



VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026**

SLOT:C1+TC1

Programme Name & Branch : B.Tech CSE
Course Code and Course Name : BCSE409L , Natural Language Processing
Faculty Name(s) : Dr. SARAVANAKUMAR K, Dr. KANNADASAN R,
 Dr. RAJESHKANNAN R, Dr. SATHYARAJ R,
 Dr. JABANJALIN HILDA J, Dr. ARIVOLI A
Class Number(s) : VL2025260502034, 2035, 2032, 2033, 2039, 2270
Date of Examination : 17/03/26
Exam Duration : 90 minutes **Maximum Marks: 50**

General instruction(s): Answer All Questions

- M - Max mark; CO – Course Outcome; BL – Blooms Taxonomy Level (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyse, 5 – Evaluate, 6 – Create)
- Course Outcomes
 - CO3 - Check the syntactic and semantic correctness of natural language.
 - CO4 – Select a suitable language modelling and feature representation to develop real-world applications.

Q. No	Question	M	CO	BL																									
1.	Consider the sentence: “The young scientist analyzed the complex data in the laboratory.” Using transition-based dependency parsing, derive the dependency structure for the given sentence. Present the step-by-step parsing process in a transition table, listing the stack, input word list (buffer), action taken, and dependency relation added at each step. Finally, draw the resulting dependency tree/graph.	10	3	3																									
	Transition-based dependency parsing - Transition table <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Step</th> <th>Stack</th> <th>Buffer</th> <th>Action</th> <th>Relation Added</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>[ROOT]</td> <td>[The, young, scientist, analyzed, the, complex, data, in, the, laboratory]</td> <td>SHIFT</td> <td></td> </tr> <tr> <td>1</td> <td>[ROOT, The]</td> <td>[young, scientist, analyzed, the, complex, data, in, the, laboratory]</td> <td>SHIFT</td> <td></td> </tr> <tr> <td>2</td> <td>[ROOT, The, young]</td> <td>[scientist, analyzed, the, complex, data, in, the, laboratory]</td> <td>SHIFT</td> <td></td> </tr> <tr> <td>3</td> <td>[ROOT, The, young,</td> <td>[analyzed, the, complex, data, in, the, laboratory]</td> <td>LEFT-ARC</td> <td>(scientist ← young)</td> </tr> </tbody> </table>	Step	Stack	Buffer	Action	Relation Added	0	[ROOT]	[The, young, scientist, analyzed, the, complex, data, in, the, laboratory]	SHIFT		1	[ROOT, The]	[young, scientist, analyzed, the, complex, data, in, the, laboratory]	SHIFT		2	[ROOT, The, young]	[scientist, analyzed, the, complex, data, in, the, laboratory]	SHIFT		3	[ROOT, The, young,	[analyzed, the, complex, data, in, the, laboratory]	LEFT-ARC	(scientist ← young)			
Step	Stack	Buffer	Action	Relation Added																									
0	[ROOT]	[The, young, scientist, analyzed, the, complex, data, in, the, laboratory]	SHIFT																										
1	[ROOT, The]	[young, scientist, analyzed, the, complex, data, in, the, laboratory]	SHIFT																										
2	[ROOT, The, young]	[scientist, analyzed, the, complex, data, in, the, laboratory]	SHIFT																										
3	[ROOT, The, young,	[analyzed, the, complex, data, in, the, laboratory]	LEFT-ARC	(scientist ← young)																									



