

	<b>VIT-AP</b>	<b>Continuous Assessment Test – Winter Sem(2024-25) -JAN 2025</b>	
UNIVERSITY		Maximum Marks: 50	Duration: <b>90 Mins</b>
Course Code: <b>CSE3015</b>			Course Title: <b>Natural Language Processing</b>
Set No: 5	Exam Type : <b>Closed Book</b>		School: <b>SCOPE</b>
Date: <b>31/01/2025</b>	Slot: E1+TE1		Session: <b>FJ</b>

**Keeping mobile phone/smart watch, even in ‘off’ position is treated as exam malpractice**

**General Instructions if any Open Book/Open Notebook/Closed Book:**

1. “fx series” - non Programmable calculator are permitted : YES
2. Reference tables permitted : NO

**Answer any ALL Questions, Each Question Carries 10 Marks ( $5 \times 10 = 50$  Marks)**

- Q1. Compare and contrast lemmatization and stemming in NLP. Why would one prefer lemmatization over stemming in a sentiment analysis task? Demonstrate lemmatization on the following sentence using Python: *"The cats were chasing the mice, and the mice were hiding in the small holes."* (10M)
- Q2. Compute the TF-IDF matrix for the following set of product reviews, along with a python script. Briefly discuss the limitations of TF-IDF. (10M)
- 1. This phone has an excellent camera.
  - 2. The battery life of this laptop is outstanding.
  - 3. The sound quality of these headphones is amazing.
- Q3. a. Describe the architecture of CBOW in Word2Vec embedding's. Discuss its limitations (7M)  
 b. Differentiate FastText from traditional word embedding models like Word2Vec and GLOVE (3M)
- Q4. a. Design a regular expression tagger to tag parts of speech of the following sentence using necessary regular expression patterns and print the output (7M)
- Sentence: [Tendulkar took up cricket at the age of eleven, made his test match debut on 15 November 1989 at the age of sixteen.]
- b. Identify and categorize the named entities of the following text into one of the five groups: [Person, Organization, Time, Location, Work of Art]. (3M)
- Content:** Leonardo da Vinci was an Italian polymath of the Renaissance period. He painted the *Mona Lisa*, one of the most famous works of art, and was born in Vinci, Italy, in 1452.

Da Vinci also contributed to science and engineering, leaving behind numerous sketches and notes.

- Q5. Compare the working procedure of Unigram, Bigram and Trigram taggers to tag any given string token and illustrate the working procedure with a code snippet that tags the words of any given sentence of your choice. **(10M)**

#### QP MAPPING

Q. No.	E/A/T	Module Number	Marks	BL	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped
Q1	A	1	10	3	CO1	PO1, PO2, PO3	PEO1, PEO3	
Q2	E	1	10	2	CO1	PO1, PO2, PO3	PEO1, PEO3	
Q3	T	1	10	2	CO1	PO1, PO2, PO3	PEO1, PEO3	
Q4	A	2	10	3	CO2	PO2, PO3	PEO1, PEO3	
Q5	A	2	10	3	CO2	PO2, PO3	PEO1, PEO3	

	<b>VIT-AP</b>	<b>Continuous Assessment Test – Winter Sem(2024-25) -JAN 2025</b>			
UNIVERSITY	Maximum Marks: 50		Duration: 90 Mins		
Course Code: CSE3015	Course Title: Natural Language Processing				
Set No: 6	Exam Type : Closed Book		School: SCOPE		
Date: 31-01-25	Slot: E2+TE2		Session: An		
<b>Keeping mobile phone/smart watch, even in ‘off’ position is treated as exam malpractice</b>					
<b>General Instructions if any Open Book/Open Notebook/Closed Book:</b>					
1. “fx series” - non Programmable calculator are permitted: YES 2. Reference tables permitted: NO					

**PART – A: Answer any ALL Questions, Each Question Carries 10 Marks ( $5 \times 10 = 50$  Marks)**

1. Develop a python program using NLTK to preprocess a list of sentences by: **(10 M)**
  - a. Tokenizing each sentence into individual words.
  - b. Remove stop words.
  - c. Applying stemming on the remaining words
2. Create vector representations for the following documents using One-hot encoding and BOW. **(10M)**

D1: Dog bites man.  
 D2: Man bites dog.  
 D3: Dog eats meat.  
 D4: Man eats food.
3. How does the Skip-gram architecture in the Word2Vec model transform a given text into vector representations? **(10M)**
4. What is an N-gram tagger, and how N-gram tagger assigns part-of-speech (POS) tags to words in a sentence? **(10M)**
5. Create a Named Entity Recognition (NER) tagger that identifies and labels **Person Names, Organization Names, Location Names, and Date Expressions** in a given text. **(10M)**

**Input:** "John Doe, the CEO of Google, visited New York on January 15, 2025.  
**Output:** [(John Doe, PERSON), (Google, ORGANIZATION), (New York, LOCATION), (January 15, 2025, DATE)]

