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

Enrollment No.....

Faculty of Engineering

End Sem (Odd) Examination Dec-2022

CS3EA06 / IT3EA06 Natural Language Processing

Programme: B.Tech. Branch/Specialisation: CSE All / 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. What is the field of Natural Language Processing (NLP)? **1**

(a) Computer science (b) Artificial intelligence

(c) Linguistics (d) All of these

ii. What is 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

iii. What is Morphological Segmentation? **1**

(a) Does discourse analysis

(b) Separate words into individual morphemes and identify the class of the morphemes

(c) Is an extension of propositional logic

(d) None of these

iv. Rule-based POS taggers does not possess which of the following properties- **1**

(a) The rules in Rule-based POS tagging are built auto

(b) These taggers are knowledge-driven taggers

(c) These taggers are consist of many hand written rules

(d) The information is coded in the form of rules.

v. Where does the bayes rule can be used? **1**

(a) Solving queries

(b) Increasing complexity

(c) Decreasing complexity

(d) Answering probabilistic query

P.T.O.

vi. The study of the sound patterns in natural language and the rules that govern them is- **1**
(a) Phonetics (b) Morphology (c) Phonology (d) Syntax

vii. N-grams are defined as the combination of N keywords together. How many bi-grams can be generated from the given sentence: "Gandhiji is the father of our nation". **1**
(a) 7 (b) 6 (c) 8 (d) 9

viii. The statement "Which team won the match?" can be represented as- **1**
(a) N->Wh-NP VP
(b) S->Wh-NP VP
(c) VP->Wh-NP VP
(d) S->Wh-NP NP

ix. Which of the following algorithms is widely used for text classification? **1**
(a) Decision Tree (b) Support vector machine
(c) Naive Bayes (d) All of these

x. N-Gram language models cannot be used for _____. **1**
(a) Spelling Correction
(b) Predicting the completion of a sentence
(c) Removing semantic ambiguity
(d) Speech Recognition

Q.2 i. What is natural language processing? Discuss approach of NLP. **4**
ii. Define ambiguity. Describe various types with example. **6**
OR iii. Explain different phases in natural language processing with diagram. **6**

Q.3 i. Define Morphemes, Morphology and Corpus. **2**
ii. Explain "BOW" and "TF-IDF Vectorizer" methods of feature extraction. **4**
iii. Explain POS tagging with example and also write its types. **4**
OR iv. Differentiate between inflectional and derivational morphology. **4**

Q.4 Attempt any two:
i. Define the term phonetics with their types. **5**
ii. What is the spelling error? What are their types? **5**
iii. Explain Minimum Edit Distance algorithm and also find the Minimum Edit Distance between EXECUTION and INTENTION. **5**

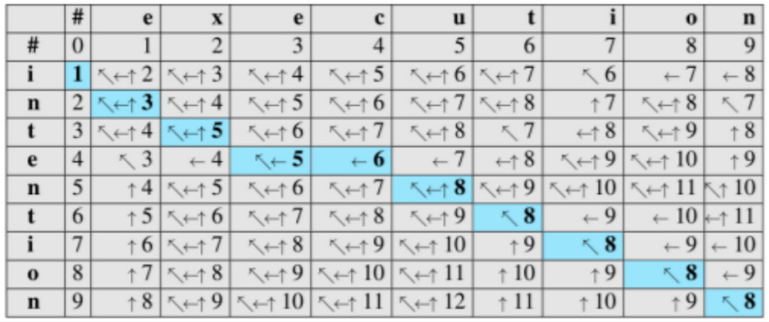
Q.5 i. Explain the terms smoothing and backoff. **4**
ii. State the advantages of bottom-up chart parser compared to top-down parsing. **6**
OR iii. What is Perplexity? Suppose a language model assigns the following conditional n-gram probabilities to a 3-word test set: 1/4, 1/2, 1/4. Then P(test-set) = 1/4 * 1/2 * 1/4 = 0.03125. Calculate the perplexity. **6**

