UPES
UNIVERSITY OF TOMORROW

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2022

Course: B.Tech (CSE + AI&ML) **Semester: 6th**
Program: Computational linguistics and NLP **Time: 03 hrs.**
Course Code: CSEG 3024 **Max. Marks: 100**

Instructions:

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	a. Consider the statement " The students went to class" . Assign POS tags for the statement. b. *Mujhe khaanna khaanna hai.* What will be tag of third word in the given sentence?	4 (3+1)	CO1
Q2	Write a regular expression to find all instances of the determiner "the":	4	CO1
Q3	Using Porter's stemming technique, find out the stem of the following words: (i) automobile (ii) automotive (iii) information (iv) informative	4	CO2
Q4	Given a document containing terms with given frequencies: A(3),B(2),C(1) Assume collection contains10,000 documents and document frequencies of these terms are: A(50),B(1300),C(250) **Find out the tf-idf value of all the terms.**	4	CO3
Q5	Differentiate text analysis, text mining, and text analytics.	4	CO1

SECTION B
(4Qx10M= 40 Marks)

Q6	There are different levels of understanding in the linguistics of any language. Discuss all the levels in details.	10	CO2
Q7	In vector space retrieval model, consider the following three sentences; d1: "new york times" d2: "new york post" d3: "los angeles times" According to the cosine similarity values, find out the final order in which the documents are presented (where total number of documents	10	CO3

	N is 3).		
Q7	CYK Parser plays an important role in NLP. Justify the statement with the help of a suitable example.	**10**	**CO2**
Q8	Consider the following example to convert CFG to CNF: S → ASB $\qquad$ A → aAS|a|ε $\qquad$ B → SbS|A|bb or Illustrate the different classes of Morphology with suitable examples.	**10**	**CO1**
SECTION-C **(2Qx20M=40 Marks)**			
Q9	Sentiment analysis is an active sub-field of NLP. Explain the full working process of sentiment analysis and discuss the different features extraction techniques.	**20**	**CO2**
Q10	Write short notes with example on the following:- a) TF-IDF $\quad$ b) Bag of Words c) n-gram model d) Word Sense Disambiguation **or** **"The students are learning the Natural Language Processing course."** Apply the Hidden Markov Model in the above-mentioned sentence to identify Parts of Speech Tagging. Illustrate the each steps of the algorithm	**20**	**CO3**