

Hall Ticket Number:

--	--	--	--	--	--	--	--	--	--

Code No. : 18232 N/O

VASAVI COLLEGE OF ENGINEERING (*AUTONOMOUS*), HYDERABAD

Accredited by NAAC with A++ Grade

B.E. (C.S.E.) VIII-Semester Main &amp; Backlog Examinations, May-2023

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.	Write a regular expression to find all the lines ending with a space character.	2	2	1	1,2
2.	Which algorithm design strategy is used to solve the Minimum Edit Distance problem?	2	1	1	1,2
3.	What are the ambiguity issues in the following phrase? "Old men and women"	2	2	2	1,2
4.	Convert the following general rule to Chomsky Normal form. A → B C D	2	2	2	1,2
5.	Give any two applications of TF-IDF vector model.	2	1	3	1,2
6.	Give any two applications of emotion detection.	2	1	3	1,2
7.	What is semantic role labeling in NLP?	2	2	4	1,2
8.	Match the pairs to identify the thematic role of each participant (shown in italic font) in the given sentences.  1. EXPERIENCER                      a. The city built <i>a regulation-size baseball diamond</i> ... 2. FORCE                                      b. <i>John</i> has a headache. 3. RESULT                                    c. I drove to <i>Portland</i> . 4. GOAL                                        d. <i>The wind</i> blows debris from the mall into our yards.	2	1	4	1,2
9.	Identify the NLP task performed by the following regular expression. s/. * I'M (depressed sad) .* / I AM SORRY TO HEAR YOU ARE \1 /	2	3	5	1,2,3
10.	With respect to automatic speech recognition, give the list of the various dimensions of the task variations.	2	2	5	1,2
<b>Part-B (5 × 8 = 40 Marks)</b>					
11. a)	It is required to write a regular expression to find the cases of the English article "the" in a given text. Is the regular expression /the/ sufficient? Does it produce any false positives and false negatives? If yes, what are those and how to address them?	4	3	1	1,2,3

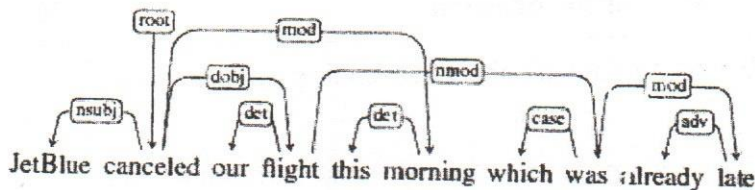
Contd... 2

b) A user wants to purchase a computer and his requirements are the following;  
 "Any machine with at least 6 GHz and 500 GB of disk space for less than \$1000".  
 Build and explain a regular expression for the above search operation?

4 3 1 1,2,3

12. a) How to identify projective arc in a given dependency tree. Are there any projective arcs in the following tree?

4 2 2 1,2



b) Construct a dependency tree for the following sentence.

4 3 2 1,2,3

"Corona virus troubled the human race a lot".

13. a) Given the following term-term matrix, compute the word similarity between the words in each of the pairs (cherry, information) and (digital, information).

4 3 3 1,2,3

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

b) Write a generic transition-based dependency parser and trace the algorithm for the sentence "Book me the morning flight"

4 3 3 1,2

14. a) Compute the tf-idf weighted term document matrix for four words in four Shakespeare plays for the term-document matrix given below. Given that the total number of documents is 37 and document frequencies for the words (battle, good, fool, wit) are (21, 37, 36, 34) respectively.

4 3 4 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

b) With the help of a neat diagram and equations, explain the implementation of encoder-decoder model using RNNs for machine translation.

4 3 4 1,2,3

15. a)	Write and explain an algorithm to generate rule based responses in chatbot applications.	4	2	5	1,2
b)	Explain corpus-based chat bots.	4	1	5	1,2
16. a)	How many tokens are there in the following text? "The U.S.A. helped low-income Pakistan about \$649.40 million for education ...". Write Python code to perform word tokenization using regular expressions.	4	2	1	1,2
b)	Design an algorithm to solve the Minimum Edit Distance problem. Using the algorithm transform "intention" to "execution"	4	3	2	1,2,3
17.	Answer any <i>two</i> of the following:				
a)	Why is the term-frequency weighing term insufficient in NLP? How to address it?	4	2	3	1,2
b)	Compute softmax for the following vector $Z = [0.6, 1.1, -1.5, 1.2, 3.2, -1.1]$	4	3	4	1,2,3
c)	Explain text-to-speech systems.	4	1	5	1,2

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

i)	Blooms Taxonomy Level - 1	20%
ii)	Blooms Taxonomy Level - 2	33%
iii)	Blooms Taxonomy Level - 3 & 4	47%

\*\*\*\*\*