Q.6 Attempt any two:
i. What is sentiment analysis? Explain the types of sentiment analysis. **5**
ii. Explain word sense disambiguation. How to evaluate WSD? **5**
iii. What is text classification? Give an example. **5**

******

Marking Scheme
IT3EA06 Natural Language Processing

Q.1	i)	What is the field of Natural Language Processing (NLP)? d) All of the mentioned	**1**
	ii)	What is Machine Translation? a) Converts one human language to another	**1**
	iii)	What is Morphological Segmentation? b) Separate words into individual morphemes and identify the class of the morphemes	**1**
	iv)	Rule-based POS taggers does not possess which of the following properties a)The rules in Rule-based POS tagging are built auto	**1**
	v)	The study of the sound patterns in natural language and the rules that govern them is: c) Phonology	**1**
	vi)	Where does the baye's rule can be used? d) Answering probabilistic query	**1**
	vii)	N-grams are defined as the combination of N keywords together. How many bi-grams can be generated from the given sentence: ***Gandhiji is the father of our nation*** b) 6	**1**
	viii)	The statement "Which team won the match?" can be represented as b) S->Wh-NP VP	**1**
	ix)	Which of the following algorithms is widely used for text classification? d) All of the above	**1**
	x)	N-Gram language models cannot be used for _____. c)Removing semantic ambiguity	**1**
Q.2	i.	What is Natural Language Processing? Discuss approach of NLP. Definition of NLP – 1 Mark Approaches – 3 Marks	**4**
	ii.	Define ambiguity. Describe various types with example. Definition – 1 Mark	**6**
		Types – 5 Marks	
OR	iii.	Explain different phases in natural language processing with diagram. Diagram – 1 Mark Description of each phase – 5 Marks	**6**
Q.3	i.	Define Morphemes, Morphology and Corpus. Definition of each – 2 Marks	**2**
	ii.	Explain *BOW* and *TF-IDF Vectorizer* methods of Feature Extraction. Definition of each – 2 Marks Example of each – 2 Marks	**4**
	iii.	Explain POS tagging with example and also write its types. Definition of POS – 1 Mark Types – 3 Marks	**4**
OR	iv.	Differentiate between Inflectional and Derivational Morphology. Definition of Inflectional Morphology – 2 Marks Definition of Derivational Morphology – 2 Marks	**4**
Q.4	i.	Define the term Phonetics with their types. Definition – 2 Marks Types – 3 Marks	**5**
	ii.	What is the spelling error? What are their types? Definition – 1 Mark Types – 4 Marks	**5**
OR	iii.	Explain Minimum Edit Distance algorithm and also find the Minimum Edit Distance between EXECUTION and INTENTION. Definition – 2 Marks Problem solution – 2 Marks Correct Answer – 1 Mark Minimum Edit Distance = 8	**5**

			
Q.5	i.	Explain the terms Smoothing and Backoff. Definition of each term – 2 Marks (2 + 2 = 4 Marks)	**4**
	ii.	State the advantages of bottom-up chart parser compared to top-down parsing. Definition of Bottom-up parser – 2 Marks Definition of Top-down parser – 2 Marks Comparison – 2 Marks	**6**
OR	iii.	What is Perplexity? 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(test-set) = 1/4 * 1/2 * 1/4 = 0.03125. What is the perplexity? Definition – 2 Marks Problem solution – 2 Marks Correct Answer – 2 Marks Solution: Given, |w| = 3, P(test-set) = 0.03125 $pp(w) = \sqrt[\|w\|]{\frac{1}{P(w_1, w_2 \dots w_{\|w\|})}} = \sqrt[\|3\|]{\frac{1}{0.03125}} = 3.175$	**6**
Q.6		Attempt any two:	
	i.	What is Sentiment Analysis? Explain the types of Sentiment Analysis. Definition – 2 Marks Types – 3 Marks	**5**
	ii.	Explain Word Sense Disambiguation and how to evaluate WSD. Definition – 2 Marks	**5**

		Evaluation of WSD – 3 Marks	
	iii.	What is Text Classification? Give an example. Definition – 2 Marks Example – 3 Marks	**5**

******