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

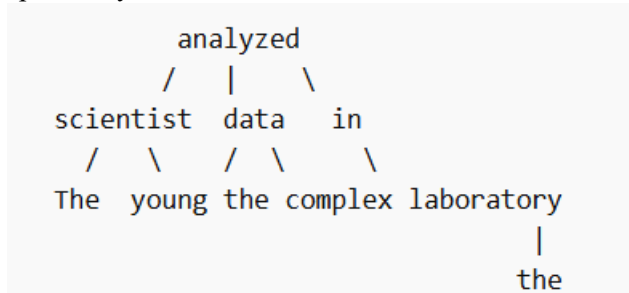
	scientist]			
4	[ROOT, The, scientist]	[analyzed, the, complex, data, in, the, laboratory]	LEFT-ARC	(scientist ← The)
5	[ROOT, scientist]	[analyzed, the, complex, data, in, the, laboratory]	SHIFT	
6	[ROOT, scientist, analyzed]	[the, complex, data, in, the, laboratory]	LEFT-ARC	(analyzed ← scientist)
7	[ROOT, analyzed]	[the, complex, data, in, the, laboratory]	SHIFT	
8	[ROOT, analyzed, the]	[complex, data, in, the, laboratory]	SHIFT	
9	[ROOT, analyzed, the, complex]	[data, in, the, laboratory]	SHIFT	
10	[ROOT, analyzed, the, complex, data]	[in, the, laboratory]	LEFT-ARC	(data ← complex)
11	[ROOT, analyzed, the, data]	[in, the, laboratory]	LEFT-ARC	(data ← the)
12	[ROOT, analyzed, data]	[in, the, laboratory]	RIGHT-ARC	(analyzed → data)
13	[ROOT, analyzed]	[in, the, laboratory]	SHIFT	
14	[ROOT, analyzed, in]	[the, laboratory]	SHIFT	



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

15	[ROOT, analyzed, in, the]	[laboratory]	SHIFT	
16	[ROOT, analyzed, in, the, laboratory]	[]	LEFT-ARC	(laboratory ← the)
17	[ROOT, analyzed, in, laboratory]	[]	RIGHT-ARC	(in → laboratory)
18	[ROOT, analyzed, in]	[]	RIGHT-ARC	(analyzed → in)
19	[ROOT, analyzed]	[]	RIGHT-ARC	(ROOT → analyzed)
20	[ROOT]	[]	DONE	

Dependency tree -



2.	<p>Consider the following sentences:</p> <ol style="list-style-type: none"> The idea slept quietly. The table opened the door with a key. The number two is red. The car read the newspaper. <p>For each sentence, perform Semantic Role Labelling (SRL) by identifying the appropriate semantic roles. Examine whether these roles satisfy the selectional restrictions of the predicate. Based on this analysis, determine whether the sentence is semantically valid or anomalous, and justify your answer.</p>	10	3	4
----	--	----	---	---



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

<p>1. Sentence: <i>The idea slept quietly.</i></p> <p>SRL Output</p> <p>Agent → idea</p> <p>Selectional Restriction Check</p> <p>sleep → requires animate/living entity idea → abstract concept</p> <p>Conclusion</p> <p>The Agent role does not satisfy the selectional restriction of the predicate <i>sleep</i>. The sentence is semantically anomalous.</p> <p>2. Sentence: <i>The table opened the door with a key.</i></p> <p>SRL Output</p> <p>Agent → table Patient → door Instrument → key</p> <p>Selectional Restriction Check</p> <p>open → requires animate / intentional agent table → inanimate object door → physical object key → instrument/tool</p> <p>Conclusion</p> <p>The Agent violates the selectional restriction of the predicate <i>open</i>. The sentence is semantically anomalous.</p> <p>3. Sentence: <i>The number two is red.</i></p> <p>SRL Output</p> <p>Theme → number two Attribute → red</p> <p>Selectional Restriction Check</p> <p>red → color property typically applied to physical objects number two → abstract numerical entity</p> <p>Conclusion</p> <p>The Theme does not satisfy the semantic constraint of the predicate. The sentence is semantically anomalous.</p>			
---	--	--	--



