Printed Page:-04		Subject Co	Subject Code:- ACSAI0712						
N									
INC	$\frac{1}{(\Delta n)}$	E OF ENGINEERING	G AND IECH Affiliated to A	NOLO	GY, Gf Luckno	(EAII)	ER NU	IDP	A
	(7 11 2	E E	B.Tech	, in the second sec	LUCKIO	•••)			
	SEN	A: VII - THEORY EX	XAMINATION	J (2024	- 2025)	)			
		Subject: Natural I	Language Proc	essing					
Time	3 Hours					N	/lax. M	lark	s: 100
General	Instructions:	a manaimad tha muantia		1			de has		
IMP: Vo	Prify that you have Question paper ce	e received the questio	n paper with th	ne corre C-It co	ect cour nsists o	rse, coa f Multi	ae, bra nl <i>e</i> Ch	ncn	<i>etc</i> .
1. 1nts Ouestio	ns (MCO's) & Su	biective type question	ions 11, D, Q .	<i>c. n co</i>	1131313 0	1111111	pie Ch	one	~
2. Maxi	mum marks for ea	ach question are indic	ated on right -	hand st	ide of ea	ach qu	estion.		
3. Illust	rate your answer.	s with neat sketches w	wherever neces.	sary.					
4. Assu	ne suitable data į	f necessary.							
5. Prefe	rably, write the a	nswers in sequential o	order.	11 1	1 , •	11 . 1	,		
0. NO SI	eet should be left d/chackad	t blank. Any written m	aterial after a	blank s	sheet wi	ll not l	<i>pe</i>		
evalual	eu/checkeu.								
<u>SECTI</u>	<u>ON-A</u>								20
1. Atter	npt all parts:-								
1-a.	Why is stop-we	ord removal important	t in NLP?(CO	1,K2)					1
	(a) To increase	the length of text							
	(b) To remove	words that don't carry	significant me	eaning					
	(c) To add emp	hasis to certain words							
	(d) To identify	important keywords							
1-b.	What does POS	S tagging stand for in	NLP?(CO1,K)	1)					1
	(a) Parts of Spe	ech tagging							
	(b) Position of	Sentence tagging							
	(c) Point of Sal	e tagging							
	(d) Positive or	Negative sentiment ta	gging						
1-c.	What does data	a normalization in NL	P refer to? (CC	D2,K1)					1
	(a) a) Convertin	ng text to lowercase	, , , , , , , , , , , , , , , , , , ,						
	(b) b) Handling	y email IDs							
	(c) c) Reducing	repeated characters							
	(d) d) Transfor	ming text to a commo	n standard						
1-d.	How can you h	andle URLs in text da	ata preprocessi	ng? (C	O2,K2)				1
	(a) a) Replacin	g with "URL"	1 1 1 1 1 1 1 1		, ,				_
	(b) b) Removin	ig them							
	(b) b) Removin	ig them							

(c) c) Converting to uppercase

•

	(d)	d) Standardizing them				
1-e.	(	Cosine similarity measures the cosine of the angle between two vectors. What loes a cosine similarity of 1 indicate? (CO3,K2)	1			
	(a)	Perfect dissimilarity				
	(b)	Perfect similarity				
	(c)	No relationship				
	(d)	Undefined relationship				
1 <b>-</b> f.	,	What is the Word Mover's Distance used for?(CO3,K1)	1			
	(a)	Calculating the similarity between documents				
	(b)	Measuring the distance between words in a sentence				
	(c)	Evaluating the performance of Word2Vec				
	(d)	Assessing the relevance of terms in a document				
1-g.	t	What is the role of a labeled dataset in training a sentiment analysis model using ext classification?(CO4,K2)	1			
	(a)	It helps in determining the font style of text.				
	(b)	It provides training data with known sentiment labels to build and evaluate the				
	(c)	It is used to convert text to audio				
	(d)	It assists in counting the number of paragraphs in a document				
1-h	(u) 1	Which popular topic modeling algorithm is often used to discover topics in a text	1			
1 11.	(	corpus?(CO4,K2)	1			
	(a)	Linear regression				
	(b)	K-means clustering				
	(c)	Latent Dirichlet Allocation (LDA)				
	(d)	Support Vector Machine (SVM)				
1-i.	1	What is the significance of word embeddings in NLP?(CO5,K1)				
	(a)	Representing words as vectors				
	(b)	Counting the frequency of words				
	(c)	Removing stop words				
	(d)	Tokenizing sentences				
1-j.	1	What is BERT in the context of transformer-based models?(CO5,K1)	1			
	(a)	Bidirectional Encoder Representations from Transformers				
	(b)	Binary Encoder Regression Transformer				
	(c)	Bidirectional Extractor for Recurrent Text				
	(d)	Basic Encoder for Recursive Text				
2. Atte	empt	all parts:-				
2.a.	]	Explain how contextual information can help resolve ambiguity.(CO1,K2)	2			
2.b.	]	Define how morphology help in understanding the evolution of	2			

