

Total No. of Questions : 8]

SEAT No. :

P6566

[Total No. of Pages : 2

[6181]-116

B.E. (Computer Engineering)
NATURAL LANGUAGE PROCESSING
(2019 Pattern) (Semester - VIII) (410252A) (Elective - V)

Time : 2½ Hours]

[Max. Marks : 70

Instructions to the candidates:

- 1) All questions are compulsory.
- 2) Figures to the right indicate full marks.

- Q1)** a) What are generative models of language? Explain any one model in detail. [4]
b) Consider the following small corpus: [8]

Training corpus:

<s> I am from Pune </s>

<s> I am a teacher </s>

<s> students are good and are from various cities </s>

<s> students from Pune do engineering </s>

Test data:

<s> students are from Pune </s>

Find the Bigram probability of the given test sentence.

- c) Explain in detail Latent Semantic Analysis for topic modelling (LSA). [6]

OR

- Q2)** a) Write short note on BERT. [4]
b) Given a document-term matrix with the following counts: [6]

	Document 1	Document 2	Document 3
Term 1	10	5	0
Term 2	2	0	8
Term 3	1	3	6

Calculate the TF-IDF score of "Term 1" in "Document 1".

- c) Describe the Latent Dirichlet Allocation (LDA) algorithm and how it is used for topic modeling? [8]

P.T.O.

- Q3)** a) Describe the concept of Information Retrieval. Explain the significance of Natural Language Processing in Information Retrieval. [4]
b) Explain reference resolution and coreference resolution with example. [8]
c) What is Cross-Lingual information Retrieval, and how is it used in Natural Language Processing? Provide an example. [6]

OR

- Q4)** a) Explain the concept of the Vector Space Model, and describe how it is used in Information Retrieval. [6]
b) Describe entity extraction and relation extraction with the help of examples. [8]
c) What is Named Entity Recognition (NER)? Describe the various metrics used for evaluation. [4]
- Q5)** a) List the tools available for the development of NLP applications? Write the features of any 3 tools. [7]
b) Describe in detail the Lesk algorithm and Walker's algorithm for word sense disambiguation. [10]

OR

- Q6)** a) Explain the following lexical knowledge networks? [10]
i) WordNet
ii) Indo WordNet
iii) VerbNets
iv) PropBank
v) Treebanks
b) Write Python code using NLTK library to split the text into tokens using whitespace, punctuation-based and default tokenization methods. [7]
- Q7)** a) Explain three stages of Question Answering system with neat diagram. [7]
b) Explain Rule based Machine Translation and Statistical Machine Translation (SMT) with suitable diagrams and example. [10]

OR

- Q8)** a) Describe following NLP applications: [10]
i) Text Entailment
ii) Dialog and Conversational Agents
b) Explain Natural Language Generation with reference architecture. [7]

