



Seventh Semester B.E./B.Tech. Degree Examination, Dec.2025/Jan.2026
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 NLP? Explain various levels of NLP with example. Explain how a sentence, student of CSE like pizza is processed at various levels. (10 Marks)
b. Explain transformational grammar. Represent the following sentence in s-structure and d-structure. "Mukhesh was killed". (10 Marks)

OR

- 2 a. Explain n-gram model. What are the problems associated with n-gram model? How are these problems handled? (10 Marks)
b. Explain Karaka Theory with example. (10 Marks)

Module-2

- 3 a. What is Morphological Parser? Explain 2-step morphological parser with example. (10 Marks)
b. Write minimum edit distance algorithm. Compute the minimum edit distance between 'tutor' and 'tumour'. (10 Marks)

OR

- 4 a. What is POS Tagging? Explain different tagging methods. (10 Marks)
b. Explain disadvantages of PCFG. (05 Marks)
c. How unknown words can be handled in the tagging process. (05 Marks)

Module-3

- 5 a. Explain how the relation pattern can be captured with a string kernel. (10 Marks)
b. Explain dependency path kernel for relation extraction with example. (10 Marks)

OR

- 6 a. With neat diagram explain learning frame work architecture. (10 Marks)
b. Explain the strategies used in active learning approach for acquiring labels using committee based classification schema. (10 Marks)

Module-4

- 7 a. Define the following : (10 Marks)
i) Structure
ii) Cohesion
iii) Interestingness
iv) Coherence
v) Coverage.
b. With neat diagram, explain the evolutionary model for knowledge discovery from texts. (10 Marks)

OR

- 8 a. What is Parsing? Explain different types of parser. (10 Marks)
b. Explain various approaches to analyzing texts. (10 Marks)

Module-5

- 9 a. What is IR? Explain the design features of IR with neat diagram. (10 Marks)
b. Consider a document represented by the five terms {tornado, swirl, wind, desert, water} With raw tf (termed frequency) 4, 1, 1, 2 and 3 respectively. In a collection of 100 documents, 15 documents contains the term tornado, 20 contains swirl, 40 contains wind, 25 contains desert and 30 contains water. Calculate idf (inverse document frequency) and weight of each term. (10 Marks)

CMRIT LIBRARY BANGALORE - 560 037

OR

- 10 Write a note on : (20 Marks)
a. WORDNET
b. FRAMENET
c. STEMMER
d. POS tagger.

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.