Hall Ticket Number:

Code No. : 18222

VASAVI COLLEGE OF ENGINEERING *(AUTONOMOUS)*, HYDERABAD

Accredited by NAAC with A++ Grade

B.E. (C.S.E.) VIII-Semester Main & Backlog Examinations, June-2022

Natural Language Processing (PE-VI)

Time: **3 hours** Max. Marks: **60**

*Note: Answer all questions from **Part-A** and any **FIVE** from **Part-B***

Part-A (10 × 2 = 20 Marks)

Q. No.	Stem of the question	M	L	CO	PO
1.	Identify the NLP task performed by the following UNIX command line. tr -sc 'A-Za-z' '\n' < sh.txt	2	3	1	1,2,3
2.	What is Minimum Edit Distance Problem?	2	1	1	1,2
3.	Convert the following production rules to Chomsky Normal form S -> Aux NP VP S -> VP VP -> book | include | prefer	2	3	2	1,2,3
4.	What is the effect of the following operators on stack and input buffer in a transition based dependency parser? 1. LEFTARC 2. RIGHTARC	2	2	2	1,2
5.	Define inverse document frequency (idf) term weight.	2	1	3	1,2
6.	Give a vector representing the document "As you Like It" given the following term-document matrix.	2	2	3	1,2
7.	Which one of the following tasks is performed by Google Translate? a. Parts of speech tagging b. Machine Translation c. Semantic role labeling	2	1	4	1,2
8.	Differentiate between the semantic role and thematic role.	2	2	4	1,2

	As you like it	Twelfth Night	Julius Caesar	Henry V
Battle	1	0	7	13
Good	114	80	62	89
Fool	36	58	1	4
wit	20	15	2	3

Contd... 2

9.	What is a chatbot?	2	1	5	1,2
10.	Transform the following using ELIZA rule based chatbot. "You hate me"	2	3	5	1,2,3
	Part-B (5×8 = 40 Marks)				
11. a)	Explain the following regular expression in detail and identify the NLP task performed by it. s/.* I'M (depressed|sad) .*/I AM SORRY TO HEAR YOU ARE \1/	4	3	1	1,2,3
b)	Design an algorithm for solving the Minimum Edit Distance Problem.	4	3	1	1,2,3
12. a)	Draw CKY chart for the sentence "Book that flight" using the following mini-grammar	4	3	2	1,2,3

Rules
S -> NP VP
S -> V NP
NP -> Det N
VP -> V NP
Det -> the | a | that
N ->book| meal | flight
V -> book | includes

b)	Explain Probabilistic Context Free Grammar with an example.	4	2	2	1,2
13. a)	The total number of documents is 37 and document frequencies for the words (battle, good, fool, wit) are (21, 37, 36, 34) respectively. Compute the tf-idf weighted term document matrix for four words in four Shakespeare plays for the term-document matrix given below.	6	3	3	1,2,3

	As you like it	Twelfth Night	Julius Caesar	Henry V
Battle	1	0	7	13
Good	114	80	62	89
Fool	36	58	1	4
wit	20	15	2	3

Contd... 3

Code No. : 18222

b)	Given the following term-term matrix, find word similarity between the following words using cosine method. 1) digital and information 2) digital and cherry	2	3	3	1,2,3

	pie	data	computer
cherry	442	8	2
digital	5	1683	1670
information	5	3982	3325

14. a)	Define RNN. Give formulae for finding the hidden vector, out vector and softmax in RNN.	4	1	4	1,2
b)	Explain RNN Language Model	4	2	4	1,2
15. a)	Explain rule based chatbots	4	2	5	1,2
b)	Give applications of chatbots	4	2	5	1,2,3
16. a)	What is Markov assumption? With the help of suitable equations, explain how does the Markov assumption simplify the problem of finding the n-gram?	4	3	1	1,2,3
b)	Consider the following configuration of a transition based parser.	4	3	2	1,2,3

Stack	Word List	Relations
[root, book, the, morning, flight]	[]	(book → me)

The above configuration indicates one dependency relation. Construct the remaining dependency relations.

17.	Answer any ***two*** of the following:				
a)	Give the applications of TF-IDF vector model.	4	1	3	1,2
b)	Discuss about supervised learning of word sentiment.	4	2	4	1,2
c)	How to convert a raw wavefile to the most commonly used features, sequences of log mel spectrum vectors?	4	3	5	1,2,3

M : Marks; L: Bloom's Taxonomy Level; CO; Course Outcome; PO: Programme Outcome

i)	Blooms Taxonomy Level – 1	20%
ii)	Blooms Taxonomy Level – 2	30%
iii)	Blooms Taxonomy Level – 3 & 4	50%

*****