M. Sc IT (Sem - VI) 09/05/2023

Paper / Subject Code: 92910 / Information Technology Natural Language Processing (R 2021)

(Time: $2\frac{1}{2}$ hours)

[Total Marks: 60]

N. B.: (1) All questions are compulsory.
(2) Make suitable assumptions wherever necessary and state the assumptions made.
(3) Answers to the same question must be written together.
(4) Numbers to the right indicate marks.
(5) Draw neat labeled diagrams wherever necessary.
(6) Use of Non-programmable calculator is allowed.

Q. 1 **Attempt *any two* of the following:** 12

a. Discuss the challenges of text preprocessing.
b. What is tokenization? Discuss the tokenization in Space-Delimited Languages.
c. List and explain the contextual factors that assist sentence segmentation in difficult cases.
d. Explain finite state transducer with example.

Q. 2 **Attempt *any two* of the following:** 12

a. Write a short note on Discourse Representation Theory.
b. What is context free grammar? Explain with suitable example.
c. How unary rules are handled in the CKY algorithm?
d. Write a short note on Natural language generation.

Q. 3 **Attempt *any two* of the following:** 12

a. What is part-of-speech? What are the difficulties of part-of-speech tagging?
b. What is the use of Hidden Markov Model? Explain the use of HMM for part of Speech Tagging.are the issues with it?
c. Explain any two translation model in detail.
d. Write a short note on Word Alignment.

Q. 4 **Attempt *any two* of the following:** 12

a. Discuss word processing with respect to Chinese Word.
b. What is probabilistic context free grammar (PCFG)? How PCFG is used as statistical parsing model?
c. Write a short note on MALT parser.
d. Discuss Local Discriminative Models.

Q. 5 **Attempt *any two* of the following:** 12

a. What is idiomaticity? Enlist and explain in detail the subtype of idiomaticity.
b. Write a short note on Multiword Expressions classification.
c. Explain any two methods for measuring the word similarity.
d. Discuss any one application of word sense disambiguation.
