Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

May 2022

Max. Marks	**: 60**	**Duration: 120 Minutes**	
Class	**: TE Computer**	**Semester: VI**	
Course Code	**: CS 312**	**Branch : CS/IT**	
Name of the Course:	**Natural Language Processing(PE)**		

Instructions:

(1) All Questions are Compulsory

(2) Draw neat diagrams

Question No.		Max. Marks	BL
Q1a)	i) Assume that there are 10000 documents in a collection. Out of these, 50 documents contain the terms "difficult task". If "difficult task" appears 3 times in a particular document, what is the TFIDF value of the terms for that document? ii) Given a set of unigram and bigram probabilities, what is the probability of the following sequence ' do Sam I like' according to the bigram language model? P(do|) = 2/11, P(do|Sam) = 1/11, P(Sam|) = 4/11, P(Sam|do) = 1/8, P(I|Sam) = 4/11, P(Sam|I) = 2/9, P(I|do) = 2/8, P(I|like) = 2/7, P(like|I) = 3/11, P(do) = 3/8, P(Sam) = 2/11, P(I) = 4/11, P(like) = 5/11 iii) Consider the example and identify the type of ambiguity that exist in the conversation. Justify your answer. Tourist (checking out of the hotel): Waiter, go upstairs to my room and see if my sandals are there; do not be late; I have to catch the train in 15 minutes. Waiter (running upstairs and coming back panting): Yes sir, they are there.	2+2+2	3

Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

Q1b)	Distinguish between semantics, pragmatics and discourse in NLP. OR Explain 3 different POS tagging methods with example	9	2	CO1
Q2a)	Define Context Free Grammar. When is CFG said to be ambiguous? Extend the given CFG and generate all possible parse trees for the sentence "Astronomers saw stars with ears" S→NP VP PP→P NP VP→V NP VP→VP PP	2+2+2 +4	4	CO2
Q2b)	Using HMM & Viterbi decode a simple DNA sequence. Show all steps. A DNA sequence is a series of components from {A, C, G, T}. There is one hidden variable S that controls the generation of DNA sequence. S takes 2 possible states {S1, S2}. Given the following transition probabilities for HMM M P(S1|S1) = 0.8, P(S2|S1) = 0.2, P(S1|S2) = 0.2, P(S2|S2) = 0.8 emission probabilities as following P(A|S1) = 0.4, P(C|S1) = 0.1, P(G|S1) = 0.4, P(T|S1) = 0.1 P(A|S2) = 0.1, P(C|S2) = 0.4, P(G|S2) = 0.1, P(T|S2) = 0.4 and start probabilities as following P(S1) = 0.5, P(S2) = 0.5 Assume the observed sequence is x = CGTCAG, calculate the sequence of hidden state :	5	4	CO3
Q3a)	"Pattern matching by regular expressions are greedy." Justify with an example and give non greedy expression for the same scenario Write a regular expression to find cases of determiner "the" Explain the expression.Give one example each how it takes care of false positive and false negative.	5 + 5	3	CO
Q3b)	Explain chunking and named entity recognition with example	5	2	CO

Sardar Patel Institute of Technology

Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058-India
(Autonomous College Affiliated to University of Mumbai)

Q4a)	Write First Order Predicate logic representation for following sentence. List 2 applications of FOPL a)There are exactly 2 purple mushrooms. b)Every gardener likes the sun. c)Some prime number is even. d) X is above Y if X is on directly on top of Y or else there is a pile of one or more other objects directly on top of one another starting with Y and ending with X.	9	3	CO2
Q4b)	How is LSTM better than RNN in Natural Language Processing? Draw and explain LSTM architecture and discuss how it is used for classifying the sentiments for amazon Alexa product reviews into 'positive' and 'negative' categories. **OR** List 2 limitations of Encoder Decoder Network in machine translation. Draw and explain Attention model for translating Spanish text to English text.	6	3	CO4