VIT
Vellore Institute of Technology
CHENNAI

| Reg. Number: | 22 BCE 1178 |
|---|---|

# Continuous Assessment Test (CAT) – I (January 2025)

| | | | | | |
|---|---|---|---|---|---|
| Programme | : | B.Tech. Computer and Science Engineering | Semester | : | Winter-2025 |
| Course Code & Course Title | : | BCSE409L & Natural Language Processing | Class Number | : | CH2024250501967 CH2024250501965 |
| Faculty | : | **Dr.Krithiga.R** **Dr. Vijayaprabakaran K** | Slot | : | D2+TD2 |
| Duration | : | 1½ Hours | Max. Mark | | 50 |

**General Instructions:**

- Write only your registration number 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 | | Consider the strings A = "internationalization" and B = "intercontinentalization".<br>a) Compute the transformation cost matrix for each character pair using the dynamic programming approach to find the minimum edit distance. [6 Marks]<br>b) Identify the sequence of operations (insertion, deletion, or substitution) needed to convert A into B, including the specific positions for each operation.[4 Marks] | 10 |
| 2 | | Design a Natural Language Processing (NLP) system for a healthcare provider to automatically extract important information from patient medical records, such as diagnoses, symptoms, treatments, and medication prescriptions.<br>a) Discuss key NLP techniques useful for extracting and organizing medical data from unstructured text, and briefly explain the role of each technique. [8 Marks]<br>b) Mention the challenges in processing healthcare data using an NLP system. [2 Marks] | 10 |

| | | | |
|---|---|---|---|
| 3 | | Classify each of the following words as undergoing either inflectional or derivational changes, and provide their base forms. For each word, explain the morphological transformation.<br>a)"cats"<br>b)"teacher"<br>c)"running"<br>d)"happiness"<br>e)"walked"<br>f)"unfriendly"<br>g)"quickly" | 10 |
| 4 | | Design a finite-state transducer (or a set of transducers) to model the consonant-doubling rule in English spelling for verbs that double a single final consonant when adding the suffix "-ing." For example, "beg" + "ing" becomes "begging," while "run" + "ing" becomes "running." However, certain verbs, like "sing," remain "singing" without consonant doubling. The transducer should accept any input string, transforming those that require consonant doubling while preserving exceptions. Provide a diagram representing the transducer's state transitions, demonstrating its operation on sample words, including cases with and without consonant doubling. | 10 |
| 5 | | Use the Hidden Markov Model (HMM) tagger to tag the phrase "Knowledge is power" with the following probabilities:<br>• Initial Probabilities:<br>P(NN) = 0.3, P(VB) = 0.7<br>• Transition Probabilities:<br>P(NN → NN) = 0.6, P(NN → VB) = 0.4; P(VB → NN) = 0.5, P(VB → VB) = 0.5<br>• Emission Probabilities:<br>P(Knowledge \| NN) = 0.7, P(Knowledge \| VB) = 0.3 ; P(is \| NN) = 0.2, P(is \| VB) = 0.8; P(power \| NN) = 0.1, P(power \| VB) = 0.9<br>Apply the Viterbi algorithm to analyze the tags. Assume all other conditional probabilities not mentioned are zero. | 10 |

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