Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

VII SEMESTER B.TECH END SEMESTER EXAMINATION, NOV/DEC 2023

SUBJECT: PE-III NATURAL LANGUAGE PROCESSING [CSE 4061]

(- / - /2023)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Missing data may be suitably assumed.

Q No.	Questions		CO	AHEP LO	Blooms level
1A.	"Natural language is extremely rich in form and structure, and very ambiguous". Justify your answer through suitable examples for ambiguities at various levels.	4M	CO1	1,2,5	5
1B.	Illustrate the following i) Morphological parsing ii) Stemming iii) Lemmatization iv) Tokenization	4M	CO1	1,2,5	2
1C.	Construct a Finite State Automata (FSA) for recognizing adjective morphology. The FSA must handle words like clear, real, big, red for the prefix 'un' and suffixes 'er', 'ly', 'est'.	2M	CO1	1,2,5	3
2A.	Write the algorithm for finding the minimum edit distance between two strings. Illustrate the working with an example.	4M	CO5	1, 2, 3, 4	3
2B.	Given a N-gram model with the count $C(W_1,W_2,....W_n)$ having non-zero value, how would you use interpolated model and back-off model?. Give suitable equations. Which of these models use information from lower order models?	3M	CO5	1, 2, 3, 4	3
2C.	Given the bigram frequency Table 2C1 with a vocabulary size of V=1446, and unigram count in Table 2C2 having unigram counts, apply add-1 smoothing to obtain smoothed bigram probabilities table. Table 2C1	3M	CO5	1, 2, 3, 4	3

		i	want	to	eat	chinese	food	lunch	spend				
		i	5	827	0	9	0	0	0	2			
		want	2	0	608	1	6	6	5	1			
		to	2	0	4	686	2	0	6	211			
		eat	0	0	2	0	16	2	42	0			
		chinese	1	0	0	0	0	82	1	0			
		food	15	0	15	0	1	4	0	0			
		lunch	2	0	0	0	0	1	0	0			
		spend	1	0	1	0	0	0	0	0			
	Table 2C2 i want to eat chinese food lunch spend 2533 927 2417 746 158 1093 341 278												
3A.	Given the transition probabilities and the observation likelihood, resolve the ambiguity in POS tagging the sentence *Flies like a flower*. Which one of the following is the correct tagging using HMM model? Show mathematically. The symbol <S> is the start-of-sentence symbol. Flies/NN like/VB a/DT flower/NN Flies/VB like/ADJ a/DT flower/NN <S> ADJ DT NN VB ADJ 0.01 0.13 0.33 0.18 DT 0 0.008 0.11 0 NN 0.21 0 0.4 0.12 VB 0.09 0 0.12 0.13 0.0023 0.009 0.09 0.08 a like Flies Flower ADJ 0 0.02 0 0.08 DT 0.12 0 0 0 NN 0 0.003 0.13 0.091 VB 0 0.12 0.08 0.002										4M	CO4	1, 2, 4	3
3B.	What are Treebanks?. Compare dependency grammar and phrase structure grammar										4M	CO4	1, 2, 4	2
3C	Mention the word class which resembles preposition or an adverb but is used in combination with a verb. Explain with a suitable example.										2M	CO2	1,2,3,4, 5	3
4A.	Construct the parse trees for the sentence "John called Sue from Denver" using CKY parsing. S -> NP VP VP -> V NP NP -> NP PP VP -> VP PP PP -> P NP NP -> John NP -> Sue NP -> Denver V -> called V -> sue P -> from Given a context-free grammar, how do you convert it to Chomsky Normal Form (CNF).										4M	CO3	1,2,3,4, 5	3
4B.	Given the Probabilistic Context Free Grammar (PCFG) below, find the parse of the sentence "The flight includes a meal"										3M	CO3	1,2,3,4, 5	3

	S → NP VP .80 Det → The .40 NP → Det N .30 Det → a .40 VP → V NP .20 N → meal .01 V → includes .05 N → flight .02 Calculate the probability of the constructed parse tree.					
4C.	"Probabilistic parsing is to solve the problem of disambiguation". Justify your answer with suitable example.	3M	CO3	1,2,3,4, 5	5
5A.	Given the sentence type handling yes-no questions, how do you modify the grammar to handle subject verb agreement?	4M	CO2	1,2,3,4, 5	2
5B.	Given the bracketed notation, construct the equivalent parse tree [$_S$ [$_{NP}$ [$_{PRO}$ I]] [$_{VP}$ [$_V$ prefer] [$_{NP}$ [$_{Det}$ a] [$_{Nom}$ [$_N$ morning] [$_N$ flight]]]]]	3M	CO2	1,2,3,4, 5	3
5C.	What are sentential complements for the verbs? Give example for transitive verbs and intransitive verbs.	3M	CO5	1,2,3,4	2

¤¤¤¤¤¤¤¤¤¤¤¤¤¤¤¤