



QUESTION PAPER

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

Course Code: CSE 3015

Course Title: Natural Language Processing

Set number: 4

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) He drew one card.
- (ii) Mr Spock was charged with illegal alien recruitment.
- (iii) He crushed the key to my heart.
- (iv) Time flies like an arrow
- (v) John said he would come

Q2. (a) Design a program to remove rareword from a given string "Coca-Cola and Pepsi are fierce competitors, superior to Sprite" using Natural Language Tool Kit platform. Also, print the tokens before and after the removal of rareword. **[5M]**

(b) Differentiate between stemming and lemmatization. Provide proper theoretical and practical explanation to support the assertions. Based on Lancaster stemming find the result of the tokens listed as: **{cats, walked, swimming, better, quickest}** **[5M]**

Q3. (a) Explain the working procedure of Affix tagger and illustrate any two rules that is employed to tag a token. Provide proper explanation, snippet representation and reasons to support the explanation. **[5M]**

(b) Correct the following words using edit distance and Jaccard from nltk library compare the results and explain each step involved. **[5M]**

[soe comon challnges i natral language procesing, an hw can thy b adressed]

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]

India Independence Day is celebrated on August 15th each year to commemorate the nation's independence from British rule. On this day in 1947, after a prolonged struggle for freedom, India finally gained independence from British colonial rule, ending nearly 200 years of British domination.

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

5. Find the best sequence label using Viterbi Algorithm for the following phrase

“Nustin will stop will”

[10M]

Nustin/**noun** Justin/ **noun** can/ **model** watch//**verb** will/ **noun**

Spot/ **noun** will/ **model** watch /**verb** Nustin / **noun**

Will/ **model** Justin/ **noun** spot/ **verb** Nustin / **noun**

Nustin/ **noun** will **mode** pat/ **verb** Spot/ **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 |