

- Q.6 Attempt any two:
- Explain the process of sentiment analysis and discuss its importance in various real-world applications. Provide examples of how sentiment analysis can be used in business, social media, and customer feedback analysis. **5**
 - Discuss the challenges involved in word sense disambiguation and describe common techniques used to address these challenges. Provide examples to illustrate how word sense disambiguation can improve the accuracy of natural language processing tasks such as information retrieval and machine translation. **5**
 - Explain the concept of text classification and discuss its role in organizing and categorizing large volumes of textual data. Describe common approaches to text classification, such as supervised learning algorithms and deep learning models. **5**

Total No. of Questions: 6

Total No. of Printed Pages:4

Enrollment No.....



Faculty of Engineering
End Sem Examination May-2024
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. Assume suitable data if necessary. Notations and symbols have their usual meaning.

- Q.1
- What is the main approach of Natural Language Processing (NLP)? **1**
(a) Statistical analysis (b) Symbolic rules
(c) Neural networks (d) All of these
 - Which of the following best describes regular expression? **1**
(a) A method for defining patterns in text data
(b) A programming language for AI development
(c) An algorithm for speech recognition
(d) A model for sentiment analysis
 - Which technique is used to break down text into smaller meaningful units such as words or subwords? **1**
(a) Morphological parsing (b) Tokenization
(c) Feature extraction (d) Part of speech tagging
 - Which approach to Part of Speech (POS) tagging relies on predefined linguistic rules and patterns? **1**
(a) Rule-Based POS Tagging
(b) Stochastic POS Tagging
(c) Transformation-Based Tagging
(d) None of these
 - Which method in speech processing utilizes Bayesian principles to model and correct spelling errors in text? **1**
(a) Spelling error
(b) Minimum edit distance
(c) Bayesian method to spelling
(d) Bayesian method of pronunciation variation

[2]

- vi. Which model in speech processing is commonly used to capture the uncertainty and variability in speech data? **1**
 (a) Phonological rules and transducer
 (b) Probabilistic models
 (c) Minimum edit distance
 (d) Bayesian method of pronunciation variation
- vii. What is used to measure the uncertainty or unpredictability of a language model? **1**
 (a) N-Grams (b) Perplexity
 (c) Smoothing (d) Entropy
- viii. Which technique is used to handle the problem of unseen N-Grams in language models? **1**
 (a) Simple N-Gram (b) Smoothing
 (c) Parsing (d) Treebank
- ix. Which application in natural language processing focuses on determining the sentiment or emotional tone of a piece of text? **1**
 (a) Sentiment analysis (b) Spelling correction
 (c) Machine translation (d) Text classification
- x. Which task involves determining the intended meaning of a word based on its context in a sentence? **1**
 (a) Sentiment analysis
 (b) Word sense disambiguation
 (c) Text classification
 (d) Question answering system
- Q.2 i. Define ambiguity in the context of Natural Language Processing (NLP). **2**
 ii. Compare and contrast the main approaches of Natural Language Processing (NLP), namely statistical analysis, symbolic rules, and neural networks. **3**
 iii. Discuss the importance of formal languages and natural languages in the context of NLP, highlighting their differences and how they intersect. **5**
- OR iv. Explain the role of models and algorithms in the field of Natural Language Processing (NLP), highlighting their significance in addressing language-related challenges. **5**

[3]

- Q.3 i. Explain the concept of tokenization in text preprocessing. **2**
 ii. Discuss the significance of Finite State Transducers (FSTs) in morphological analysis and parsing, highlighting their role, advantages, and applications in natural language processing. **8**
- OR iii. Compare and contrast three approaches to Part of Speech (POS) tagging: Rule-Based POS Tagging, Stochastic POS Tagging, and Transformation-Based Tagging. Discuss the underlying principles, advantages, limitations, and typical applications of each approach in natural language processing. **8**
- Q.4 i. Explain the concept of minimum edit distance in the context of speech processing and its application in tasks such as speech recognition or spelling correction. **3**
 ii. Discuss the application of Bayesian methods in modeling and addressing pronunciation variation in speech processing. Provide examples of how Bayesian principles are used to handle uncertainty and variability in spoken language. **7**
- OR iii. Discuss the role of probabilistic models in speech processing, highlighting their significance in tasks such as speech recognition and phonetic analysis. Provide examples of probabilistic models commonly used in these applications and explain how they address the challenges of variability and uncertainty in speech data. **7**
- Q.5 i. Explain the concept of entropy in the context of NLP language models and its relationship with the predictability of a text. Provide examples to illustrate how entropy is calculated and interpreted in N-Gram models. **4**
 ii. Compare and contrast the concepts of smoothing and backoff in N-Gram language models. Discuss their respective roles in addressing the problem of data sparsity and improving the performance of language models. Provide examples to illustrate their usage in practical applications. **6**
- OR iii. Describe the significance of treebanks in statistical parsing and natural language processing. Discuss how treebanks are constructed, annotated, and utilized in training and evaluating parsing algorithms. Provide examples of commonly used treebanks and explain their role in advancing research in syntactic analysis and parsing. **6**

