



# VIT

Vellore Institute of Technology  
(Approved by the University Grants Commission of India, Act 1956)

REG. NO.: 22BCE3210

SLOT: E2

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING  
CONTINUOUS ASSESSMENT TEST - II  
FALL SEMESTER 2024-2025

Programme Name & Branch : BTech CSE  
 Course Code and Course Name : BCSE409L, Natural Language Processing  
 Faculty Name(s) : Dr. Varalakshmi, Dr. Suganthini, Dr. Padmavathy,  
 Dr. Lydia Jane, Dr. Hiteshwar Kumar, Dr. Rajeshkannan  
 Class Number(s) : VL2024250101856, 1899, 1896, 1876, 1881, 1861  
 Date of Examination : 17/10/2024  
 Exam Duration : 90 minutes Maximum Marks: 50

Answer All Questions

### Course Outcomes

- Check the syntactic and semantic correctness of natural language
- Select a suitable language modelling and feature representation to develop real-world applications

Q. No	Question	M	CO	BL
1.	Apply CYK algorithm to the two input sentences given below and demonstrate the use of parse trees in accepting syntactically correct sentences and rejecting wrong sentences. Use the following Context-Free Grammar. $S \rightarrow NP VP$ $NP \rightarrow DT NN$ $PP \rightarrow IN NP$ $VP \rightarrow VB NP$ $VP \rightarrow VP PP$ $DT \rightarrow the$ $NN \rightarrow boy$ $NN \rightarrow car$ $VB \rightarrow drove$ (i) the boy drove the car   (ii) drove boy the	10	3	3
2.	For constructing dependency tree using transition-based parsing, assume that you are provided with two different parsers ( <i>oracle</i> ) to predict the transitions. Generate the dependency tree for the sentence 'The white cat drank the milk' using both the parsers and determine which parser is more accurate. Justify your answer. Following are the lists of transitions predicted by the two parsers at each step for the example sentence. i) SHIFT, SHIFT, SHIFT, LEFTARC, LEFTARC, SHIFT, LEFTARC, SHIFT, SHIFT, LEFTARC, RIGHTARC ii) SHIFT, SHIFT, SHIFT, SHIFT, LEFTARC, SHIFT, SHIFT, LEFTARC, RIGHTARC	10	3	4
3.	a) Analyze if simple LESK algorithm will be successful or extended LESK algorithm should be applied to disambiguate the meaning of the word 'bank' in the sentence 'The river bank has plenty of fish'. Use the following glosses of its different senses and the wordnet for your analysis. i) an establishment for the custody, loan, exchange, or issue of money ii) the land alongside a waterbody Meronyms of sense 1 -- locker, ATM, cash counter Meronyms of sense 2 -- water, fish, boat b) Assume the word embedding of some target word to be [1,2] and there are two senses for the word – sense 1 with 2 words and sense 2 with 3 words whose embeddings are given below. Apply 1-nearest-neighbor algorithm to	7          3	3	4



# VIT

Vellore Institute of Technology

REG.NO.: 22BCE3216

SLOT: E2

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING  
CONTINUOUS ASSESSMENT TEST - II  
FALL SEMESTER 2024-2025

	<p>disambiguate the meaning of the target word and determine the nearest neighbor sense for the word.</p> <p>Embeddings for the 2 words in sense 1: [-3,0],[-1,2]</p> <p>Embeddings for the 3 words in sense 2: [3, 3],[2, 4], [1,2]</p>			
4.	<p>a) Given the vocabulary with 5 unique words "He", "poured", "the", "waste", "water", after applying word2vec model to generate the embeddings, which of the following could be the possible trained weight matrix between the input layer and hidden layer? What is the embedding for the word "waste" and what is its dimension?</p> $\begin{bmatrix} 1 & 7 & 3 \\ 2 & 5 & 6 \\ 7 & 4 & 3 \\ 6 & 9 & 2 \\ 8 & 1 & 5 \end{bmatrix} \text{ or } \begin{bmatrix} 5 & 2 & 1 & 8 & 3 \\ 4 & 6 & 0 & 7 & 5 \\ 7 & 2 & 9 & 4 & 3 \end{bmatrix}$	3		
	<p>b) The binary classifier of the mention-pair coreference model should be trained with both positive and negative coreference samples to determine whether or not a mention is an antecedent of the anaphor. Given below are an example sentence and all possible pairs of mentions for the anaphor 'it'. Select the positive and negative samples out of those pairs and justify your choice.</p> <p>'Sunil poured water from the bottle into the cup as he wanted to drink it'.</p> <p>Mention-pairs: (it, Sunil), (it, water), (it, bottle), (it, cup), (it, he)</p> <p>Assuming the model does not use the mention-pair (it, Sunil) for training, state the possible reason.</p>	3	2	2
	<p>c) It is important to identify and filter out singleton mentions from coreferent mentions as it reduces the search space and improves the accuracy of downstream coreference resolution.</p> <p>Mary was at her apartment reading a book. She finished it at 8 PM.</p> <p>Identify the mentions in the aforementioned example sentence. Which is the singleton mention in the above example? Give reasons.</p> <p>How does the mention-ranking coreference model help to remove the singleton mention from the example sentence?</p>	4		
5.	<p>&lt;s&gt; the black cat chased the rat &lt;/s&gt;</p> <p>&lt;s&gt; the rat chased the chick &lt;/s&gt;</p> <p>&lt;s&gt; the man caught the chick &lt;/s&gt;</p> <p>&lt;s&gt; the black cat ate the rat &lt;/s&gt;</p> <p>Based on the tiny corpus given above, apply the bi-gram and tri-gram models to predict the probability of the sentence " &lt;s&gt; the cat chased &lt;/s&gt;". Use the necessary technique to handle the zero-probability in the calculations. Choose the best of the two models based on the perplexity measure.</p>	10	4	3

\*\*\*\*\*