ourdiu.com | Shared by Md. Ferdous Foysal Hridoy

Daffodil International University

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Midterm Examination, Fall-2023

Course Code: CSE445, Course Title: Natural LanguageProcessing

Level: 4 Term: 1, 2 Batch: 56, 57

Time: 1 Hour and 30 Minutes **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)***	Define Natural Language Processing. Write down the major areas of research and development in NLP.	**3**	**CO1**
	b)	Write down the regular expression based on the following scenario: "First match the word Column, then followed by a number and optional spaces, the whole pattern repeated any number of times"	**2**	
2.	***a)***	Describe the difference between lemmatization and stemming with proper examples.	**3**	**CO2**
	b)	Analyze the best suitable algorithm for calculating minimum edit distance to convert "EDITING" to "DISTANCE" (using insertion cost = 1, deletion cost = 1 and substitution cost = 2).	**5**	
3.	***a)***	Analyze Naive Bayes Theorem to predict the category of this testing data: Bangladesh cricket team get excellent victory	**7**	**CO2**

words	**class**
Bangladesh cricket team victory	Pos
~~Cricket~~ matches Bangladesh	Pos
Cricket players skillful	Pos
Proud of Bangladesh team	Pos
Bangladesh cricket match defeated	Neg
Bangladesh ~~cricket~~ team get struggled	Neg
Excellent batting Bangladesh team	Pos
Injuries in Bangladesh ~~cricket~~ team	Neg

4.	***a)***	Explain shortly about "discounting". Mention the problems regarding HMM and give proper solutions.	**2+3**	**CO2**