## Page 2 of 4

•

	languages?(CO2,K2)		
2.c.	How do vector space models represent documents in a high-dimensional space?(CO3,K2)		
2.d.	Explain the concept of text classification in NLP(CO4,K2)	2	
2.e.	Discuss the advantages of using Hugging Face Transformers for working with transformer-based models such as BERT and GPT.(CO5,K1)	2	
<b>SECTIO</b>	<u>N-B</u>	30	
3. Answe	r any <u>five</u> of the following:-		
3-а.	Discuss the purpose of tokenization in NLP?(CO1,K2)	6	
3-b.	Discuss how NLP relate to other fields in artificial intelligence, and what distinguishes it as a distinct discipline?(CO1,K2)		
3-с.	Discuss the difference between semantics and syntax in linguistic analysis?(CO2,K2)	6	
3-d.	Describe different strategies for handling URLs in text data preprocessing and discuss the trade-offs between these strategies.(CO2,K2)		
3.e.	Explain the Bag-of-Words model and how it represents documents in text vectorization. Discuss its advantages and limitations.(CO3,K2)		
3.f.	Explain the techniques commonly employed in spam detection using natural language processing. How can machine learning models be trained to effectively identify spam messages? (CO4,K2)	6	
3.g.	Examine the limitations of traditional bag-of-words models in capturing sequential dependencies in text. Propose alternative approaches or techniques that address these limitations.(CO5,K3)	6	
<u>SECTIO</u>	<u>N-C</u>	50	
4. Answe	r any <u>one</u> of the following:-		
4-a.	Analyze the applications of NLP in sentiment analysis and its significance in understanding public opinion and market trends.(CO1,K3)	10	
4-b.	Elaborate on the concept of lemmatization and how it differs from stemming in terms of linguistic accuracy and applications. (CO1,K3)	10	
5. Answe	r any <u>one</u> of the following:-		
5-a.	Elaborate step-by-step guide to a comprehensive data preprocessing pipeline using Python, including the handling of email IDs, HTML tags, URLs, and emojis.(CO2,K3)	10	
5-b.	Explain the concept of word embeddings and their role in capturing word semantics. Compare and contrast word embeddings with traditional vector space models.(CO2,K3)	10	
6. Answe	r any <u>one</u> of the following:-		
б-а.	Provide a detailed outline of the training process for Word2Vec in generating word embeddings (CO3 K3)	10	
	(cos, ic)		

•

word structure, inflections, and morphological typology. Provide examples of languages with different morphological characteristics.(CO3,K2)

- 7. Answer any one of the following:-
- 7-a. Explain the process of implementing sentiment analysis using text classification. 10
  Provide an example scenario and discuss the significance of sentiment analysis in applications like social media monitoring or customer feedback analysis. (CO4,K2)
- 7-b. Explore the potential applications of text summarization in legal document
  10 analysis. Discuss how summarization techniques can facilitate efficient
  information retrieval and decision-making in the legal domain.(CO4,K3)

8. Answer any one of the following:-

- 8-a. Describe the architecture of BERT (Bidirectional Encoder Representations from 10 Transformers). Discuss the pre-training objectives, tokenization strategies, and how BERT captures bidirectional contextual information in language understanding tasks.(CO5,K3)
- 8-b. Provide an in-depth overview of the GPT (Generative Pre-trained Transformer) 10 model architecture. Discuss the autoregressive nature of GPT, its training objectives, and how it generates coherent and contextually relevant text.(CO5,K3)