

Enrollment No.....



Faculty of Engineering
End Sem (Odd) Examination Dec-2019
IT3EA06 Natural Language Processing
Programme: B.Tech. Branch/Specialisation: 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. If Regular Expression is $[a-z][A-Z]$. Then accepted strings contain: **1**
 (a) All lower case character (b) All upper case character
 (c) Both (a) and (b) (d) All digit
- ii. Which of the following is not an approach of NLP? **1**
 (a) Rule Based (b) Algorithmic based
 (c) Deep learning (d) Probabilistic Based
- iii. Which is not a feature of finite state transducer: **1**
 (a) Generator (b) Set Relater
 (c) Recognizer (d) Tokenizer
- iv. Text Normalization is process of: **1**
 (a) Transform text into a single canonical form
 (b) Chopping text into smaller pieces
 (c) Parsing data into different language
 (d) Extraction of text from structured data
- v. Which of the following is not a Phonological rule? **1**
 (a) Assimilation (b) Dissimilation
 (c) Neutralization (d) Submission
- vi. The minimum Edit distance on two similar character is..... **1**
 (a) 1 (b) 0 (c) 2 (d) 3
- vii. N-grams are defined as the combination of N keywords together. **1**
 How many bi-grams can be generated from given sentence:
 “Regular Expression is a formula in a special language”
 (a) 7 (b) 8 (c) 9 (d) 10

P.T.O.

[2]

- viii. Suppose a language model assigns the following conditional n-gram probabilities to a 3-word test set: $1/4$, $1/2$, $1/4$. Then $P(\text{test-set}) = 1/4 * 1/2 * 1/4 = 0.03125$. What is the perplexity?
(a) 0.02 (b) 0.03 (c) 0.04 (d) 0.05 **1**
- ix. Which of the following analysis can perform tweet classification with regards to context mentioned above? **1**
(a) Spelling Correction
(b) Sentiment Analysis
(c) Word sense Disambiguation
(d) Machine Translation
- x. Machine Translation **1**
(a) Converts one human language to another
(b) Converts human language to machine language
(c) Converts any human language to English
(d) Converts Machine language to human language
- Q.2 i. How regular expression plays an important role to process natural language? **4**
ii. What do you mean by ambiguity? Explain it with its type & example. **6**
- OR iii. Explain knowledge in speech & language processing by pyramid structure. **6**
- Q.3 i. How Text pre-processing helps to processed natural language? **4**
ii. Explain part of Speech Tagging with its different type. **6**
- OR iii. Explain Morphology with its type & also explain why Finite state transducer used over finite state automata in morphological parsing. **6**
- Q.4 i. Justify the Statement “Probabilistic model is more accurate to detect spelling & pronunciation errors”. **3**
ii. Evaluate the Levenshtein distance of the following strings where Insertion, Substitution & deletion cost will be 1:
S1: Intention
S2: Execution
Also write the operation used in above strings. **7**

[3]

- OR iii. Write an algorithm to find out minimum edit distance between two strings? Illustrate with following two string & find out minimum edit distance. **7**
S1: abcdef
S2: azced
- Q.5 i. Why maximum likelihood estimation used over markov assumption & language modelling to solve N-gram probability? **4**
ii. Find out probability, perplexity & entropy of the Test sentence i.e. $\langle s \rangle$ I I am not $\langle /s \rangle$ through maximum likelihood estimation in bi-gram. Where training sentences are: **6**
 $\langle s \rangle$ I am a human $\langle /s \rangle$
 $\langle s \rangle$ I am not a stone $\langle /s \rangle$
 $\langle s \rangle$ I live in Indore $\langle /s \rangle$
- OR iii. What do you mean by parsing? Explain different type of parsing with example. **6**
- Q.6 Attempt any two:
i. What are different kinds of methods used to analyse sentiment of natural language? **5**
ii. Explain different application of natural language processing. **5**
iii. Write a short note on: **5**
(a) Machine translation (b) Word sense disambiguation

Marking Scheme

IT3EA06 Natural Language Processing

Q.1	i.	If Regular Expression is [a-z][A-Z]. Then accepted strings contain: (c) Both (a) and (b)	1		
	ii.	Which of the following is not an approach of NLP? (b) Algorithmic based	1		
	iii.	Which is not a feature of finite state transducer: (d) Tokenizer	1		
	iv.	Text Normalization is process of: (a) Transform text into a single canonical form	1		
	v.	Which of the following is not a Phonological rule? (d) Submission	1		
	vi.	The minimum Edit distance on two similar character is..... (b) 0	1		
	vii.	N-grams are defined as the combination of N keywords together. How many bi-grams can be generated from given sentence: "Regular Expression is a formula in a special language" (b) 8	1		
	viii.	Suppose a language model assigns the following conditional n-gram probabilities to a 3-word test set: 1/4, 1/2, 1/4. Then $P(\text{test-set}) = 1/4 * 1/2 * 1/4 = 0.03125$. What is the perplexity? (b) 0.03	1		
	ix.	Which of the following analysis can perform tweet classification with regards to context mentioned above? (b) Sentiment Analysis	1		
	x.	Machine Translation (a) Converts one human language to another	1		
Q.2	i.	Regular expression plays an important role to process natural language As per explanation	4		
	ii.	Definition of ambiguity Its type Example	2 marks 2 marks 2 marks	6	
	OR iii.	Knowledge in speech & language processing by pyramid structure. Stepwise marking		6	
Q.3	i.	Text pre-processing helps to processed natural language		4	
	ii.	Definition of part of Speech Tagging Its different types	2 marks 4 marks	6	
	OR iii.	Definition of Morphology Its type Reason	2 marks 2 marks 2 marks	6	
Q.4	i.	Justification of Statement		3	
	ii.	Find out minimum edit distance Operation	5 marks 2 marks	7	
	OR iii.	Algorithm To find out minimum edit distance	3 marks 4 marks	7	
Q.5	i.	Reason		4	
	ii.	Find out probability Perplexity Entropy of the Test sentence	3 marks 1.5 marks 1.5 marks	6	
	OR iii.	Definition of parsing Different type of parsing Example	2 marks 3 marks 1 mark	6	
Q.6		Attempt any two:			
	i.	Kinds of methods used to analyse sentiment of natural language As per the explanation		5	
	ii.	Application of natural language processing 1 mark for each application		5	(1 mark *5)
	iii.	Write a short note on: (a) Machine translation (b) Word sense disambiguation		5	2.5 marks 2.5 marks
