|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

****

**Presidency University**

**Bengaluru**

**School Of Computer Science and Engineering & Information Science**

**Summer Term End-Term Examinations, Aug 2024**

**Date**: 07.08.2024

**Time**: 01:00 pm to 04:00 pm

**Max Marks**: 100

**Weightage**: 50%

**Summer Term**: 2023 - 24

**Course Code**: CSE3014

**Course Name**: Fundamentals of Natural Language Processing

**Department: School of CSE & IS**

**Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *Do not write any matter on the question paper other than roll number.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q.No** | **Questions** | **Marks** | **CO** | **RBT** |
| 1 | 1. What is NLP? List and Explain applications of NLP. | 4 | CO1 | L1 |
| 1. Identify the number of bigram and Trigram in the following sentence.   A big bug bites a little beetle but the little beetle bit the big bug back. | 6 | CO1 | L2 |
| 1. List and explain different phases of analysis in Natural Language Processing with an example for each. | 10 | CO1 | L3 |
| OR | | | | |
| 2 | 1. Out of the following evaluation metrics, select the six that CAN be used for part-of-speech tagging: (a) Accuracy, (b) BLEU, (c) Cohen's Weighted Kappa, (d) Cohen's Unweighted Kappa, (e) Classification Error, (f) F1-Score, (g) Precision, and (h) Recall. NOTE: If you write any wrong evaluation metric, you will be getting 0 marks for this question. | 4 | CO1 | L1 |
| 1. Find one tagging error in each of the following sentences that are tagged with   the Penn Treebank tagset:  1. I/PRP need/VBP a/DT flight/NN from/IN Atlanta/NN  2. Does/VBZ this/DT flight/NN serve/VB dinner/NNS  3. I/PRP have/VB a/DT friend/NN living/VBG in/IN Denver/NNP  4. Can/VBPyou/PRPlist/VBthe/DTnonstop/JJafternoon/NNflights/NNS | 6 | CO1 | L2 |
| 1. Explain Different types of document representation in nlp. | 10 | CO1 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3 | 1. Stopwords are words which are very frequently used in NLP. Consider a situation where weigh the counts of words by their tf-idf values. Mention the value of a the weighted count (weighted by the product of the tf and the idf) of a stop word, that is present in all the documents of a corpus. | 4 | CO2 | L1 |
| 1. Assume that we are using a small, 26-dimension vector to represent our words, such that each dimension represents the count of the character (from a to z) of our words. Eg. "sandeep" = [1, 0, 0, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]. For each word pair, compute the dot product and cosine similarity.  * word1 = sitting, word2 = kitten * word1 = donkey, word2 = money * word1 = grain, word2 = grail * word1 = table, word2 = stall * word1 = hello, word2 = helm | 6 | CO2 | L2 |
| 1. Find the cosine similarity of cos(cherry, information) using the following data.  |  |  |  |  | | --- | --- | --- | --- | |  | **pie** | **data** | **computer** | | cherry | 442 | 8 | 2 | | digital | 5 | 1683 | 1670 | | information | 5 | 3982 | 3325 | | 10 | CO2 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 4 | 1. Mention the term which describes the number of documents in a corpus that a particular token is present in. | 4 | CO2 | L1 |
| 1. What is POS Tagging? POS tag the following sentence: “Ram went to school after two months”. | 6 | CO2 | L2 |
| 1. Discuss about the evaluation metrics of Machine Translation. | 10 | CO2 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5 | 1. Expand GPE in the context of named entity recognition | 4 | CO3 | L1 |
| 1. Write a short note on 1)Sentiment Analysis 2)Machine Translation | 6 | CO3 | L2 |
| 1. Write the algorithm for Minimum edit distance and compute the minimum edit distance between Intention and Execution. | 10 | CO3 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 6 | 1. Explain sequence labelling. | 4 | CO3 | L1 |
| 1. Explain the TF-IDF algorithm in NLP. Find TF-IDF vectors use log (with base 10) for calculating IDF.   D1=The best team plays the finals  D2=India won a medal in the finals. | 6 | CO3 | L2 |
| 1. Discuss about text classification and its applications. | 10 | CO3 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 7 | 1. List any two activation functions, their formulae and the range of values that they take. | 4 | CO4 | L1 |
| 1. Consider a sentiment analysis classifer that classifies texts into 3 classes - positive, negative, and neutral. The results of the classification are as follows in the given confusion matrix.   Confusion Matrix for 300 documents, of which 100 documents are positive, 100 documents are neutral and 100 documents are negative.    Assuming that each class actually has 100 documents, calculate the accuracy of the classifier, as well as the precision, recall, and F1-scores of all 3 classes. | 6 | CO4 | L2 |
| 1. Consider 2 strings x, y. Let x = “sandeep” and y = “albert”. Let the value of n (size of maximum n-grams) = 3. Calculate HISK(x,y). | 10 | CO4 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 8 | 1. Morphological segmentation involves splitting a word into individual units. Mention the name of those units. | 4 | CO4 | L1 |
| 1. Given an observation matrix O of size n\*n, write the expression to calculate the value of the ith row and jth column (i.e. E[i][j]) in the expectation matrix. | 6 | CO4 | L2 |
| 1. Consider the following documents (Yes, each bullet point is a document):  * Principles of Artificial Intelligence * Artificial Intelligence for Gaming * Artificial Intelligence and Machine Learning * Artificial Intelligence for Game Development   Assume only the following **terms:**   * Principles * Artificial * Intelligence * Gaming * Machine * Learning * Game * Development   Write down the raw counts matrix, and generate the TF-IDF matrix, whose elements are weighted by the product of the TF and the IDF. Consider that the logarithm we are using is in base 10. | 10 | CO4 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 9 | 1. Identify the morphological type (Noun phrase, Verb Phrase, Adjective Phrase) of following sentence segments   1. important to Bill  2. looked up the tree | 4 | CO1 | L1 |
| 1. Identify and describe the ambiguities in the following sentences.   i. The man kept the dog in the house.  ii. Book that flight. | 6 | CO1 | L2 |
| 1. Write a detailed note on Part of Speech Tagging with an example. | 10 | CO1 | L3 |

OR

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 10 | 1. What is meant by the semantics of a natural language, and how this differs from   the pragmatics? | 4 | CO2 | L1 |
| 1. Perform parsing using simple top-down parsing for the sentence “The dogs cried” using the grammar given below:   S->NP VP  NP->ART N  NP->ART ADJ N  VP->V  VP->V NP | 6 | CO2 | L2 |
| 1. Calculate the BLEU score between the following pair of sentences. Consider that we use only unigrams and bigrams, with weights of 0.5 each (i.e. no trigrams or 4-grams).   Candidate: the dog dog on the grass grass  Reference: the dog is on the grass | 10 | CO2 | L3 |