CS6030

Natural language Processing

Question Paper

2 marks:

1. Given the sentence: “The quick brown fox jumps over the lazy dog”. Perform Tokenization.
2. Explain the difference between the morphology and syntax with an example.
3. Distinguish between Dataset and Corpus
4. Define a bigram and provide example using the sentence: “Artificial Intelligence is the future”
5. State any four use cases for Word Embedding.
6. Draw block diagram for N-Gram topic modelling
7. Describe the role of Hidden Markov models in Parts-of-speech tagging.
8. Describe a scenario where dependency parsing would be essential.
9. What is Tokenization in data preprocessing?
10. What is WordNet, and how is it useful in Word sense Disambiguation?

8 Marks:

1. Discuss various POS Tags in English. Given the text do POS tagging  
     
   “Sara, a single mother, launched a home bakery with $200. Through social media marketing and exceptional cakes, her business flourished. Within a year, she expanded to a storefront, creating jobs and inspiring her community. Sara’s passion turned her dream into a successful business.”
2. Compare and Contrast Word Embedding model
3. Explain N-grams and identify all the Bigram and Trigram for the following text

“Natural language processing involves the interaction between computers and human language. It is used to apply algorithms to identify and extract the natural language rules such that the unstructured language data is converted into a form that computers can understand”

1. What is pre-trained language model? Will topic model will be impacted using pre-trained LM? How about using word embedding using pre-trained LM? Justify
2. Draw Ambiguous parse trees for “She saw the man with binoculars ” and “Visiting relatives can be boring”.
3. What is morphological analyzer? Explain the types of Morphological Analysis with examples.
4. Explain Word2VEC embedding model with example.
5. Discuss the concepts of Lexicon and Lexeme in linguistic theory. Additionally, elaborate on the various types of lexical relations that exist between lexemes, providing examples to illustrate your points**.**
6. Discuss three different types of evaluation metrics used in text classification. Provide a comprehensive explanation of each metric, including its calculation method, significance, and potential limitations. Illustrate your points with relevant examples..
7. Perform NER for the following Text.

புதிய தமிழகம் ஆர்ப்பாட்டம் திருநெல்வேலி : கூடங்குளத்தில் உற்பத்தியாகும் மின்சாரம் முழுவதையும் தமிழகத்திற்கு வழங்க வலியுறுத்தி புதிய தமிழகம் கட்சியினர் ஆர்ப்பாட்டத்தில் ஈடுபட்டனர். அணுமின் நிலையத் திற்கு வெளியே காமராஜர் சிலைமுன்பாக நடந்த ஆர்ப்பாட்டத்திற்கு, கட்சி தலைவர் கிருஷ்ணசாமி தலைமை வகித்தார். அவர் கூறுகையில், "தமிழகத்தில் மின்சார தட்டுப்பாடு நிலவுகிறது. கூடங்குளம் அணுஉலையில் ஏதாவது பாதிப்பு ஏற்பட்டால் தமிழக மக்களைத்தான் பாதிக்கும்.  மின் தட்டுப்பாடுள்ள தமிழகத்திற்கு இங்கு உற்பத்தியாகும் மின்சாரம் முழுவதையும் வழங்க வேண்டும்,'' என்றார்.

1. Give a simple paragraph and explain the process of obtaining N-grams from the text.
2. Explain Vector space model with neat illustrated example.
3. Explain Word Sense Disambiguation with the example   
   “The fisherman went to the bank to catch some fish.”

Part-c

1.Explain Dependency parsing with 2 example of your choice

2.Explain and show examples for distinguishing information Extraction vs. information Retrieval.