


Name:			
Enrolment No:			
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES End Semester Examination, December 2022			
Course: Natural Language Processing Program: MCA Course Code: CSAI 8004P		Semester: III Time : 03 hrs. Max. Marks: 100	
Instructions: All questions are compulsory. Question no. 9 of Section B and Question no. 11 of Section C have internal choice.			
SECTION A (5Qx4M=20Marks)			
S. No.		Marks	CO
Q 1	“Rice flies like sand”. Which type of ambiguity exists in this sentence?	4	CO1
Q 2	Write POS tags probability for sequence “NLP is very interesting subject”.	4	CO2
Q 3	‘Wagen’ can be translated into ‘Wagon’ or ‘Vehicle’. How probability-based machine translation is used to translate it.	4	CO3
Q 4	How semantic representation is used for NLU?	4	CO4
Q 5	Write various Question answering platforms.	4	CO5
SECTION B (4Qx10M= 40 Marks)			
Q 6	“Pre-processing of text is required in NLP to remove stop-words, punctuations. It is used for tokenization, stemming, lemmatization, and POS tagging”. Perform NLP phases on this set of sentences.	10	CO1
Q 7	How statistical machine translation is different from neural machine translation? Which is most commonly used these days by professionals and organizations.	10	CO3
Q 8	“I prefer a morning flight”. Verify the correctness of this sentence using CKY parser.	10	CO4
Q 9	How NLP techniques are applied in text summarization? or How NLP techniques are applied in sentiment analysis?	10	CO5
SECTION-C (2Qx20M=40 Marks)			

<p>Q 10</p>	<p>How probabilistic based grammar is used for resolving ambiguity?</p> <p>If $P(\Phi \text{Art})=0.7$, $P(\text{N} \text{Art})=1$, $P(\text{V} \text{N})=0.4$, $P(\text{N} \text{V})=0.3$, $P(\Phi \text{N})=0.30$, $P(\text{Art} \text{V})=0.7$, $P(\Phi \text{V})=0.01$, $P(\text{P} \text{V})=0.1$, $P(\text{V} \text{V})=0.1$, $P(\text{Art} \text{P})=0.2$</p> <p>and occurrences of terms are as follows.</p> <table border="1" data-bbox="240 558 1146 787"> <thead> <tr> <th></th> <th>N</th> <th>V</th> <th>ART</th> <th>P</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Flies</td> <td>37</td> <td>30</td> <td>0</td> <td>0</td> <td>67</td> </tr> <tr> <td>like</td> <td>5</td> <td>25</td> <td>0</td> <td>12</td> <td>42</td> </tr> <tr> <td>A</td> <td>2</td> <td>0</td> <td>30</td> <td>0</td> <td>30</td> </tr> <tr> <td>flower</td> <td>50</td> <td>10</td> <td>0</td> <td>0</td> <td>60</td> </tr> <tr> <td>Total</td> <td>94</td> <td>65</td> <td>30</td> <td>12</td> <td>199</td> </tr> </tbody> </table> <p>How ambiguity in sentence “ Flies like a flower” can be resolved using Probabilistic grammar.</p>		N	V	ART	P	Total	Flies	37	30	0	0	67	like	5	25	0	12	42	A	2	0	30	0	30	flower	50	10	0	0	60	Total	94	65	30	12	199	<p>20</p>	<p>CO4</p>
	N	V	ART	P	Total																																		
Flies	37	30	0	0	67																																		
like	5	25	0	12	42																																		
A	2	0	30	0	30																																		
flower	50	10	0	0	60																																		
Total	94	65	30	12	199																																		
<p>Q 11</p>	<p>Describe various characteristics of corpus? How corpus is different from dataset. Describe most commonly used corpus in NLP. Write code to calculate similarity between two words using Wordnet and synsets in NLTK.</p> <p style="text-align: center;">Or</p> <p>Write Penn TreeBank symbols for POS tagging and their detailed explanation and examples. Write code for tokenization and POS tagging using NLTK.</p>	<p>20</p>	<p>CO2</p>																																				