Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – May / June 2026

Date: 20- 05-2026 **Time:**01:00pm – 04:00pm

School: SOCSE	**Program:** B. Tech	
Course Code: CAI2504	**Course Name:** Natural Language Processing	
Semester: VI	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	CO1	CO2	CO3	CO4	CO5
Marks	22	22	24	06	26

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.	What is Part-of-Speech (PoS) tagging?	**2 Marks**	**L1**	**CO1**
2.	Explain Bag-of-Words model.	**2 Marks**	**L2**	**CO2**
3.	Explain the concept of sequence labeling in NLP.	**2 Marks**	**L2**	**CO3**
4.	Explain the concept of sequence labeling in NLP.	**2 Marks**	**L2**	**CO3**
5.	Explain any 2 standard algorithms used for Named Entity Recognition.	**2 Marks**	**L2**	**CO4**
6.	Explain BLEU score calculation for machine translation evaluation.	**2 Marks**	**L2**	**CO4**
7.	Recall any 2 examples of Named Entities.	**2 Marks**	**L1**	**CO4**
8.	Define Constituency with respect to NLP.	**2 Marks**	**L1**	**CO5**
9.	Compare any 1 feature each of NER and PoS tagging.	**2 Marks**	**L2**	**CO5**
10.	Describe the attention mechanism in transformer models.	**2 Marks**	**L2**	**CO5**

Part B

Answer the Questions. **Total Marks 80M**

<table>
<tr><td>11.</td><td>a.</td><td>Illustrate dynamic programming approach to calculate the Minimum Edit Distance between
X = “algorithm” and Y = “altruistic”
Assume the cost of: Insertion = 1; Deletion = 1; Substitution = 2</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>Demonstrate string kernel similarity function on the following two strings
x = “machine” and y = “learning” to calculate:
a. HISK (x, y).
b. HISK normalised (x, y).</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>12.</td><td>a.</td><td>Compare various document representation techniques used for natural language processing.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>A chatbot system receives the user input:
"Book me a flight from Bangalore to Delhi tomorrow morning."
Demonstrate how the sentence is handled at each stage (Lexical Analysis, Syntax Analysis etc and explain the transformation of the input through these stages.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
</table>

<table>
<tr><td>13.</td><td>a.</td><td>Infer the values of TP, FP, FN, and TN from the given confusion matrix. Also, compute Accuracy, Precision, Recall and F1-score from the given confusion matrix:
<table>
<tr><td></td><td>Predicted Positive</td><td>Predicted Negative</td></tr>
<tr><td>Actual Positive</td><td>180</td><td>20</td></tr>
<tr><td>Actual Negative</td><td>30</td><td>270</td></tr>
</table></td><td>10 Marks</td><td>L2</td><td>CO2</td></tr>
<tr><td></td><td>b.</td><td>A word embedding model represents words as vectors in a 2D space. Given the following word vectors:
$\overrightarrow{Doctor} = (0.9,0.5), \overrightarrow{Hospital} = (0.6,0.3), \overrightarrow{School} = (0.4,0.7)$
Demonstrate vector arithmetic to compute the analogy
"Doctor" – "Hospital" + "School" and infer:
(i) Resultant vector (ii) Most appropriate analogous word.</td><td>10 Marks</td><td>L2</td><td>CO2</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>14.</td><td>a.</td><td>A Natural Language Processing model uses a context window of ±1 to learn word relationships. Relate this concept to the text: “Machine learning builds intelligent systems.”
(i) Show all possible (target, context) word pairs using the given context window.
(ii) Show the word–context co-occurrence matrix for the words: Machine, learning, builds, intelligent, systems.</td><td>10 Marks</td><td>L2</td><td>CO2</td></tr>
<tr><td></td><td>b.</td><td>A word–context matrix is given below representing the co-occurrence frequencies of words:
<table>
<tr><td></td><td>network</td><td>data</td><td>model</td></tr>
<tr><td>neural</td><td>2100</td><td>1800</td><td>1950</td></tr>
<tr><td>deep</td><td>2050</td><td>1750</td><td>1900</td></tr>
</table></td><td>10 Marks</td><td>L2</td><td>CO2</td></tr>
</table>

		<table><tr><td>random</td><td>120</td><td>60</td><td>80</td></tr></table> Demonstrate the concept of cosine similarity to compute the degree of similarity between the two vectors: (neural and network) and (data and network).			

