Q. Code:112100

Reg. No. | | | | | | | | | | | | | |

B.TECH. DEGREE EXAMINATIONS, MAY 2024

Sixth Semester

AD18603 – NATURAL LANGUAGE PROCESSING TECHNIQUES

(Artificial Intelligence and Data Science)

(Regulation 2018/2018A)

TIME: 3 HOURS **MAX. MARKS: 100**

COURSE OUTCOMES	STATEMENT	RBT LEVEL
CO 1	Tag a given text with basic Language features.	**3**
CO 2	Design an innovative application using NLP components.	**3**
CO 3	Implement a rule based system to tackle morphology/syntax of a language.	**3**
CO 4	Design a tag set to be used for statistical processing for real-time applications.	**3**
CO 5	Compare and contrast the use of different statistical approaches for different types of NLP applications.	**3**

PART- A (10 x 2 = 20 Marks)

(Answer all Questions)

		CO	RBT LEVEL
1.	Summarise the closure property of Finite state transducers.	**1**	**2**
2.	Define period symbol in regular expression with suitable example.	**1**	**2**
3.	Define Markov assumption.	**2**	**2**
4.	Infer the chain rule of probability in N-Gram model.	**2**	**2**
5.	Define the purpose of treebank in NLP.	**3**	**2**
6.	Infer the concept of Phrase structure grammar.	**3**	**2**
7.	Why Feature extraction from textual data is difficult?	**4**	**2**
8.	Discuss the significance of dictionary and thesaurus in NLP	**4**	**2**
9.	Formulate cosine similarity measure in computational discourse.	**5**	**2**
10.	Define conference resolution.	**5**	**2**

PART- B (5 x 14 = 70 Marks)

			Marks	CO	RBT LEVEL
11. (a)	**(i)**	Identify various challenges of NLP in machine learning.	**(7)**	**1**	**3**

(ii)	Write the regular expression for the following: i. Language starting and ending with a and having any having any combination of b's in between. ii. Starting with a but not having consecutive b's as L = {a, aba, aab, aba, aaa, abab, ...} iii. Language accepting all the string in which any number of a's is followed by any number of b's is followed by any number of c's	**(7)**	**1**	**3**
	(OR)			
(b)	Explain the algorithm of Minimum edit distance and Compute the edit distance (using insertion cost 1, deletion cost 1, and substitution cost 2) of "LEDA" to "DEAL". Show your work (using the edit distance grid).	**(14)**	**1**	**3**
12. (a)	Describe the Hidden Markov Chain PoS tagging technique in detail.	**(14)**	**2**	**2**
	(OR)			
(b)	Illustrate in detail about different word classes with suitable examples.	**(14)**	**2**	**2**
13. (a)	Elaborate on the CKY algorithm and illustrate probabilistic CKY for the following statement using the given grammar. "The price includes a facemask" ¬ S->NP VP [0.80] ¬ NP->Det N [0.3] ¬ VP->V NP [0.20] ¬ V->includes [0.05] ¬ Det->the [0.4] ¬ Det->a [0.4] ¬ N->price [0.01] ¬ N->facemask [0.02]	**(14)**	**3**	**3**
	(OR)			
(b)	Demonstrate in detail about dependency grammar and illustrate the types of ambiguity with suitable example.	**(14)**	**3**	**3**
14. (a)	Discuss in detail about various Techniques used for Feature Extraction.	**(14)**	**4**	**3**
	(OR)			
(b)	Explain selectional restrictions and provide appropriate justification for the requirement for selectional preferences.	**(14)**	**4**	**3**
15. (a)	Elaborate on Anaphora Resolution using Hobbs and Centering Algorithm.	**(14)**	**5**	**3**
	(OR)			
(b)	Using appropriate applications describe any three resources of NLP.	**(14)**	**5**	**3**

PART- C (1 x 10 = 10 Marks)

(Q.No.16 is compulsory)

		Marks	CO	RBT LEVEL
16.	With suitable explanation construct a parse tree for the following sentence using CFG rules: **The man read this book**	**(10)**	**3**	**5**

Rules:
S->NP VP
S->VP
NP->Det Nom
Nom->Noun
VP->verb NP
Verb->read
Det->the , this
Noun->book, man
Verb->book, read

****