U.S.N.										

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

Autonomous Institute Affiliated to VTU

October 2024 Supplementary Examinations

Programme: B.E. **Semester: VI**

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

Course Code: 24AM6PENLP **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.

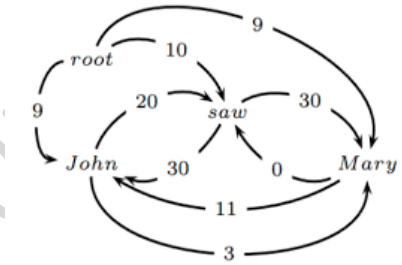
Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	CO	PO	Marks
1	a)	Elucidate the Phases of Natural Language Processing (NLP)with neat diagram.	CO1	PO1	**6**
	b)	Illustrate the ambiguities in NLP by suitable example for each.	CO1	PO1	**8**
	c)	Predict the regular expressions for the following languages. i. Date in MM-DD-YYYY format. ii. To identify the email IDs present in the given text. iii. Put angular brackets around all integers in text	CO2	PO2	**6**
		UNIT - II			
2	a)	Explain POS tagging with example. What are the 2 main classes of tagging algorithms in which they can be grouped into? Explain each one in detail.	CO2	PO2 PO9	**5**
	b)	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	PO3 PO5	**5**
	c)	Elaborate smoothing techniques with relevant mathematical equations.	CO1	PO1	**10**
		OR			
3	a)	Write Viterbi algorithm and Identify the POS tagging for the sentence "**James can drive the car**" given the transition and emission probabilities.	CO2	PO2 PO9	**10**

	NNP	MD	VB	JJ	NN	RB	DT
<s>	0.2767	0.0006	0.0031	0.0453	0.0449	0.0510	0.2026
NNP	0.3777	0.0110	0.0009	0.0084	0.0584	0.0090	0.0025
MD	0.0008	0.0002	0.7968	0.0005	0.0008	0.1698	0.0041
VB	0.0322	0.0005	0.0050	0.0837	0.0615	0.0514	0.2231
JJ	0.0366	0.0004	0.0001	0.0733	0.4509	0.0036	0.0036
NN	0.0096	0.0176	0.0014	0.0086	0.1216	0.0177	0.0068
RB	0.0068	0.0102	0.1011	0.1012	0.0120	0.0728	0.0479
DT	0.1147	0.0021	0.0002	0.2157	0.4744	0.0102	0.0017

Jamus can dire the car

NNP	0.000032	0	0	0.000048	0
MD	0	0.308431	0	0	0
VB	0	0.000028	0.000672	0	0.000028
JJ	0	0	0.000340	0	0
NN	0	0.000200	0.000223	0	0.002337
RB	0	0	0.010446	0	0
DT	0	0	0	0.506099	0

	b)	Consider the given chocolate corpus to answer the following questions: d1:Chocolate is a kind of sweet made from cocoa beans. d2:Cocoa originated in ancient Mexico and is now grown around the globe. d3:Cocoa contains an active compound that makes our body produce oxytocin. i. State chain rule ii. Construct a unigram and Bigram language model with probabilities.	*CO3*	*PO3*	**10**
		UNIT - III			
4	a)	Apply Chu-Liu Edmonds algorithm for finding a maximum spanning tree to the given weighted directed graph. Write the algorithm for the same.  	*CO1*	*PO1*	**10**
	b)	Parse the below sentences using dependency parsing: 1. Maria baked a big cake for john. 2. John can hit the ball. 3. The cat eats tasty fish.	*CO1*	*PO1*	**5**
	c)	Illustrate the working of transition based dependency parsing with a neat sketch.	*CO1*	*PO1*	**5**
		UNIT - IV			
5	a)	Derive the equation for Cosine similarity measure and illustrate the same to find whether ‘college’ and ‘natural’ is similar to ‘campus’ using given table.	*CO3*	*PO3* *PO5*	**10**

	students	language	canteen
college	442	8	2
natural	5	1683	1670
campus	5	3982	3325

	b)	Discuss different ways to convert word into vector with suitable examples.	*CO1*	*PO1*	**10**
		OR			
6	a)	Derive equation to calculate TF-IDF score. Calculate the Inverse Document Frequency for the following words by considering the given term frequency and total documents as 37.	*CO3*	*PO3* *PO5*	**10**
	b)	Explain the importance of word sense disambiguation for language technology with examples.	*CO1*	*PO1*	**5**
	c)	Infer the ways in which embeddings can be visualized also articulate the semantic properties of embeddings.	*Co1*	*PO1*	**5**
		UNIT - V			
7	a)	Comprehend the architectures for coreference resolution algorithms with its applications.	*CO1*	*PO1*	**10**
	b)	Discuss on the following pre-train models: i. BERT ii. RoBERTa	*CO3*	*PO3*	**10**

Word	df
Romeo	1
Salad	2
Forest	12
Battle	21
Fool	36
Good	37
sweet	37
Juliet	4

******