Name: **Enrolment No:**	

UPES

End Semester Examination, December 2024

Course: Natural Language Processing **Semester: III**

Program: MCA **Time : 03 hrs.**

Course Code: CSAI8004P **Max. Marks: 100**

SECTION A

S. No.		Marks	CO
Q 1	List Linguistic and Mathematical steps for building a Distributional Semantic Model (DSM).	**4**	**CO3**
Q 2	Describe context weighting when documents are used as contexts.	**4**	**CO3**
Q 3	Explain two different tools available for machine translation with their advantages and limitations.	**4**	**CO5**
Q 4	Explain how does question-answering model works.	**4**	**CO4**
Q 5	State the intuition behind Viterbi algorithm. Describe the three steps of Viterbi algorithm.	**(1+3=4)**	**CO2**
	SECTION B		
Q 6	Explain the five key components of Natural Language Processing (NLP). Discuss two roles of Natural Language Understanding (NLU) and Natural Language Generation (NLG) in NLP systems.	**(5+5=10)**	**CO1**
Q 7	Explain two importance of smoothing in language models. Discuss Good-Turing and Kneser-Ney Smoothing techniques with mathematical formulas.	**(5+5=10)**	**CO2**
Q 8	Explain the components of information retrieval system. List three differences between information retrieval and data retrieval.	**(7+3=10)**	**CO4**
Q 9	Explain different methods to word sense disambiguation (WSD). Discuss three applications of WSD. **(7+3)** OR Explain syntactic parsing. Describe context free grammar (CFG) and probabilistic context free grammar (PCFG) with example. **(3+7)**	**10**	**CO3**
	SECTION-C		
Q 10	Tag the phrase “a light book” using HMM tagger with Viterbi algorithm, where we have the following probabilities: P(the/DT) = 0.3, P(the/NN) = 0.1, P(light/NN) = 0.003,	**20**	**CO2**

	P(light/JJ) = 0.002, P(light/VB) = 0.06, P(book/NN) = 0.003, P(book/VB) = 0.01, P(VB/DT) = 0.00001, P(NN/DT) = 0.5, P(JJ/DT) = 0.3, P(NN/NN) = 0.2, P (JJ/NN) = 0.002, P(NN/JJ) = 0.2, P(NN/VB) = 0.3, P(VB/NN) = 0.3, P(VB/JJ) = 0.001, P(VB/VB) = 0.1 Assume all other conditional probabilities not mentioned to be zero. Also, assume that all tags have the same probability to appear in the beginning of a sentence.		
Q 11	Given the following grammar: S ⟶ NP VP VP ⟶ VBG NNS VP ⟶ VBZ VP NP ⟶ DT NN NP ⟶ JJ NNS DT ⟶ a NN ⟶ pilot VBZ ⟶ likes VBG ⟶ flying JJ ⟶ flying NNS ⟶ planes Use the CKY algorithm to parse the sentence "a pilot likes flying planes" OR (a) Differentiate Top-Down parsing and Bottom-Up parsing. **(5)** (b) Consider the following four sentences as a document corpus: "AI improves learning." "Learning improves intelligence." "AI fosters innovation." "Intelligence boosts innovation." Calculate: i. The Term Frequency (TF) for each word in each sentence. **(5)** ii. The Inverse Document Frequency (IDF) for each unique word in the corpus. **(5)** iii. The TF-IDF feature vector for each unique word across the sentences. **(5)**	**20**	**CO3**