Paper / Subject Code: 42175 / NATURAL LANGUAGE PROCESSING (DLOC - III)

BE SemVII (R2019, C Scheme) "Computer" Dec'2023

30/12/2023

Duration: 3hrs **[Max Marks: 80]**

N.B.: **(1) Question No 1 is Compulsory.**
(2) Attempt any three questions out of the remaining five.
(3) All questions carry equal marks.
(4) Assume suitable data, if required and state it clearly.

1 Attempt any FOUR **[20]**

a What is the rule-based and stochastic part of speech taggers?

b Explain Good Turing Discounting?

c Explain statistical approach for machine translation.

d Explain with suitable example the following relationships between word meanings: Hyponymy, Hypernymy, Meronymy, Holynymy

e What is reference resolution?

2 a Explain FSA for nouns and verbs. Also Design a Finite State Automata (FSA) for the words of English numbers 1-99. **[10]**

b Discuss the challenges in various stages of natural language processing. **[10]**

3 a Consider the following corpus **[10]**

<s> the/DT students/NN pass/V the/DT test/NN<\s>
<s> the/DT students/NN wait/V for/P the/DT result/NN<\s>
<s> teachers/NN test/V students/NN<\s>

Compute the emission and transition probabilities for a bigram HMM. Also decode the following sentence using Viterbi algorithm.

"The students wait for the test"

b What are five types of referring expressions? Explain with the help of example. **[10]**

4 a Explain dictionary-based approach (Lesk algorithm) for word sense disambiguation (WSD) with suitable example. **[10]**

b Explain the various challenges in POS tagging. **[10]**

5 a Explain Porter Stemming algorithm in detail. **[10]**

b Explain the use of Probabilistic Context Free Grammar (PCFG) in natural language processing with example. **[10]**

6 a Explain Question Answering system (QAS) in detail. **[10]**

b Explain how Conditional Random Field (CRF) is used for sequence labeling. **[10]**

****