60009210199

4/3/24
12:30 to 1:30

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Computer Science and Engineering (Data Science)

Term Test – I (AY- 2023-2024)

Class: Third Year	**Semester:** VI
Course Name: Computational Linguistics	**Course Code:** DJ19DSC602
Total marks: 25	**Time:** 1 hr.

Q. No.	Question	Max. Marks
Q1A	Calculate the edit distance using Levenshtein distance for the string's "algorithm" and "logarithm," considering insertion and deletion costs as 1 and substitution cost as 2.	05
Q2A	Consider the following Training Data <S>I am Jack</S> <S>Jack I am</S> <S>Jack I like</S> <S>Jack I do like</S> <S>do I like Jack</S> Apply Bigram language model on the above data and find what is the most probable next word predicated by the model for the following. <S>do I like __? _	05
	OR	
Q2B	Demonstrate how two-level morphology can be applied to analyze and generate lexical forms of words using surface forms. Draw diagram for surface level, intermediate level and lexical level for the following words. 1. Foxes 2. Played	05
Q3A	Differentiate between stemming and lemmatization using suitable example.	05
	OR	
Q3B	Explain Substitutions and Capture Group in Regular expression. Write the regular expression for the following sentences using substitution and capture group. A) The harder he worked, the harder she worked. B) The happier he became, the happier she became.	05
Q4A	Justify with the suitable example "Parts of Speech Tagging is disambiguation task"	05
Q4B	Discuss with suitable example how the HMM algorithm is trained for Parts of Speech (POS) tagging. What kind of information is required during the training phase?	05