



CBCS SCHEME

18CS743

Seventh Semester B.E. Degree Examination, Jan./Feb. 2023 Natural Language Processing

Time: 3 hrs.

Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Module-1

- 1 a. What is Natural Language Processing (NLP)? Explain different challenges of NLP with examples. (10 Marks)
- b. What are Karaka relations? Explain Karaka theory with example. (10 Marks)

OR

- 2 a. Explain 'n-gram' model. How data sparseness problem handled in 'n-gram' model. (10 Marks)
- b. Explain with example binding theory. (04 Marks)
- c. Consider the following training set:
 - i) The Arabian Knights </S>
 - ii) These are the fairy tales of the east </S>
 - iii) The stories of the Arabian Knights are translated in many languages </S>Find the probability of the following Test sentence using 'bi-gram' model.
"The Arabian Knights are the fairy tales of the east" (06 Marks)

Module-2

- 3 a. Explain minimum edit distance algorithm. Compute minimum edit distance between 'tumour' and 'tutor'. (10 Marks)
- b. List POS tagging methods. Explain Rule-based tagger, with example. (10 Marks)

OR

- 4 a. Explain Probabilistic CYK algorithm. List any two problems associated with PCFG. (10 Marks)
- b. Write a note on :
 - (i) Phrase level construction
 - (ii) Sentence level construction(10 Marks)

Module-3

- 5 a. Explain how relation pattern can be captured with a string Kernel. (10 Marks)
- b. Explain shortest path hypothesis with example. (10 Marks)

OR

- 6 a. Explain the strategies used in active learning approach. (10 Marks)
- b. Explain functional overview of Infact system with neat diagram. (10 Marks)

Module-4

- 7 a. With neat diagram, explain the evolutionary model for KDT (Knowledge Discovery from Text). (10 Marks)
- b. Explain word matching feedback system. (10 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8=50, will be treated as malpractice.

OR

- 8 a. Define the following :
(i) Structure (ii) Cohesion (iii) Interestingness (iv) Coherence (v) Coverage (10 Marks)
- b. Write a short note on :
(i) LSA
(ii) Sequence model (10 Marks)

Module-5

- 9 a. Explain six criteria that can be used for evaluation of IR (Information Retrieval) system. (10 Marks)
- b. Explain WORDNET and its applications. (10 Marks)

OR

- 10 Write a short note on :
a. Indexing.
b. Eliminating stop words.
c. Stemming.
d. Zipf's law. (20 Marks)
