Total No. of Questions: 8]

SEAT No. :

PB2258

[Total No. of Pages :3

[6263]-96
B.E.(Computer Engineering)
NATURAL LANGUAGE PROCESSING
(2019 Pattern) (Semester-VIII) (Elective-V) (410252 A)

Time : 2½ Hours] ***[Max. Marks : 70***

Instructions to the candidates:

1) ***Answer Q1 or Q2, Q3 or Q4, Q5 or Q6, Q7 or Q8.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right side indicate full marks.***
4) ***Assume Suitable data if necessary.***

Q1) a) What are generative models of language, and how do they differ from discriminative models? Provide an example of a generative model and describe how it can be used in NLP. **[9]**

b) Define Latent Dirichlet Allocation (LDA) and explain how it is used for topic modeling in text data. Discuss the key components of LDA, including topics, documents, and word distributions. **[9]**

OR

Q2) a) Describe the concept of contextualized representations, such as those generated by BERT, and how they are used in natual language processing. Discuss the advantages and disadvantages of contextualized representations. **[10]**

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]**

P.T.O.

Q3) a) Describe the Vector Space Model (VSM) for information retrieval. How does VSM represent documents and queries, and how are similarities calculated? Discuss the strengths and weaknesses of VSM. **[9]**

b) Discuss the different methods used for evaluating NER systems. What are common metrics for measuring NER system performance, and how can the results be analyzed to improve the system? **[9]**

OR

Q4) a) Define Cross-Lingual Information Retrieval (CLIR) and discuss the challenges involved in retrieving information from languages different from the query language. How do machine translation techniques assist in CLIR? **[9]**

b) Explain the importance of entity extraction in NLP. How does entity extraction differ from named entity recognition, and provide examples of real-world applications where entity extraction is crucial. **[9**

Q5) a) Describe Walker's algorithm for word sense disambiguation. How does it differ from other disambiguation techniques like Lesk's Algorithm, and what are the scenarios where it can be most effective? **[9]**

b) Compare the Indo Word Net with the traditional WordNet. What are the key differences and advantages of IndoWordNet, especially in the context of Indian languages? **[8]**

OR

Q6) a) Compare and constrast the natural Language Toolkit (NLTK), spaCy, and TexBlob. what are their main features and in what use cases are they most suitable? **[9]**

b) What is the significance of PropBank and VerbNet in linguistic resources? Provide examples of how these resoures can be used to extract semantic information from text. **[8]**

Q7) a) Define natural language generation and its role in NLP. How does NLG differ from text-to-speech synthesis, and what are the applications of NLG in data reporting and storytelling? **[9]**

b) Discuss the challenges in cross-lingual translation and provide examples of how it is beneficial in real-world applications. **[8]**

OR

Q8) a) Explain the key principles of rule-based machine translation. How do rule-based techniques differ from statistical approaches in machine translation? provide an example of a rule-based translation. **[9]**

b) Discuss the key components of a conversational agent, such as chatbots or virtual assistants. How do natural language generation and understanding play a role in creating effective conversational agents? **[8]**