B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

Fifth Semester

Artificial Intelligence and Data Science

(Common to Computer Science and Engineering (AIML))

20AIPC503 - NATURAL LANGUAGE PROCESSING AND CHAT BOT

Regulations - 2020

Duration: 3 Hours   Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)
Answer ALL Questions

		Marks	K–Level	CO
1.	Which of the following is an application of NLP?	1	K1	CO1
	(a) Image compression (b) Machine translation (c) Circuit design (d) Data encryption			
2.	What is meant by Tokenization in NLP ?	1	K1	CO1
	(a) Dividing text into sentences or words (b) Removing duplicate data			
	(c) Translating a document (d) Generating new tokens for authentication			
3.	Choose an option based on which atrigram model predicts the next word?	1	K1	CO2
	(a) Previous one word (b) Previous two words			
	(c) Entire sentence (d) Random selection			
4.	Laplace smoothing is used to _____	1	K1	CO2
	(a) Increase memory storage (b) Avoid zero probabilities			
	(c) Improve tokenization (d) Accelerate computation			
5.	Which model predicts context words from a centre word?	1	K1	CO3
	(a) Skip-gram (b) CBOW (c) LDA (d) PCA			
6.	Select an advantage of word embedding over one-hot encoding from the given options.	1	K1	CO3
	(a) Fixed dimensionality (b) Captures semantic relationships			
	(c) Simpler computation (d) No need for training			
7.	Where does a top-down parser begin parsing from?	1	K1	CO4
	(a) Leaf nodes (b) Start symbol (c) Bottom words (d) POS tags			
8.	What is meant by unification in parsing?	1	K1	CO4
	(a) Resolving feature mismatches (b) Tokenizing text			
	(c) Generating embedding (d) Removing ambiguity			
9.	Choose the correct option that represents a key advantage of Neural MT over SMT?	1	K1	CO5
	(a) Simpler computation (b) Captures long-range dependencies			
	(c) Requires no training data (d) Avoids ambiguity			
10.	Which of the following is an ethical concern in chatbot use?	1	K1	CO6
	(a) User privacy and data security (b) Parsing complexity			
	(c) Low BLEU score (d) Grammar inconsistencies			

PART - B (12 × 2 = 24 Marks)
Answer ALL Questions

11.	Define lemmatization with an example.	2	K1	CO1
12.	Write the features of NLTK and SpaCy library.	2	K1	CO1
13.	Define perplexity in language modelling.	2	K1	CO2
14.	Interpret is the purpose of smoothing in language models.	2	K2	CO2
15.	Recall word sense disambiguation (WSD) and its types.	2	K1	CO3
16.	Explain Lesk algorithm and its significance.	2	K2	CO3
17.	Compare Top-Down and Bottom up parsing techniques.	2	K2	CO4
18.	Write the limitations of CFG in NLP.	2	K1	CO4
19.	Why is SMT not effective for long sentences?	2	K1	CO5

K1 – Remember; K2 – Understand; K3 – Apply; K4 – Analyze; K5 – Evaluate; K6 – Create   **13978**

20. How does BLEU score help in evaluating MT systems?	2	K1	CO5
21. What is the role of Rasa framework?	2	K1	CO6
22. Define user intent in chatbot systems.	2	K1	CO6

PART - C (6 × 11 = 66 Marks)
Answer ALL Questions

23.	a)	Explain in detail the steps involved in the normalization of text in NLP with use cases.	11	K2	CO1
		OR			
	b)	Interpret the functions of at least seven Python libraries used for Natural Language Processing	11	K2	CO1
24.	a)	Explain statistical N-gram language models in Natural Language Processing (NLP) and discuss how these models are evaluated using entropy and perplexity.	11	K2	CO2
		OR			
	b)	Describe any four techniques that redistribute probability mass from seen words to unseen ones, preventing zero-probability estimates and improving the reliability of models.	11	K2	CO2
25.	a)	Explain CBOW model with neat diagram. Compare CBOW and Skip-gram approaches.	11	K2	CO3
		OR			
	b)	Compare and contrast the contextual and non- contextual embedding techniques in NLP. Outline the importance of ELMo and BERT in embedding techniques.	11	K2	CO3
26.	a)	Apply the rules for converting a Context-Free Grammar (CFG) into Chomsky Normal Form (CNF) and determine which of the given grammar sets (A, B, or C) satisfies the CNF conditions, justifying your answer.	11	K3	CO4

Set A:	Set B:	Set C:
S→AB|aB	S→AB|BC	S→ASA|aB
A→a|b	A→BA|a	A→S|B
B→b	B→CC|b	B→b
	C→AB|a	

		OR			
	b)	Apply the rules of a Probabilistic Context-Free Grammar (PCFG)tocalculate the probability of a parse tree for a given sentence --- "The cat sleeps".	11	K3	CO4
27.	a)	Explain the architecture and working of the Statistical and Neural Machine Translation techniques.	11	K2	CO5
		OR			
	b)	Explain the attention mechanism in the transformer model, including the concepts of query, key and value vectors.	11	K2	CO5
28.	a)	Build a conversational AI agent for a travel booking platform that can assist users with flight, hotel and car rental bookings. Explain the key features, functionalities, and benefits of the design.	11	K3	CO6
		OR			
	b)	Apply your understanding of conversational AI platforms to analyze and differentiate between Microsoft Bot Framework, Amazon Lex, and Google Dialogflow in terms of their features, strengths, weaknesses, and use cases.	11	K3	CO6