VIT-AP UNIVERSITY

QUESTION PAPER

Name of the Examination: Win 2023-24 Semester – CAT-1

Course Code: CSE 3015   **Course Title: Natural Language Processing**

Set number: 5   **Date of Exam:**

Duration: 90 Min   **Total Marks: 50**

Instructions:

1. Assume data wherever necessary.

2. Any assumptions made should be clearly stated.

Q1. Identify the sources of multiple meanings in the provided sentence and determine whether these meanings arise from structural ambiguity, semantic ambiguity, or pragmatic ambiguity   **[10M]**

    (i)   She told him the story sitting on the bench.
    (ii)   They saw the man with the telescope.
    (iii)   You have to fly to Delhi to attend that interview.
    (iv)   Raja saw Ravi with his binocular.
    (v)   The old house had leaks.

Q2. Find the root word of each token for the below sentence using lemmatization with their respective POS of each token using NLTK library and explain each step involved.   **[10M]**

"Natural language processing (NLP) is a subfield of artificial intelligence (AI) that focuses on the interaction between computers and humans through natural language. NLP techniques enable computers to understand, interpret, and generate human language in a way that is both meaningful and useful."

Q3. Design the regular expression tagger to tag the following tokens using the necessary POS rules. Use the POS tags   **[10M]**

["Running","Car","Is","Beautiful","Dog","yellow","Quickly","Jumping","over","Lazy",Fox"]

Q4. (a) Identify and classify various types of named entities within the provided content, placing them into one of five distinct categories: person, organization, time, location, and/or work of art.   **[5M]**

"Google, founded by Larry Page and Sergey Brin, is a multinational technology company specializing in internet-related services and products. It was incorporated on September 4, 1998, in Menlo Park, California."

   (b) Compare and contrast the Affix and N-gram taggers. Then, outline a program for implementing one of these taggers.   **[5M]**

Q5. Develop an HMM tagger by estimating the necessary transition and emission probabilities from a provided set of training sentences. Then, utilize the trained HMM tagger to perform POS tagging on the sentence "Will Will google CampusX" .   **[10M]**

Training sentences:
can/model   Laasya/noun   google/verb   CampusX/noun
will/model   Mahira/noun   google/verb   Campusx/noun
Mahira/noun   loves/verb   Will /noun
Will /noun   loves/verb   google/noun

QP MAPPING

Q. No.	Module Number	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	1	CO1	2,3,5	PEO2	PSO1	10
Q2	1	CO1	2,3,5	PEO2	PSO1	10
Q3	2	CO2	2,3,5	PEO2	PSO1	10
Q4	2	CO2	2,3,5	PEO2	PSO1	10
Q5	2	CO3	4,5,6	PEO3	PSO3	10