Emissions	Healthy	Fever	Normal	0.5	0.1	Cold	0.4	0.3	Dizzy	0.1	0.6			
Fever	0.4	0.6																		
Emissions	Healthy	Fever																		
Normal	0.5	0.1																		
Cold	0.4	0.3																		
Dizzy	0.1	0.6																		
	b.	Explain the phrase structure (constituency) rules for the sentence and construct parse tree: “the cat sleeps quietly”	20 Marks	L2	C03															

17.	a.	Identify the BLEU score between the following pair of sentences. Assume n till 4-grams and assume equal weights for each precision. Candidate: Do not let them do not let them see Reference: Do not let them in do not let them see	10 Marks	L3	C04																																	
	b.	Define Machine Translation. Explain Machine Translation Paradigms in detail with examples.	10 Marks	L2	C04																																	
Or																																						
18.	a.	Define Machine Translation. Explain Machine Translation Paradigms in detail with examples.	10 Marks	L2	C04																																	
	b.	Calculate the 3 different Kappas by make use of 2 annotators, A1 and A2 annotate a set of 10 documents as either positive or negative, denoted as P and N in the table below. Their ratings are as follows: <table border="1"><tr><th>Annotator</th><th>D1</th><th>D2</th><th>D3</th><th>D4</th><th>D5</th><th>D6</th><th>D7</th><th>D8</th><th>D9</th><th>D10</th></tr><tr><td>A1</td><td>P</td><td>P</td><td>P</td><td>N</td><td>N</td><td>N</td><td>P</td><td>P</td><td>N</td><td>N</td></tr><tr><td>A2</td><td>N</td><td>P</td><td>N</td><td>P</td><td>N</td><td>P</td><td>P</td><td>P</td><td>N</td><td>P</td></tr></table> Calculate the observation and expectation matrices, and calculate the unweighted Kappa (UWK), linear weighted Kappa (LWK) and quadratic weighted Kappa (QWK) between the 2 annotators.	Annotator	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	A1	P	P	P	N	N	N	P	P	N	N	A2	N	P	N	P	N	P	P	P	N	P	10 Marks	L3	C04
Annotator	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10																												
A1	P	P	P	N	N	N	P	P	N	N																												
A2	N	P	N	P	N	P	P	P	N	P																												