Marking Scheme

NATURAL LANGUAGE PROCESSING (NLP) IT3EA06

Q.1	i)	D) All of the above	1
	ii)	A) A method for defining patterns in text data	1
	iii)	B) Tokenization	1
	iv)	A) Rule-Based POS Tagging	1
	v)	C) Bayesian Method to Spelling	1
	vi)	B) Probabilistic Models	1
	vii)	B) Perplexity	1
	viii)	B) Smoothing	1
	ix)	A) Sentiment Analysis	1
	x)	B) Word Sense Disambiguation	1
Q.2	i.	Define ambiguity in the context of Natural Language Processing (NLP).	2
	ii.	Compare and contrast the main approaches of Natural Language Processing (NLP), namely statistical analysis, symbolic rules, and neural networks.	3
		1 Mark each	
	iii.	Discuss the importance of formal languages and natural languages in the context of NLP, highlighting their differences and how they intersect.	5
OR		Importance	2.5 Marks
		Differences	2.5 Marks
	iv.	Explain the role of models and algorithms in the field of Natural Language Processing (NLP), highlighting their significance in addressing language-related challenges.	5
		Role of model and algorithm	2.5 Marks
Q.3		Addressing	2.5 Marks
	i.	Explain the concept of tokenization in text preprocessing.	2
		2 Marks	
	ii.	Discuss the significance of Finite State Transducers (FSTs) in morphological analysis and parsing, highlighting their role, advantages, and applications in natural language processing.	8
		FST	4 Marks
		Advantage	2 Marks

OR	iii.	Application	2 Marks	8
		Compare and contrast three approaches to Part of Speech (POS) tagging: Rule-Based POS Tagging, Stochastic POS Tagging, and Transformation-Based Tagging. Discuss the underlying principles, advantages, limitations, and typical applications of each approach in natural language processing.		
		Three approaches	3 Marks	
		Principle advantage	2 Marks	
Q.4	i.	Applications	3 Marks	3
		Explain the concept of Minimum Edit Distance in the context of speech processing and its application in tasks such as speech recognition or spelling correction.		
		Concept	1.5 Marks	
		Application	1.5 Marks	
OR	ii.	Discuss the application of Bayesian methods in modeling and addressing pronunciation variation in speech processing. Provide examples of how Bayesian principles are used to handle uncertainty and variability in spoken language.		7
		Application of Bayesian	3.5 Marks	
		Example	3.5 Marks	
		Discuss the role of probabilistic models in speech processing, highlighting their significance in tasks such as speech recognition and phonetic analysis. Provide examples of probabilistic models commonly used in these applications and explain how they address the challenges of variability and uncertainty in speech data.		
Q.5	i.	Probabilistic models	3 Marks	4
		Significance	2 Marks	
		Applications	2 Marks	
		Explain the concept of entropy in the context of NLP language models and its relationship with the predictability of a text. Provide examples to illustrate how entropy is calculated and interpreted in N-Gram models.		
	ii.	Concept	1.5 Marks	6
		Entropy calculation	2.5 Marks	
		Compare and contrast the concepts of smoothing and backoff in N-Gram language models. Discuss their respective roles in		

[2]

addressing the problem of data sparsity and improving the performance of language models. Provide examples to illustrate their usage in practical applications.

Comparison 2 Marks
Roles in addressing data sparsity 3 Marks.
Example 1 Marks

- OR iii. Describe the significance of Treebanks in statistical parsing and natural language processing. Discuss how Treebanks are constructed, annotated, and utilized in training and evaluating parsing algorithms. Provide examples of commonly used Treebanks and explain their role in advancing research in syntactic analysis and parsing. **6**

Significance of treebanks 2 Marks
Treebank construction 3 Marks
Example 1 Marks

- Q.6 i. Attempt any two:
i. Explain the process of sentiment analysis and discuss its importance in various real-world applications. Provide examples of how sentiment analysis can be used in business, social media, and customer feedback analysis. **5**

Process 2.5 Marks
Example 2.5 Marks

- ii. Discuss the challenges involved in word sense disambiguation and describe common techniques used to address these challenges. Provide examples to illustrate how word sense disambiguation can improve the accuracy of natural language processing tasks such as information retrieval and machine translation. **5**

Word sense disambiguation and techniques 3 Marks
Example 2 Marks

- iii. Explain the concept of text classification and discuss its role in organizing and categorizing large volumes of textual data. Describe common approaches to text classification, such as supervised learning algorithms and deep learning models. **5**

Text classification 1.5 Marks
Its role 1 Marks
Approaches 2.5 Marks

[3]