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

	<p>4. Sentence: <i>The car read the newspaper.</i></p> <p>SRL Output</p> <p>Agent → car Patient → newspaper</p> <p>Selectional Restriction Check</p> <p>read → requires animate/intelligent agent car → inanimate object newspaper → readable object</p> <p>Conclusion</p> <p>The Agent violates the selectional restriction of the predicate <i>read</i>. The sentence is semantically anomalous.</p>			
3.	<p>Consider the following sentences: Sentence 1: The young boy met Ravi near the park. Sentence 2: He greeted him with a friendly smile. The following Context-Free Grammar (CFG) is provided: S → NP VP VP → V NP V NP PP NP → Pronoun ProperNoun Det N Det Adj N PP → P NP Det → the a Adj → young friendly V → met greeted N → boy park smile ProperNoun → Ravi Pronoun → he him P → near with</p> <p>Construct the parse trees for the sentences using the given grammar and apply Hobbs' Algorithm to resolve the anaphors "he" and "him." Identify the possible antecedents and demonstrate the step-by-step traversal of the parse tree used by the algorithm to determine the most likely antecedents. Finally, rewrite the sentences by replacing the pronouns with their resolved antecedents.</p>	10	3	3
	<p>Draw parse Tree</p> <p>Hobbs' Algorithm</p> <p>Hobbs' Algorithm resolves pronouns by:</p> <ol style="list-style-type: none"> Starting at the NP node of the pronoun Moving up the parse tree to the nearest NP or S 			



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

<p>3. Searching left-to-right breadth-first for candidate NPs</p> <p>4. If no suitable NP is found, the algorithm moves to previous sentences</p> <p>The first NP satisfying agreement constraints is chosen as the antecedent.</p> <p>3. Resolving “he”</p> <p>Pronoun: he</p> <p>Candidate antecedents from the previous sentence:</p> <ul style="list-style-type: none"> • boy • Ravi <p>Hobbs traversal examines the previous sentence NPs.</p> <p>Search order:</p> <p>NP → boy</p> <p>NP → Ravi</p> <p>The first suitable NP encountered is “boy.”</p> <p>Therefore:</p> <p>he → boy</p> <p>4. Resolving “him”</p> <p>Pronoun: him</p> <p>Candidate antecedents:</p> <ul style="list-style-type: none"> • boy • Ravi <p>Search order again:</p> <p>boy</p> <p>Ravi</p> <p>Since boy already corresponds to the subject pronoun “he”, the next valid candidate is selected.</p> <p>Therefore:</p> <p>him → Ravi</p> <p>Final Resolution</p> <table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; padding: 5px;">Pronoun</th> <th style="text-align: left; padding: 5px;">Antecedent</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;">he</td> <td style="padding: 5px;">boy</td> </tr> </tbody> </table>	Pronoun	Antecedent	he	boy			
Pronoun	Antecedent						
he	boy						



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

	<p>him Ravi</p> <p>Rewritten Sentences</p> <p>The young boy met Ravi near the park. The young boy greeted Ravi with a friendly smile.</p>																											
4	<p>Consider the following sentences:</p> <p>Sentence A: "The tiger chased the deer." Sentence B: "The lion hunted the deer." Sentence C: "The tiger hunted the deer." The vocabulary is: {tiger, lion, deer, chased, hunted} The sentences are represented using Term frequency (TF) vectors:</p> <table border="1" data-bbox="279 869 1002 1111"> <thead> <tr> <th>Term</th> <th>Sentence A</th> <th>Sentence B</th> <th>Sentence C</th> </tr> </thead> <tbody> <tr> <td>tiger</td> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>lion</td> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>deer</td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>chased</td> <td>1</td> <td>0</td> <td>0</td> </tr> <tr> <td>hunted</td> <td>0</td> <td>1</td> <td>1</td> </tr> </tbody> </table> <p>a) Compute the Inverse Document Frequency (IDF) for each Term. b) Compute the TF-IDF vectors for Sentence A, Sentence B, and Sentence C. c) Using the TF-IDF vectors, compute the cosine similarity between: i) Sentence A and Sentence B ii) Sentence A and Sentence C d) Interpret the results. Which pair of sentences is more similar and why?</p>	Term	Sentence A	Sentence B	Sentence C	tiger	1	0	1	lion	0	1	0	deer	1	1	1	chased	1	0	0	hunted	0	1	1	10	4	5
Term	Sentence A	Sentence B	Sentence C																									
tiger	1	0	1																									
lion	0	1	0																									
deer	1	1	1																									
chased	1	0	0																									
hunted	0	1	1																									
	<p>Total Number of Sentences</p> <p>$N = 3$</p> <p>Document Frequency (df)</p> <p>Term Appears in df</p> <p>tiger A, C 2 lion B 1 deer A, B, C 3 chased A 1 hunted B, C 2</p>																											



