

**UNIVERSITY OF PETROLEUM AND ENERGY
STUDIES**

End Semester Examination, December 2021

Course: Computational Linguistic & NLP
Program: BTech CSE OSSOS
Course Code: CSEG 3024

Semester : V
Duration : 03 hrs.
Max. Marks: 100

Instructions: Answer all questions from Sec. A, B & C.

SECTION A

(5Qx 4M = 20 Marks)

		Marks	COs
Q 1	Provide 4 applications for NLP.	4	CO1
Q 2	Difference in Information Retrieval and Information Extraction systems.	4	CO1
Q 3	What are the Features used for Text Classifications?	4	CO2
Q 4	“They fish in the river. In December there is no fish, as fishing is already over.” Tag the sentence using POS.	4	CO2
Q 5	Select the correct options and justify your answer. a. Constituency parsing is based on grammar rules. b. Dependency parsing is based on word relationship. c. Cosine similarity is maximum for 90 degree angle between documents.	4	CO1

SECTION B

(Answer in 500-800 words/1-1.5 pages)

(4Qx10M = 40 Marks)

Q 6	Discuss TFIDF with Example.(Formula & Design)	10	CO2
Q 7	What is Sentiment Analysis? Why is it useful, discuss with use case.	10	CO1
Q 8	What are Text Similarity measures and uses? Discuss 3 methods to find text similarity.	10	CO1
Q 9	Explain use of POS tagging in NLP. What are different symbols used in tagging, discuss with an example. Or Discuss Naïve Bayes Classification Algorithm. Provide working Principles.	10	CO2

SECTION-C

(Answer in 1200-1500 words/2-3 pages)

(2Qx 20M= 40 Marks)

Q 10	Explain NLP pipeline for text classification with example.(Data injection, preprocessing, feature set, any method for text classification using feature set)	20	CO1
Q 11	What is Topic Modeling? Explain Latent Dirichlet Allocation (LDiA) as a topic modeling method. How Linear Discrimination Analysis (LDA) works? Discuss use case difference of LDiA vs LDA. Or Text = [‘The Stop words are removed after the process of Tokenization’, ‘Using Lemmatization, we can find the root of a word’, ‘The preprocessing and feature selection are important steps, for using cleaned pipeline.’, ‘Natural Language Processing covers modern aspects of machine translations, and also functional of catboats.’] Perform (i) Tokenization, (ii) Remove the Stop words, (iii) Lemmatization, (iv) Stemming, (v) Bigrams, (vi) POS tags Differentiate between Lemmatization and Stemming. Discuss with example what Hypernym & Hyponym are.	20	CO2