KUMARAGURU
college of technology
character is life

M.TECH DEGREE EXAMINATIONS: APRIL /MAY 2024

(Regulation 2018)

Second Semester

DATA SCIENCE

P18ITE0019**:** Natural Language Processing

COURSE OUTCOMES

CO1:   Explain the language models.

CO2:   Analyze the natural language text.

CO3:   Generate the natural language.

CO4:   Do machine translation.

CO5:   Apply information retrieval techniques.

Time: Three Hours   **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1.   Identify the correct sequence structure of given Grammar to Generate Parse tree   CO3   [K₃]

for the following sentence "The police will catch the snatcher" .

(i)V-> Aux+Verb

(ii)S-> NP+VP

(iii)NP->Det+Noun

(iv)VP->V+NP

a)   (i),(iii),(ii),(iv)   b)   (ii),(i),(iii),(iv)

c)   (ii),(iv),(iii),(i)   d)   (iii),(iv),(i),(ii)

2.   The word "window" is a ____ of "room" .   CO3   [K₂]

a)   hyponym   b)   hypernym

c)   holonym   d)   meronym

3.   CO5   [K₁]

Which of the following is related to Zipf's law.

a)   The frequency of a word type is proportional to its rank by frequency.

b)   The frequency of a word type is inversely proportional to its rank by frequency.

c)	The frequency of a word type does not influence its ranking.	d)	The frequency of a word type influence its ranking.

4. Matching type item with multiple choice code　　　　　　　　　　CO5　[K₂]

List I	List II
A. Early NLP system	i. Python's NLTK library
B. Lexical resource for IR research	ii. Paninian grammar
C. NLP tool	iii. Eliza
D. Indian language framework	iv. WordNet

a)	A-ii　B-i　C-iv　D-iii	b)	A-iii　B-iv　C-i　D-ii
c)	A-ii　B-iv　C-iii　D-i	d)	A-iii　B-i　C-ii　D-iv

5. "Rohan Was with Her, They Both Go Together", In the Given Sentence Who Is Her Is Unclear, Specify the Type of Ambiguity?　　CO2　[K₁]

a)	Semantic Ambiguity	b)	Anaphoric Ambiguity
c)	Pragmatic Ambiguity	d)	Lexical Ambiguity

6. Which approach is commonly used for machine translation involving Indian languages?　　CO4　[K₁]

a)	Statistical approach	b)	Neural network-based approach
c)	Rule-based approach	d)	a,b & c options

7. Assertion (A): Viterbi algorithm used to find the best tag sequence given the word sequence and a HMM model.　　CO2　[K₂]

Reason (R): It uses dynamic programming to arrive at the solution by constructing solutions for the smaller problems.

a)	Both A and R are Individually true and R is the correct explanation of A	b)	Both A and R are Individually true but R is not the correct explanation of A
c)	A is true but R is false	d)	A is false but R is true

8. Which of the following approaches to machine translation relies on linguistic rules to translate text?　　CO4　[K₁]

a)	Rule-based Machine Translation (RBMT)	b)	Statistical Machine Translation (SMT)
c)	Neural Machine Translation (NMT)	d)	Hybrid Machine Translation (HMT)

9. What is the most significant challenge in processing Indian languages in NLP?　　CO1　[K₁]

a)	Lack of data	b)	Limited grammar rules

c) Multiple scripts and dialects   d) High computational cost

10. What is the purpose of a corpus in natural language processing?   CO1   [K₁]

    a) To represent a language model   b) To store and organize large amounts of text data

    c) To measure the accuracy of a language model   d) To train a machine learning algorithm

PART B (10 x 2 = 20 Marks)

#	Question	CO	K
11.	Define Sentiment Analysis.	CO1	[K2]
12.	Differentiate between stemming and lemmatization.	CO2	[K2]
13.	What is a stochastic tagger?	CO1	[K2]
14.	List the steps involved in Supervised learning for disambiguation.	CO4	[K3]
15.	Define hypernymy. Give an example.	CO4	[K2]
16.	List two major characteristics of Indian languages.	CO3	[K1]
17.	Identify Noun and verb phrase in the following sentence "Time flies like an arrow".	CO2	[K2]
18.	Use the Discourse planner and analyze the given example.	CO3	[K3]
19.	For a query, a retrieval system retrieves 30 relevant documents and 14 irrelevant documents from a document collection that consists of 100 relevant documents. Calculate the precision of the retrieval system.	CO5	[K3]
20.	Distinguish between classical and non-classical models of information retrieval.	CO5	[K2]

PART C (6 x 5 = 30 Marks)

21. Explain probabilistic Grammar using bigram model for the sentences: CO1   [K3]

"The Arabian Knights"

"These are the fairy tales of the east"

"The stories of the Arabian knights are translated in many languages".

22. Discuss about two step morphological parsing in detail with example.   CO3   [K2]

23. Define Context free grammar. For the sentence "The child ate a cake with fork", develop the CFG .   CO2   [K3]

24.	Define cohesion and reference resolution in discourse processing. How do they contribute to the coherence and structure of a text?	CO4	[K$_2$]
25.	Write the steps involved in memory-based learning for disambiguation.	CO5	[K$_2$]
26.	Discuss Zipf's law with an example.	CO5	[K$_2$]

Answer any FOUR Questions
PART D (4 x 10 = 40 Marks)

27	Explain about Rule based Machine Translation approach. Identify common problems encountered in machine translation and suggest possible solutions.	CO4	[K$_2$]
28.	Compute the minimum edit distance between these two strings "grey "and "dog" using Levenshtein method.	CO2	[K$_3$]
29.	Discuss the role of regular expressions and finite-state automata in word-level analysis. Provide an example of how they are used in NLP.	CO3	[K$_2$]
30.	Compare and contrast different grammar-based language models used in NLP. Provide examples to illustrate your points.	CO1	[K$_2$]
31.	Scenario: You are tasked with improving the information retrieval (IR) system of a digital library to better serve users who search in multiple languages. Question: Discuss the use of lexical resources such as WordNet and FrameNet in enhancing the IR system. Explain how these resources can improve the accuracy of search results and user satisfaction.	CO5	[K$_3$]

****