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

Inverse Document Frequency (IDF)

Formula:

$$IDF(t) = \log \left(\frac{N}{df(t)} \right)$$

Using log base 10:

Term df IDF

tiger 2 $\log(3/2) = 0.176$

lion 1 $\log(3/1) = 0.477$

deer 3 $\log(3/3) = 0$

chased 1 0.477

hunted 2 0.176

TF-IDF Vectors

TF × IDF

Sentence A

Term TF IDF TF-IDF

tiger 1 0.176 0.176

lion 0 0.477 0

deer 1 0 0

chased 1 0.477 0.477

hunted 0 0.176 0

Vector A

A=(0.176, 0, 0, 0.477, 0)



**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026**

Sentence B

Term TF IDF TF-IDF

tiger	0	0.176	0
lion	1	0.477	0.477
deer	1	0	0
chased	0	0.477	0
hunted	1	0.176	0.176

Vector B

B=(0, 0.477, 0, 0, 0.176)

Sentence C

Term TF IDF TF-IDF

tiger	1	0.176	0.176
lion	0	0.477	0
deer	1	0	0
chased	0	0.477	0
hunted	1	0.176	0.176

Vector C

C=(0.176, 0, 0, 0, 0.176)

Sentence A and Sentence B Cosine Similarity

Formula:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

Sentence A vs Sentence B =

Dot product

$$A \cdot B = 0$$

Therefore



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

	<p>$\cos(A, B) = 0$</p> <p>Sentence A vs Sentence C Cosine Similarity</p> <p>Dot product</p> $0.176 \times 0.176 = 0.031$ <p>Vector magnitudes</p> $\ A\ = \sqrt{0.176^2 + 0.477^2} = 0.508$ $\ C\ = \sqrt{0.176^2 + 0.176^2} = 0.249$ <p>Cosine similarity</p> $\cos(A, C) = \frac{0.031}{0.508 \times 0.249}$ $\cos(A, C) \approx 0.246$ <p>Similarity(A,C) > Similarity(A,B)</p>			
5.	<p>Consider the following training corpus:</p> <p><s> birds build small nests </s></p> <p><s> birds eat small insects </s></p> <p><s> cats chase small mice </s></p> <p><s> dogs chase small cats </s></p> <p><s> birds watch small insects </s></p> <p>Test sentence: <s> birds chase small mice </s></p> <p>(a) Compute the Bigram and Trigram probabilities of the test sentence using Laplace smoothing.</p> <p>(b) Compute the perplexity of the test sentence using both the Bigram and Trigram models after applying Laplace smoothing.</p> <p>(c) Compare the results and justify which model performs better for this case.</p>	10	4	5
	<p>V=13</p> <p>Bigram Sentence Probability</p> $P = 0.222 \times 0.0625 \times 0.20 \times 0.111 \times 0.143$ $P \approx 0.000044$			



VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

SLOT:C1+TC1

	<p>Trigram Sentence Probability</p> $P = 0.0625 \times 0.0769 \times 0.133 \times 0.143$ $P \approx 0.0000916$ <p>Bigram Perplexity</p> $N = 5$ $PP = (0.000044)^{-1/5}$ $PP \approx 7.37$ <p>Trigram Perplexity</p> $PP = (0.0000916)^{-1/5}$ $PP \approx 6.45$ <p>The Trigram model performs better because it has lower perplexity. The trigram model uses two previous words as context, allowing it to capture more contextual information compared to the bigram model. Therefore, it provides better prediction for the test sentence.</p>			
--	--	--	--	--
