VIT

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

Reg. Number:	21BCE5067

Continuous Assessment Test (CAT) – I (August 2024)

Programme	:	B. Tech. Computer Science and Engineering	Semester	:	**Fall Sem 24-25**
Course Code & Course Title	:	BCSE409L Natural Language Processing	Class Number	:	CH2024250101611 CH2024250101619 CH2024250101621 CH2024250101636 CH2024250101638
Faculty	:	**Dr.Manjula.D** **Dr.R.Krithiga** **Dr. Lakshmi Harika Palivela** **Dr.S.Sharmiladevi** **Dr.Gayathri.R**	Slot	:	B2+TB2
Duration	:	1½ Hours	Max. Mark		50

General Instructions:

- Write only your Reg. No. on the question paper in the box provided and do not write other information.
- Use statistical tables supplied from the exam cell as necessary
- Use graph sheets supplied from the exam cell as necessary
- Only non-programmable calculator without storage is permitted

Answer all questions

Q. No	Sub Sec.	Description	Marks
1		Read the following sentences and illustrate the different types of ambiguities present in each: *a) The bank was crowded.* *b) Visiting relatives can be annoying.* *c) John said he would give a ride to Tom, but he didn't show up.* *d) The professor discussed the student's paper in the office.* *e) The chicken is ready to eat.*	10
2		Assume that you are developing a text analysis system for a historical research application. This system needs to process and interpret texts related to significant events and figures from history or mythology. One such text is about the historical figure Krishna and his actions. Consider the following historical narrative: *In the epic Mahabharata, Krishna played a crucial role in the Kurukshetra battle and is renowned for killing the warrior Karna, a pivotal moment in the conflict.* Illustrate the various phases of Natural Language Processing (NLP) applied to the given scenario.	10

3		Consider the following sentences: • *It is going to sunny today.* • *I am going to watch the season premiere.* • *Today I am not going Chennai.* • *Chennai is very hot today.* • *"The quick brown fox jumps over the lazy dog. The dog, which was very lazy, slept all day."* List the steps required to tokenize these sentences and obtain the tokens. Calculate the type-token ratio for each Sentence.	10
4		Consider the following text corpus: • *The cat eats fish.* • *The dog barks.* • *A cat plays.* • *A dog eats.* • *The cat sleeps.* • *The dog runs.* • *A cat jumps.* Compute the emission and transition probability for the given corpus.	10
5		For the following transformations, draw a two-level transducer diagram illustrating the process: a. *Take → Taking* b. *Fire → Fires* c. *Large → Larger* d. *Bake → Baked* For each word transformation, identify which rule is applied from the base form. Specifically, explain the rule for e-insertion or retention, and describe how it affects the transformation process for the chosen example.	10

**********All the best **********