



QUESTION PAPER

Name of the Examination: WINTER 2023-2024 – CAT-1

Course Code: CSE3015

Course Title: Natural language Processing

Set number: 9

Date of Exam:

Duration: 90 minutes

Total Marks: 50

Instructions:

1. Assume data wherever necessary.
2. Any assumptions made should be clearly stated.

Q1. a. Write a program snowball stemmer for the below list
[“generous','generate','generously','generation, friendships”]

b. Explain the different steps in NLP with examples **(5+10)M**

Q2. a. Perform stemming on the following tokens using appropriate regular expression patterns from NLTK library and explain the problems in stemmed words.

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

b. Calculate the number of operations required to convert the string from `str1=“GEEXSFRGEEKKS”` and `str2=“GEEKSFORGEEKS”` using `editDistance()`

(6+6)M

Q3. Design HMM model for below training sentences and calculate transition and emission probability

And assign appropriate POS Tagging for the below sentence.

[15]M

“Will had spot Mary”

Training sentence

mary/**noun** can/**model** see/**verb**.

will/**model** spot/**verb** she/**preposition** pat/**verb** great/**adverb**.

spot/**noun** learn/**verb** jane/**noun**.

So/**conjunction**, learn/**verb**, had/**verb**.

Q4. Explain about Brill and sequence Tagger in detail with an an example program

(12) M

QP MAPPING

Q. No.	Module Number	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	1	1, 2	1,2	1		15
Q2	1	1, 2	1,2	1,3		12
Q3	2	2, 2	1,2	1,3		15
Q4	2	3	1,2,3	1,3		12