



QUESTION PAPER

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

Course Code: CSE3015

Course Title: NLP

Set number: 02

Date of Exam: 11/09/2023 (EN) (B)

Duration: 90 Minutes

Total Marks: 50M

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) looked up the tree
- (ii) Josh stirs the soup with a wooden spoon
- (iii) You have to fly to Delhi to attend that interview
- (iv) Raja saw Ravi with his binocular
- (v) There is a fly in my cup

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]

“The challenges in analyzing big data encompass a wide spectrum, including data capture, storage, analysis, search, sharing, transfer, visualization, querying, updating, maintaining information privacy, and sourcing data. Initially, big data was primarily characterized by three fundamental aspects: volume, variety, and velocity. Analyzing big data poses difficulties in the realm of sampling, which traditionally limited the focus to observations and sample-based approaches.”

Q3. Using the NLTK library's proper regular expression patterns, stem the following tokens and then describe the issues in stemmed words. [10M]

[Advisable, Drinking, Eating, Swimming, Comparable, Computers, Eats, Considered, Informed, Asked]

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]

"Dr. Bhimrao Ambedkar is a very popular personality in India. He was popularly known as Dr Babasaheb Ambedkar. He was a leading politician, activist, social reformer, philosopher, writer, politician, jurist, and economist and had an impactful personality. He made everyone realize the necessity of education, discipline, and well-being. Dr BR Ambedkar became the first law minister of the country in 1947. He was awarded the highest citizen award – Bharat Ratan, in 1990 posthumously."

(b) Differentiate between Regex and N-gram tagger. Design a program for both Regex and N-gram tagger and explain the reason for similarity or difference in the output.

[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 "Rama will see Ravana." [10M]

Training sentences:

Rama/noun can/model see/verb flight/noun

Lakshmana/noun will/model see/verb Hanuma/noun

Will/model Ravana/noun spot/verb Hanuma/noun

Hanuma/noun will/model pat/verb Spot/noun

QP MAPPING

Q. No.	Module Number	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	1	1, 2	1,2	1	-	10
Q2	1	1, 2	1,2	1	-	10
Q3	1	1, 2	1,2	1	-	10
Q4	2	3	1,2,3	1,3	-	10
Q5	2	3	1,2,3	1,3	-	10