(ii) Given the sentence:

"*Steve Jobs founded Apple in California.*"

Identify and classify named entities in this sentence.

[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 7292 **L**

Unique Paper Code : 2343010029

Name of the Paper : Natural Language Processing

Name of the Course : **B.Sc. (H) / B.A. (P) DSE / BSc. (P)**

Semester : VIII

Duration : 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Question No. **1 (Section A)** is compulsory.
3. Attempt any **4 (four)** questions from **Section B**.
4. Parts of a question should be attempted together.

Section A

1. (a) Explain the three different editing operations allowed in Minimum Edit Distance Calculation. (3)

(b) How is a Recurrent Neural Network different from a Feed Forward Neural Network? (3)

(c) Explain how Word2Vec model is used for computing embeddings? (3)

(d) Describe the purpose of any two of the following methods offered by the 're module': (4)

(i) re.match()

(ii) re.sub()

(iii) re.finditer ()

(e) Differentiate between Static and Contextual embeddings. (4)

(c) What is the role of attention in the transformer model? Explain the three matrices using which the attention head computes the attention vector. (5)

7. (a) Compare : (5)

(i) Parallelized vs Sequential attention computation

(ii) Autoregressive models and Bidirectional models

(b) What is the XOR problem? Why can it not be solved using a single-layer perceptron? Explain the role of non-linear activation functions in solving the XOR problem. (5)

(c) (i) Given the sentence: (5)

The quick brown fox jumps over the lazy dog."

Perform Part of Speech POS tagging for each word.

(b) Why are word embeddings needed in NLP? Differentiate between one-hot encoding vectors and word embedding vectors. (5)

(c) Why is cosine similarity used in NLP? Compute cosine similarity for two vectors.

A =[1, 2, 3], B = [2, 3, 4] (5)

5. (a) Give advantages and disadvantages of the Bag of Words model for representing text documents. (5)

(b) What is the utility of residual stream in transformer model? (5)

(c) Briefly explain any two ethical and any three safety issues related to modem language models. (5)

6. (a) What is temperature sampling? How is it different from greedy sampling? (5)

(b) Explain any five preprocessing steps in NLP. (5)

(f) Why is lemmatization more accurate in NLP than stemming? Justify using an example. (4)

(g) Write regex patterns for the following: (4)

(i) 10-digit phone number

(ii) Only alphabets

(h) Apply Laplace Smoothing to compute the smoothed unigram probabilities for the vocabulary words given for the following data: (5)

Vocabulary = {‘Mouse’, ‘Keyboard’, ‘computer’},

Total training words in the vocabulary is 5,

Counts: ‘Mouse’ = 3, ‘Keyboard’ = 2,’computer’ = 0.

What happens if smoothing is not used in n-gram language modelling?

Section B

2. (a) Name the three main architectures used in Transformer-based language models and briefly describe each. (5)

(b) Apply Byte Pair Encoding (BPE) on the word: (5)

"low low lower"

Assume initial symbols are split into characters and </w> marks end of word:

l o w </w>

l o w </w>

l o w e r </w>

(c) What is prompting in large language models? Explain any two prompt engineering techniques with examples. (5)

3. (a) A document D has 100 words. Find which term is more important according to its TF-IDF.

Term	Count of terms in document D	Number of document containing term (Out of 10)
A	10	5
B	5	1

What happens if a term appears in all documents? Justify your answer. (5)

(b) What is the purpose of an activation function? Explain the following activation functions using their formulae : (10)

- Sigmoid
- Tanh
- ReLU
- Softmax

4. (a) What is the n-gram language modelling technique in NLP? Generate Unigrams, Bigrams and trigrams for the sentence 'I love my Country India'. (5)