Total No. of Questions : 8]

SEAT No. :

P560

[Total No. of Pages : 2

[6004]-495
B.E. (Computer Engg.)
NATURAL LANGUAGE PROCESSING
(2019 Pattern) (Semester - VIII) (Elective - V) (410252(A))

Time : 2½ Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***Solve Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6 and Q.7 or Q.8.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right indicate full marks.***
4) ***Use of calculator is allowed.***
5) ***Assume suitable data, if necessary.***

Q1) a) Describe the process of building a simple Markov model for predicting the next word in a sentence with the help of example. **[6]**

b) Suppose you have a text corpus of 10,000 words, and you want to build a bigram model from this corpus. The vocabulary size of the corpus is 5,000. After counting the bigrams in the corpus, you found that the bigram "the cat" appears 50 times, while the unigram "the" appears 1000 times and the unigram "cat" appears 100 times. Using the add-k smoothing method with k=0.5, what is the probability of the sentence "the cat sat on the mat"? **[8]**

c) Write a short note on Latent Semantic Analysis (LSA). **[4]**

OR

Q2) a) What are generative models of language, and how do they differ from discriminative models? **[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 system in Natural Language Processing. **[4]**

b) What is Named Entity Recognition (NER)? Describe the various metrics used for evaluation. **[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 Coreference Resolution? Give examples. **[4]**

Q5) a) Write a note on : WordNet. **[10]**

b) List the tools available for the development of NLP applications? Write features NLTK and TextBlob? **[7]**

OR

Q6) a) Describe in detail the Lesk algorithm and Walker's algorithm for word sense disambiguation. **[10]**

b) Which types of tasks are performed by the Gensim library? Give an example. **[7]**

Q7) a) Write a note on : Sentiment Analysis. **[10]**

b) Explain Statistical Machine Translation (SMT) with suitable diagrams and example. **[7]**

OR

Q8) a) Describe various Machine Translation Approaches. **[10]**

b) Explain Natural Language Generation with reference architecture. **[7]**