JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2026

M.Tech-II Semester (Data Science)

COURSE CODE (CREDITS): 22M1WCI236 (3)

MAX. MARKS: 15

COURSE NAME: Natural Language Processing

COURSE INSTRUCTOR: HRI   MAX. TIME: 1 Hour

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

(c) Use of calculators is allowed.

<table>
<tr><th>Q.No</th><th>Question</th><th>CO</th><th>Marks</th></tr>
<tr><td>Q1</td><td>Assuming a corpus follows Heap's law, derive the relationship describing how the TTR changes as a function of the corpus size N.</td><td>CO1</td><td>3</td></tr>
<tr><td>Q2</td><td>Draw a complete Edit Distance (Levenshtein Distance) matrix between the strings S1= brisk and S2 = brick, assuming a cost of 1 for insertion, deletion, and substitution. What are the minimum cost and the operation?</td><td>CO1</td><td>4</td></tr>
<tr><td>Q3</td><td>The probabilities $P(q_{n+1}/q_n)$ of tomorrow's weather based on today's weather is given below. Given that today the weather is sunny, what is the probability that tomorrow is sunny and day after is rainy according to the Markov Chain transition table?
<table>
<tr><td></td><td>Tomorrow's weather</td><td></td><td></td></tr>
<tr><td>Today's weather</td><td>Sunny</td><td>Rainy</td><td>Foggy</td></tr>
<tr><td>Sunny</td><td>0.8</td><td>0.05</td><td>0.15</td></tr>
<tr><td>Rainy</td><td>0.2</td><td>0.6</td><td>0.2</td></tr>
<tr><td>Foggy</td><td>0.2</td><td>0.3</td><td>0.5</td></tr>
</table>
</td><td>CO2</td><td>3</td></tr>
<tr><td>Q4</td><td>Suppose you want to use a HMM tagger to tag the phrase, "the light book", where we have the following probabilities:
P(the|Det) = 0.3, P(the|Noun) = 0.1, P(light|Noun) = 0.003, P(light|Adj) = 0.002, P(light|Verb) = 0.06, P(book|Noun) = 0.003, P(book|Verb) = 0.01
P(Verb|Det) = 0.00001, P(Noun|Det) = 0.5, P(Adj|Det) = 0.3, P(Noun|Noun) =0.2, P(Adj|Noun) = 0.002, P(Noun|Adj) = 0.2, P(Noun|Verb) = 0.3, P(Verb|Noun) = 0.3, P(Verb|Adj) = 0.001, P(Verb|Verb) = 0.1

Work out in details the steps of the Viterbi algorithm. Assume all other conditional probabilities, not mentioned to be zero. Also, assume that all tags have the same probabilities to appear in the beginning of a sentence.</td><td>CO2</td><td>5</td></tr>
</table>

JUIT TEST-1 EXAMINATION FEB-2026

158