

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2024

(B. Tech) Program: Computer Engineering Scheme II
Regular Examination: LY Semester: VII

Course Code: CEDLC7034 and Course Name: Natural Language Processing
Duration: 02.5 Hours
Max. Marks: 60

Date of Exam: 21/11/2024

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

Q. No.	Question	Max. Marks	CO	BT level
Q 1	Solve any two questions out of three: (05 marks each)	10		
a)	List all bi-grams and compute the first 5 bigram probabilities for the sentence given below. Explain how spelling correction is done using N-grams. Sentence: <s> the students pass the test </s>, <s> the students wait for the pass </s>, <s> teachers test students </s>		CO2	Ap
b)	Describe the difference between inflectional and derivational morphology with examples		CO2	U
c)	Using the following CFG, Parse the sentence "Book the flight" S → NP VP S → VP NP → Det N NP → Det Adj N VP → V VP → V NP		CO3	Ap
Q 2	Solve any two questions out of three: (05 marks each)	10		
a)	Explain the concept of polysemy and how it differs from homonymy. Provide examples to illustrate your answer.		CO4	U
b)	i) Analyze the following sentence for structural ambiguity: • "Time flies like an arrow." Explain how different structural interpretations lead to different meanings. ii) Analyze the following sentence for pragmatic ambiguity: • "He crushed the key to my heart." Explain how context impacts the interpretation.		CO5	An

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

Nov – Dec 2024	
(B. Tech) Program: Computer Engineering Scheme II	
Regular Examination: LY Semester: VII	
Course Code: CEDLC7034 and Course Name: Natural Language Processing	
Date of Exam: 21/11/2024	Duration: 02.5 Hours
Max. Marks: 60	

c)	Write a note on Attachment for fragment of English- sentences		CO4	U
Q.3	Solve any two questions out of three. (10 marks each)	20		
a)	i) Discuss the stages in NLP from text preprocessing to semantic analysis, explaining the main function of each stage. (6 marks) ii) Provide examples to illustrate how ambiguity can affect the output at the syntactic and semantic levels. (4 marks)		CO1	U, Ap
b)	i) Explain how sequence labeling is applied in POS tagging and the use of Hidden Markov Models (HMM) in this process. (5 marks) ii) Describe the challenges associated with POS tagging, such as handling unknown words and multiple possible tags. (5 marks)		CO3	U
c)	i) Explain the phenomenon of co-reference resolution and why it is critical for discourse understanding. (5 marks) ii) Discuss the syntactic and semantic constraints that help in resolving co-references in text. Provide examples to support your answer. (5 marks)		CO5	U
Q.4	Solve any two questions out of three. (10 marks each)	20		
a)	i) Define morphological parsing, and explain how is it performed using finite state transducers (FSTs)? (6 marks) ii) Explain the difference between lemmatization and stemming with examples. (4 marks)		CO2	U
b)	i) Describe Word Sense Disambiguation (WSD) and its importance in NLP. (5 marks) ii) Apply dictionary-based method for WSD on any suitable data and discuss its limitations. (5 marks)		CO4	Ap
c)	i) Describe Machine Translation as an application of NLP(4 marks). ii) Discuss one method used for machine translation (e.g., statistical or neural translation). (3 marks) iii) comments on “challenges associated with translating languages with different grammatical structures” (3 marks)		CO6	U
