U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

January / February 2025 Semester End Main Examinations

Programme: B.E. **Semester: V**

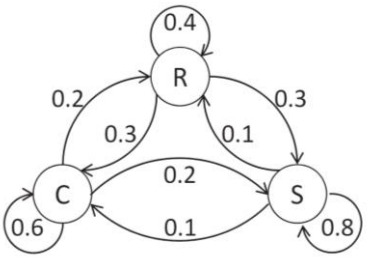
Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**

Course Code: 22AM5PENLP **Max Marks: 100**

Course: Natural Language Processing

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

		UNIT - I	*CO*	*PO*	Marks
1	a)	Find the probability of the test sentence P **(<s> students are from Vellore </s>)** in the following training set using Bi-gram model. Training set: <s> I am from Vellore </s> <s> I am a teacher </s> <s> students are good and are from various cities</s> <s> students from Vellore do engineering</s>.	*Co3*	*Po 3*	**10**
	b)	Summarize the smoothing techniques and evaluation metrics of language models.	*Co1*	*Po1*	**10**
		OR			
2	a)	Outline different phases of Natural Language Processing (NLP) with a neat diagram by giving suitable example at each phase.	*Co1*	*Po1*	**8**
	b)	Identify the type of ambiguity and all possible meanings arise from each of the sentences given. i. Time flies like an arrow. ii. Visiting relatives can be boring. iii. Are you still watching? iv. The teacher asked the student to bring his book to class. v. She observed the painting with a sharp eye.	*Co3*	*Po3*	**6**
	c)	Given a corpus with the following sentences: I like NLP I like machine learning Machine learning is fun i. Calculate the bigram probabilities. ii. Given the previous word "like", predict the next word using the bigram probabilities.	*Co3*	*Po3*	**6**

			UNIT - II			
	3	a)	Justify the role of regular expressions in NLP. Provide regular expressions for the following: i. Email address ii. Date in the format DD/MM/YYYY with a year between 1900 and 2099	*Co3*	*Po3*	**6**
		b)	Analyze whether "Campus" is closer to "Counter" by hand computing the minimum edit distance cost using the algorithm.	*Co3*	*Po3*	**6**
		c)	List and explain grammar rules with respect to English helps in constructing phrase structures.	*Co1*	*Po1*	**8**
			OR			
	4	a)	Once a day (e.g. at noon), the weather is observed as one of states 1: rainy(R) 2: cloudy(C) 3: sunny (S). Answer the following question based on the given state transition diagram:  Given that the weather on day 1 (t = 1) is sunny (state 3), what is the probability that the weather for the next 7 days will be "sun-sun – rain – rain – sun – cloudy - sun"?	*Co3*	*Po3*	**6**
		b)	Define Hidden Markov Model (HMM). Derive its components with assumptions.	*Co2*	*Po2*	**8**
		c)	List out atleast 5 different Part-of-speech tags in Penn Treebank. Identify part-of- speech to the word "back" in the sentences given below: i. A small building in the back ii. A clear majority of senators back the bill iv. Bob began to back toward the car v. Earnings growth took a back seat	*Co3*	*Po3*	**6**
			UNIT - III			
	5	a)	Construct the parse tree for the sentences/phrases using the grammar below. S -> NP VP ... PP -> P NP ... NP -> Det N | Det N PP | 'I' ... VP -> V NP | VP PP	*Co3*	*Po3*	**5**

			... Det -> 'an' | 'my' ... N -> 'elephant' | 'pajamas' ... V -> 'shot' ... P -> 'in' Check the given sentence "I shot an elephant in my pajamas" is ambiguous. Justify your answer.			
		b)	Explain the following with suitable examples: 1. Vectors and documents: term-document matrix 2. Words as vectors: document dimensions	*Co2*	*Po2*	**5**
		c)	With the help a diagram explain the working of transition based dependency parsing. Show how the word list "[book, me, the, morning, flight]" would be parsed using transition based dependency parsing.	*Co3*	*Po3*	**10**
			OR			
	6	a)	Parse the below sentences using dependency parsing: 1. I saw the queen England's hat. 2. The angry squirrel stole the mixed nuts from wallmart. 3. The cat eats tasty fish 4. Turn on the office lights.	*Co3*	*Po3*	**8**
		b)	Apply CYK parsing algorithm on the sentence "The Flight includes a meal" and the grammar for this sentence are as follows: - S-> NP VP, VP-> V NP, NP-> Det N, Det-> a, Det-> The, N->Flight, V -> includes, N -> meal.	*Co3*	*Po3*	**7**
		c)	Explain the following with examples: i. TF-IDF ii. Cosine similarity	*Co2*	*Po2*	**5**
			UNIT - IV			
	7	a)	With the help of a diagram explain key terms of co reference resolution.	*Co2*	*Po2*	**7**
		b)	Illustrate the following terms with an example i. Anaphora ii. Mention iii. Entity linking iv. Structured Polysemy v. Vector embeddings	*Co2*	*Po2*	**5**
		c)	Infer and explain the steps involved in unsupervised approach to word sense disambiguation.	*Co2*	*Po2*	**8**
			OR			
	8	a)	Comprehend the architectures for Coreference Algorithms with its applications.	*Co2*	*Po2*	**10**

		b)	Demonstrate the working of Alternate WSD algorithms and Tasks.	*Co1*	*Po1*	**10**
			UNIT - V			
	9	a)	Define machine translation. Illustrate why machine translation is hard?	*Co1*	*Po1*	**10**
		b)	Explain the following: i. Information extraction ii. Automatic text categorization	*Co1*	*Po1*	**10**
			OR			
	10	a)	Elucidate Information Retrieval and Information Extraction with neat diagram	*Co1*	*Po1*	**10**
		b)	Alex wants to automate the process of summarization for the text documents that he has. Analyze the various approaches that he could follow to summarize the text.	*Co3*	*Po3*	**10**

****