KUMARAGURU
college of technology
character is life

Register Number:……

M.TECH. DEGREE EXAMINATIONS: APRIL / MAY 2023

(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 which one of the following sequence is correct for human. CO1 [K_2]

   a) Analysis and Implications -> wisdom-> Unstructured data -> information -> knowledge

   b) information -> Analysis and Implications -> Unstructured data -> knowledge -> wisdom

   c) Unstructured data -> information -> Analysis and Implications -> knowledge -> wisdom

   d) knowledge -> wisdom -> Unstructured data -> information -> Analysis and Implications

2. Which one of the following sequence is correct for processing text? CO1 [K_2]

   a) Vectorization ->Normalization - > Tokenization -> vocabulary -> Noise Removal

   b) Vocabulary -> Noise Removal -> Normalization - > Tokenization -> vectorization

   c) Noise Removal -> Normalization - > Tokenization -> vocabulary -> vectorization

   d) Tokenization -> Noise Removal -> Normalization - > vocabulary -> vectorization

3. Which one of the following is incorrect with respect to non-deterministic finite automata? CO2 [K_4]

   a) it has finite set of states

   b) it has transition function

   c) it has finite set of symbols

   d) it can determine the state to which the machine will move.

4. Match the regular expression against corresponding regular set: CO2 [K_4]

List I	List II
A. (10)*	i. {0, 100, 10100, …..}
B. (10)*0	ii. { ∈, 10, 1010, …..}
C. (1+0)*	iii. {11,00, 1100, ….}
D. (11+00)*	iv. { ∈, 1, 0, 11, 10, 00,……}

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

5. Which one of the following is more suitable output of the regular expression (1+01)* + (1+01)*1 CO3 [K_4]

a) no two 1's come together
b) no two 0's come together
c) no two 10's come together
d) no two 110's come together

6. Which one of the following is suitable for the language L = { ∈, 0, 1, 11, 01, 010, 10, 101, 1010, 0101, …….}? CO3 [K_4]

a) 0(1+1)*+(1+10)*
b) 0(1+10)*+(1+10)*
c) 0(1+10)*+(1+0)*
d) (1+10)*+(1+10)*

7. While using machine translation, the number of words in translated sentence may becomes different. Which of the following is/are application? CO4 [K_3]
(i) Word-based translation (ii) Phrase-based translation (iii) syntax-based translation

a) (i) and (ii)
b) (ii) and (iii)
c) (i)
d) (ii)

8. Assertion (A) : Language translator is not an application of NLG but NLP application. CO4 [K_3]
Reason (R) : It is used to find out what a particular word or phrase in a different language.

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

9. Which of the design features are not related to information retrieval systems? CO5 [K_2]

a) Inverted Index
b) stemming
c) Stop word Elimination
d) inverting stems

10. Which of the following model are not related with information retrieval systems? CO5 [K_1]

a) Classical
b) Hybrid
c) Traditional
d) Non classical

PART B (10 x 2 = 20 Marks)

11.	Recall the two levels needed for text preprocessing based on records of data.	CO1	[K_2]
12.	Compare human vs machine with regard to language processing.	CO1	[K_2]
13.	Show the use of PoS tagging.	CO2	[K_2]
14.	List the techniques that are used for error detection.	CO2	[K_1]
15.	Tell the scope of the lexical semantics.	CO3	[K_2]
16.	Create an example of discourse.	CO3	[K_2]
17.	Tell any four applications of NLG.	CO4	[K_2]
18.	Identify any two problems in machine translation.	CO4	[K_2]
19.	Apply the stemming on “eating”, “eats” and “eaten” and find the root stem.	CO5	[K_3]
20.	Differentiate stemming and lemmatization.	CO5	[K_2]

PART C (6 x 5 = 30 Marks)

21.	Identify the steps involved in the preprocessing of text.	CO1	[K_2]
22.	Paraphase the limitations and problems of natural language processing.	CO1	[K_2]
23.	Summarize the way of performing syntactic analysis in a regular language.	CO2	[K_2]
24.	Build the hierarchical discourse structure for the following: “Somu went the bank to withdraw money. He then took a cab to Shyam’s cloth shop. He wanted to buy some clothes. He do not have new clothes for party. He also wanted to meet Shyam regarding his wife’s health.”	CO3	[K_3]
25.	Explore the stages of NLG working process.	CO4	[K_2]
26.	Explain the use of WordNetLemmatizer class with an example.	CO5	[K_3]

Answer any FOUR Questions

PART D (4 x 10 = 40 Marks)

27.	Illustrate the statistical based language model and explain its types and limitations.	CO1	[K_2]
28.	Explain about morphological parsing and show the use of FSA in it. Give an example.	CO2	[K_3]

29. Illustrate the various approaches used in Word Sense Disambiguation. CO3 [K_2]

30. Explain the architecture of NLG systems and their characteristics. CO4 [K_2]

31. Explore the different types of information retrieval systems available in NLP. CO5 [K_2]

****