Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science & Engineering
Mid Examination, Summer 2025
Course Code: CSE445, Course Title: Natural Language Processing
Level:3 Term: 3 Batch: 62

Time: 01:30 Hours **Marks: 25**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1.	*a)*	Natural Language Processing (NLP) systems have achieved remarkable progress with the advent of deep learning architectures; however, they still fall short of achieving comprehensive language understanding. i) Define and illustrate the types of linguistic ambiguity (lexical, syntactic, semantic, and pragmatic) and explain their impact on NLP system performance. ii) Analyze the syntactic and semantic variability across languages and domains. How do these variations affect the generalizability and adaptability of NLP models?	3	CO1
	b)	You are developing a Named Entity Recognition (NER) model to extract key entities (e.g., organizations, locations, person names, and financial terms) from informal, user-generated financial texts such as forum posts and social media updates. These texts are noisy, often unstructured, and contain inconsistent capitalization, misspellings, abbreviations, and domain-specific terminology. Critically evaluate how each of the following preprocessing steps may help or hinder the performance of your NER model: • Lowercasing • Stop-word removal • Lemmatization or stemming • Tokenization (subword vs. word-level) • Handling misspellings and abbreviations • Removing special characters and hashtags Which preprocessing choices would you avoid, customize, or enhance specifically for the NER task? Justify your decisions with respect to preserving entity boundaries, context, and semantic fidelity. Provide examples where applicable.	2	
2.	*a)*	Apply Consistency parsing for the following Sentence "The quick brown fox jumps over the lazy dog near the river." Where CFG S → NP VP NP → Det AdjP N AdjP → Adj Adj | Adj VP → V PP PP PP → P NP	3	CO2
	b)	Between Consistency and Dependency parsing which one you will depend to achieve your goal.	2	

3.	a)	Given the following three documents: Doc1: "apple banana apple" Doc2: "banana fruit banana apple" Doc3: "fruit apple orange orange" Calculate the TF-IDF values for the terms "apple," "banana," "fruit," and "orange" in each document.	5	CO3
	b)	In the context of building a text classification model for sentiment analysis, explain how overfitting and underfitting can impact model performance. Discuss the causes of each, how they might manifest in NLP tasks, and propose methods to detect and mitigate overfitting and underfitting during model training.	5	
4		Given the following training documents and their sentiment classes: **Document** **Text** **Class** D1 "love this movie" Positive D2 "hate this movie" Negative D3 "this movie is great" Positive D4 "this movie is bad" Negative Given the sentence "this movie is great", compute the posterior probabilities for both classes using the Naive Bayes formula. Based on your calculations, classify the sentence as Positive or Negative.	5	