# 315329

**12526**

**3 Hours / 70 Marks** Seat No. | | | | | | | | |

*Instructions* – (1) All Questions are *Compulsory.*

(2) Answer each next main Question on a new page.

(3) Illustrate your answer with neat sketches wherever necessary.

(4) Figures to the right indicate full marks.

(5) Assume suitable data, if necessary.

(6) Mobile Phone, Pager and any other Electronic Communication devices are not permissible in Examination Hall.

**Marks**

**1. Attempt any FIVE of the following:** **10**

a) Define Natural Language Processing (NLP).

b) Define regular expressions.

c) Write a regular expression to match a 10-digit phone number.

d) Mention any two challenges faced while using Transformers.

e) Define One-Hot Encoding.

f) Define Named Entity Recognition (NER).

g) Explain the concept of a semantic network in NLP.

P.T.O.

**Marks**

**2. Attempt any THREE of the following:** **12**

a) Define the following with one example each :–

i) Synonymy

ii) Antonymy

iii) Hyponymy

iv) Polysemy

b) Differentiate between stemming and lemmatization.

c) Explain how Hugging Face Transformers can be integrated with custom datasets for fine-tuning a sentiment analysis model. Include the role of tokenizers and Trainer API analysis model. Include the role of tokenizers and Trainer API.

d) Perform POS tagging on the sentence: 'Sachin Tendulkar played cricket for India'.

**3. Attempt any THREE of the following:** **12**

a) List and briefly explain any four real-world applications of NLP.

b) Explain dependancy parsing with example.

c) Explain Corpora Annotation and Utilities with example.

d) Define the BIO (or IOB) tagging format used in Named Entity Recognition (NER). Explain with a suitable example.

**4. Attempt any THREE of the following:** **12**

a) Explain different types of word order typologies used in natural languages. Give one example for each type.

b) Explain transfer learning in NLP. Mention its benefits.

c) Explain the process of chunking with examples.

d) Differentiate between Bag-of-Words and Word2Vec.

e) Explain Byte Pair Encoding (BPE) in tokenization

**Marks**

**5. Attempt any TWO of the following:** **12**

a) Explain Transformer architecture with encoder-decoder diagram and working.

b) Define stemming in NLP. Explain the working of the Porter Stemmer algorithm with the help of suitable examples. Stem the following words using Porter Stemmer: "running," "happily," "cats," "flying".

c) Given the following term frequency vectors for two documents:–

- Document A: [3, 0, 2, 5, 1]
- Document B: [1, 2, 0, 3, 4]

i) Calculate the cosine similarity between Document A and Document B.

ii) Interpret the result in terms of document similarity in an NLP context.

**6. Attempt any TWO of the following:** **12**

a) Describe Part-of-Speech (POS) tagging. Discuss at least four major challenges in POS tagging with suitable examples.

b) Explain vector space model with respect to words and vectors.

c) Explain the concept of N-grams vectors with types. Give examples for unigram, bigram, and trigram.

---