Name:

Enrolment No:

UPES

End Semester Examination, May 2025

Course: Computational Linguistics and NLP **Semester: VI**

Program: B. Tech-CS-AIML **Time : 03 hrs.**

Course Code: CSEG 3048 **Max. Marks: 100**

Instructions: Calculator is allowed.

SECTION A
(5Qx4M=20Marks)

S. No.		**Marks**	**CO**
Q 1	Define stemming and lemmatization? How are they different?	**2+2**	**CO1**
Q 2	Explain the concepts of Abstractive and Extractive Summarization with examples.	**4**	**CO4**
Q 3	Compare and contrast Information Retrieval and Information Extraction based on goals, techniques, and output formats.	**4**	**CO4**
Q 4	Explain the key components of a multimedia system.	**4**	**CO3**
Q 5	Define the term frequency-inverse document frequency (tf-idf).	**4**	**CO2**

SECTION B
(4Qx10M= 40 Marks)

Q 6	Explain the steps of NLP and their significance with one of the NLP applications.	**10**	**CO1**
Q 7	Consider the following context-free grammar and input string for a shift-reduce parser task: Grammar: S → S+S S → S-S S → (S) S → a Input string: a1-(a2+a3)	**10**	**CO3**

	Demonstrate the step-by-step parsing process using a Shift-Reduce parser for the given input string. Provide the sequence of shifts and reductions, including the state of the parsing stack at each step. Make clear how the parser proceeds from one step to the next and whether it shifts or reduces, following the grammar rules.		
Q 7	Corpus consisting of three documents: **Doc 1:** "apple is red and sweet" **Doc 2:** "apple is delicious" **Doc 3:** "banana is yellow and sweet" 1. Convert Doc 1 and Doc 2 into term frequency vectors and calculate the cosine similarity between them. 2. Calculate the Jaccard similarity between the set of unique words in Doc 1 and Doc 3. 3. Calculate the TF-IDF score for the word "apple" in Doc 1.	**10**	**C02**
Q 8 Q 9	Describe the question answering system? Explain the types of questions with the architecture of QA system and compare it with traditional information retrieval systems. **OR** Explain the sentiment analysis. Describe its importance in Natural Language Processing. Discuss various techniques used in sentiment analysis and explain how sentiment polarity is determined. Mention at least two applications.	**10**	**C04**
SECTION-C (2Qx20M=40 Marks)			
Q 10	You have a collection of 5 documents (D1, D2, D3, D4, and D5) and two user queries (Q1 and Q2). Using the Boolean Information Retrieval (IR) model, your task is to retrieve the most relevant documents for each query. Documents (D1, D2, D3, D4, D5): D1: "Artificial intelligence is a rapidly growing field." D2: "Machine learning techniques are used in AI." D3: "Natural language processing is a subfield of AI." D4: "Deep learning is a subset of machine learning." D5: "AI applications are transforming various industries." Queries (Q1 and Q2): Q1: "What are the applications of artificial intelligence in healthcare?" Q2: "Explain the role of deep learning in AI."	**20**	**CO4, CO3**

Q 11	**OR** Explain the Transformer architecture in NLP? describe its components such as self-attention, positional encoding, encoder-decoder structure, and multi-head attention with suitable diagrams. Discuss how transformers have revolutionized NLP with reference to models like BERT and GPT and mention their applications.		
Q 12	Describe the key challenges in building NLP systems for multilingual and low-resource languages. Discuss the techniques used for cross-lingual representation learning, including multilingual embeddings and transfer learning. Provide examples of cross-lingual data in any NLP applications.	**20**	**C02**