Total No. of Questions : 8]

SEAT No. :

PD4591

[Total No. of Pages : 2

[6404]-96

B. E. (Computer Engineering)

NATURAL LANGUAGE PROCESSING

(2019 Pattern) (Semester - VIII) (410252(A)) (Elective - V)

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, 7 or Q.8.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right indicate full marks.***

Q1) a) Explain the concept of log-linear models in NLP. How do these models work, and what are some common applications of this approach? **[6]**

b) Explain the concept of doc2vec and its use in generating document embeddings. Discuss the differences between doc2vec and word2vec. **[8]**

c) Explain Non-Negative Matrix Factorization (NMF) in the context of topic modeling **[4]**

OR

Q2) a) What is a Markov model, and how is it used to generate language sequences? Describe the process of training and using a Markov model for language generation. **[6]**

b) Explain the concept of Latent Semantic Analysis (LSA) and how it is used for identifying relationships between words. **[8]**

c) Write a short note on TF-IDF (Term Frequency-Inverse Document Frequency) representation. **[4]**

Q3) a) Explain the concept of reference resolution and coreference resolution in NLP. How do these techniques help in understanding the relationships between entities in a text? Provide examples. **[8]**

b) Explain the concept of Cross-Lingual Information Retrieval (CLIR). Discuss the challenges of retrieving information across languages, and what techniques or models are commonly used to address these challenges? **[8]**

c) What is Information Retrieval, and how is it used in NLP? **[2]**

OR

P.T.O.

Q4) a) Explain the process of Entity Extraction in Information Retrieval. Discuss the various techniques and algorithms used for Entity Extraction. **[8]**

b) 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. **[8]**

c) What is Named Entity Recognition (NER)? **[2]**

Q5) a) Explain how NLTK, spaCy, and Gensim handle tokenization and part-of-speech tagging. Discuss the advantages and disadvantages of each library's approach. **[9]**

b) Explain the concept of WordNet as a lexical knowledge network. How can WordNet be used in natural language processing, and what are its limitations? **[8]**

OR

Q6) a) Discuss the role of Treebanks and Universal Dependency Treebanks in NLP research. How do they contribute to syntactic analysis and parsing in various languages? **[9]**

b) Explain the Lesk Algorithm for word sense disambiguation . How does it work, and what are its advantages and limitations? **[8]**

Q7) 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) Explain the fundamental components of a question-answering system. How do question answering systems utilize machine learning and NLP techniques to find accurate answers to questions? **[8]**

OR

Q8) a) Define text entailment and describe its importance in natural language understanding. How does text entailment differ from text similarity or paraphrasing? Provide an example of text entailment. **[9]**

b) Explain Natural Language Generation with reference architecture. **[8]**