Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering

End Sem (Even) Examination May-2019

CS3EA06 / IT3EA06 Natural Language Processing

Programme: B.Tech. Branch/Specialisation: CSE/IT

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

Q.1 i. Which of the following statement(s) is/are true about DFA **1**

I. Infinitely many states but finitely many accept states
II. Finitely many states.
III. Transition function with finite domain and range.
IV. Transition function with possibly infinite range or domain.

(a) I and II (b) II and III (c) Only IV (d) Only I

ii. Regular expression /colou?r/ matches the string: **1**

(a) Colour only (b) Color only
(c) Colour or color (d) None of these

iii. The list of stems and affixes are called as **1**

(a) Morphotactics (b) Lexicon
(c) Clitics (d) Tokens.

iv. Choose the correct form of morphology in the following examples: **1**

I. Logic, Logician II. Bring, brought
(a) Inflectional, Derivational (b) Derivational, Inflectional
(c) Inflectional, Inflectional (d) Derivational, Derivational

v. Suppose P(w|(c) represents probability of a word given its rating. We should divide P(w|(c) by P(w) to make it: **1**

(a) Comparable across different words
(b) Comparable across different rating
(c) Both (a) and (b)
(d) None of these

P.T.O.

vi. Which of the following methods can be used to solve the edit distance problem? **1**
(a) Recursion (b) Dynamic programming
(c) Both (a) and (b) (d) None of these

vii. Parse trees are useful in application such as **1**
(a) Grammar checking (b) Semantic analysis
(c) Information extraction (d) All of these

viii. A 5-gram model is a _____ order Markov Model: **1**
(a) Six (b) Five (c) Four (d) Constant

ix. Important features to detect sentiment are: **1**
(a) Presence of negation words
(b) Presence of sentiment words
(c) Capitalisation of words
(d) All of these

x. Dictionary-based disambiguation is based on _____ resources. **1**
(a) Semantic (b) Lexical (c) Social (d) Human

Q.2 i. Give one example each of an english sentence with semantic ambiguity, syntactic ambiguity, in each case specify the ambiguity clearly. **4**

ii. Differentiate: **6**
(a) Formal language and Natural language.
(b) Data and Knowledge

OR iii. What is regular expression? How they are useful in natural language processing? **6**

Q.3 i. Distinguish between inflectional and derivational morphology. **4**

ii. Why text pre-processing is required? Write the different approaches for text pre-processing. **6**

OR iii. What are Rule based, Stochastic and Transformation based POS tagging. **6**

Q.4 i. Define: (a) Phoneme (b) Allophone. **4**

ii. Explain the Bayesian method to spelling error correction with an example. **6**

OR iii. Consider the strings “monday” and “tuesday”. What is the edit distance between the two strings? Apply minimum edit distance algorithm. **6**

Q.5 i. What is N-gram? How smoothing and backoff approaches works? **4**

ii. Discuss probabilistic parsing with an example. **6**

OR iii. Differentiate: **6**
(a) Top down parsing and bottom up parsing
(b) Tree bank and parse tree

Q.6 Attempt any two:

i. What are the Knowledge Sources in Word Sense Disambiguation (WSD)? Explain them in brief. **5**

ii. What is sentiment analysis? With the help of example explain how natural language processing approach helps in sentiment analysis. **5**

iii. Describe what is text classification with an example. **5**

******

Marking Scheme

CS3EA06 / IT3EA06 Natural Language Processing

Q.1	i.	Which of the following statement(s) is/are true about DFA I. Infinitely many states but finitely many accept states II. Finitely many states. III. Transition function with finite domain and range. IV. Transition function with possibly infinite range or domain. (b) II and III	**1**
	ii.	Regular expression /colou?r/ matches the string: (c) Colour or color	**1**
	iii.	The list of stems and affixes are called as (b) Lexicon	**1**
	iv.	Choose the correct form of morphology in the following examples: I. Logic, Logician II. Bring, brought (b) Derivational, Inflectional	**1**
	v.	Suppose P(w|(c) represents probability of a word given its rating. We should divide P(w|(c) by P(w) to make it: (a) Comparable across different words	**1**
	vi.	Which of the following methods can be used to solve the edit distance problem? (c) Both (a) and (b)	**1**
	vii.	Parse trees are useful in application such as (d) All of these	**1**
	viii.	A 5-gram model is a _____ order Markov Model: (c) Four	**1**
	ix.	Important features to detect sentiment are: (d) All of these	**1**
	x.	Dictionary-based disambiguation is based on _____ resources. (b) Lexical	**1**
Q.2	i.	Give one example each of an english sentence with semantic 2 marks for each (2 marks * 2)	**4**
	ii.	Differentiate: (a) Formal language and Natural language. Three differences 1 mark for each 3 marks (b) Data and Knowledge Three differences 1 mark for each 3 marks	**6**
OR	iii.	Regular expression 2 marks They are useful in natural language processing 4 marks	**6**
Q.3	i.	Distinguish between inflectional and derivational morphology Four differences 1 mark for each (1 mark * 4)	**4**
	ii.	Text pre-processing is required 2 marks Any four approaches for text pre-processing. 1 mark for each (1 mark * 4) 4 marks	**6**
OR	iii.	Rule based, Stochastic and Transformation based POS tagging. Two marks for each (2 marks * 3)	**6**
Q.4	i.	Define: (a) Phoneme 2 marks (b) Allophone. 2 marks	**4**
	ii.	Bayesian method to spelling error correction 4 marks Example 2 marks	**6**
OR	iii.	Consider the strings “monday” and “tuesday”. What is the edit distance between the two strings? Apply minimum edit distance algorithm. Correct answer 2 marks Table 4 marks	**6**
Q.5	i.	N-gram 2 marks Smoothing and backoff approaches works 2 marks	**4**
	ii.	Probabilistic parsing 4 marks Example. 2 marks	**6**
OR	iii.	Differentiate: (a) Top down parsing and bottom up parsing Three differences 1 mark for each 3 marks (b) Tree bank and parse tree Three differences 1 mark for each 3 marks	**6**
Q.6		Attempt any two:	
	i.	Knowledge Sources in Word Sense Disambiguation (WSD) 2 marks Explanation of them 3 marks	**5**
	ii.	Sentiment analysis 2 marks Natural language processing approach helps in sentiment analysis 3 marks	**5**
	iii.	Text classification 3 marks Example 2 marks	**5**

******