15.	**a.**	Apply the Viterbi algorithm to determine the most probable hidden state sequence for the observation sequence O = {Normal, Cold} given the following parameters: Initial probabilities: H = 0.6, F = 0.4; H being Healthy and F being Fever. Emission Probabilities <table><tr><td>State</td><td>Normal</td><td>Cold</td></tr><tr><td>H</td><td>0.5</td><td>0.4</td></tr><tr><td>F</td><td>0.1</td><td>0.3</td></tr></table> Transition Probabilities <table><tr><td>From/To</td><td>H</td><td>F</td></tr><tr><td>H</td><td>0.7</td><td>0.3</td></tr><tr><td>F</td><td>0.4</td><td>0.6</td></tr></table>	**10 Marks**	**L3**	**CO3**
	b.	Given a sequence labeling problem with prior probabilities $P(T)$and likelihood probabilities $P(W \mid T)$, where W is the word sequence and T is the best tag sequence. Apply Bayes' theorem and derive the expression for most probable tag sequence for a given word sequence. Show how the maximization simplifies by ignoring $P(W)$, and determine the final predicted tag sequence.	**10 Marks**	**L3**	**CO3**
		Or			
16.	**a.**	Consider the sentence "I like a morning flight." Let the grammar have the following rules: • S -> NP VP • VP -> VB NP • NP -> DT NP • NP -> JJ NP • NP -> NN | PN and the following PoS tags for different words: I_PN like_VB a_DT morning_JJ flight_NN Apply Context Free Grammar and generate a parse tree that shows the complete derivation of the given text.	**10 Marks**	**L3**	**CO3**
	b.	Imagine you are building a simple search engine using the four documents (D1–D4): 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" A user searches for the word: "Machine" Apply the TF–IDF method to: 1. Calculate the TF–IDF score for "Machine" and "Intelligence" across the documents.	**10 Marks**	**L3**	**CO3**

		2. Determine which document (D1, D2, D3, or D4) achieves the highest total score for these two words.			

17.	**a.**	Two annotators classify 100 text samples into three categories: Positive (P), Neutral (N), Negative (Neg). <table><tr><th></th><th>P</th><th>N</th><th>Neg</th><th>Total</th></tr><tr><td>**P**</td><td>30</td><td>10</td><td>5</td><td>45</td></tr><tr><td>**N**</td><td>8</td><td>20</td><td>7</td><td>35</td></tr><tr><td>**Neg**</td><td>2</td><td>6</td><td>12</td><td>20</td></tr><tr><td>**Total**</td><td>40</td><td>36</td><td>24</td><td>100</td></tr></table>Analyze the data and compute: 1. Observed agreement (Po) 2. Expected agreement (Pe) 3. Cohen's Kappa 4. Interpret the result	**10 Marks**	**L4**	**CO5**
	b.	**Scenario:** You are tasked with building a translation system for a "low-resource" language (a language with very little digital text available, such as a rare regional dialect) to English. You have access to a small group of fluent bilingual linguists but very few books or translated documents. Compare the **Statistical-based MT** and **Rule-based MT** paradigms and Analyse which of these two would be the most effective choice for this specific scenario.	**10 Marks**	**L4**	**CO5**
		Or			
18.	**a.**	Assuming n till 4-grams and equal weights for each precision, analyze the candidate and reference text given below and calculate the BLEU score between the following pair of sentences. Candidate: Do not let them do not let them see Reference: Do not let them in do not let them see	**10 Marks**	**L4**	**CO5**
	b.	A social media company wants to build a real-time machine translation tool for "Gen-Z" internet slang and highly sarcastic comments (e.g., translating "That's fire" or "I'm dead" into another language without losing the intended meaning). Analyse which out of Rule-based MT and Neural MT (NMT) approach is better suited for capturing the nuance and context of ever-evolving social media language.	**10 Marks**	**L4**	**CO5**