 <b>VIT-AP</b> UNIVERSITY	<b>Continuous Assessment Test –1- WINTER semester (2024-25) – January-2025</b>				
	Maximum Marks: 50		Duration: 90 Mins		
Course Code: MEC3015	Course Title: Advanced Machining Processes				
Set No: 1	Exam Type : Closed Book		School: SMEC		
Date: 30/01/25	Slot: D1	Session: FN			
<b>Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice</b>					
<b>General Instructions if any Open Book/Open Notebook/Closed Book:</b>					
1. "fx series" - non Programmable calculator are permitted : Yes 2. Reference tables permitted : Not applicable					

**PART – A: Answer ALL Questions, Each Question Carries 10 Marks (5×10=50 Marks)**

- How does the addition of abrasive materials in abrasive water jet machining enhance the cutting process, and what types of materials are typically used as abrasives? Describe. **(10 M)**
- Explain the formation of jet in abrasive water jet machining (AWJM). Describe the advantages and limitation of AWJM. **(10 M)**
- Explain magnetic abrasive finishing process in detail. **(10 M)**
- Suggest a cost effective advanced machining method for cutting aluminium in aerospace application. Describe the process. **(10 M)**
- Electropolishing is done to improve the surface finish. Justify this statement based on the process details of electropolishing. **(10 M)**

**OP MAPPING**

Q. No.	E/A/T	Module Number	Marks	BL	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped
Q1	T	1	10	5	1	1	1	1
Q2	T	1	10	5	1	1	1	1
Q3	T	1	10	6	1	1	1	1
Q4	T	2	10	5	2	1	1	1
Q5	T	2	10	6	1	1	1	1

	<b>Continuous Assessment Test – Winter Sem(2024-25) -JAN 2025</b>	
	Maximum Marks: 50	Duration: 90 Mins
Course Code: CSE3015	Course Title: Natural Language Processing	
Set No: 3	Exam Type : Closed Book	School: SCOPE
Date: 29-01-25	Slot: C1+TC1	Session: F
<b>Keeping mobile phone/smart watch, even in ‘off’ position is treated as exam malpractice</b>		
<b>General Instructions if any Open Book/Open Notebook/Closed Book:</b>		
1. “fx series” - non Programmable calculator are permitted: YES		
2. Reference tables permitted: NO		

**PART – A: Answer any ALL Questions, Each Question Carries 10 Marks ( $5 \times 10 = 50$  Marks)**

- Explain briefly why Natural Language Understanding is hard? Explain any three types of ambiguities in NLP with an example for each ambiguity. **(10 M)**
- What is a typical Text Processing pipeline for processing a web page. Explain different tokenization challenges briefly with WhatsApp / Twitter data? **(10 M)**
- Explain the concepts of Stemming and Lemmatization with examples. What are their advantages and disadvantages? **(10 M)**
- A sample corpus is given in below:  
**<S> As you like it is a play of Shakespeare </S>**  
**<S> Do you like it ? </S>**

Answer the following questions:

- What is the total count of unigrams (without <S> and </S> tags)
  - What is the total count of bigrams (including <S> and </S> tags)
  - What is the total count of trigrams (including <S> and </S> tags)
  - What is the total count of 4-grams (including <S> and </S> tags)
  - Find unigram probabilities for: P(like) and P(play)
  - Find conditional probabilities for: P (like | it) and P (it | like)
  - Find bigram probability for: P (you like)
  - Find trigram probability for: P (you like it)
  - Find 4-gram probability for: P (As you like it) **(10 M)**
- Give examples of any three POS Taggers? Explain in detail how Brill Tagger works? **(10 M)**

	<b>VIT-AP UNIVERSITY</b>	<b>Continuous Assessment Test – Winter Sem(2024-25) -JAN 2025</b>			
Maximum Marks: 50		Duration: 90 Mins			
Course Code: CSE3015		Course Title: Natural Language Processing			
Set No: 4	Exam Type : <b>Closed Book</b>		School: SCOPE		
Date: 29-01-25	Slot: C2+TC2		Session: <i>An</i>		
<b>Keeping mobile phone/smart watch, even in ‘off’ position is treated as exam malpractice</b>					
<b>General Instructions if any Open Book/Open Notebook/Closed Book:</b>					
1. “fx series” - non Programmable calculator are permitted : Yes 2. Reference tables permitted : YES / NO (if Yes, please specify: No )					

**PART – A: Answer any ALL Questions, Each Question Carries 10 Marks (5×10=50 Marks)**

1. What are the goals of preprocessing techniques in NLP? Describe the different pre-processing techniques used in NLP by providing examples of each. **(10 M)**
2. Explain the concept of one-hot encoding in word embeddings. How does it differ from other encoding techniques like Bag of Words (BoW) and TF-IDF? **(10 M)**
3. Implement a python program using Regex Tagger for the following sentence **“Book the flight” and “I like to read NLP book”.**
  - a. Function should take a sentence as input and return a dictionary, where the keys are words and values are their corresponding tags.
  - b. Modify the same function to handle punctuations and parenthesis **(10 M)**
4. Explain the following.
  - a. Describe the training process of FastText with a C-BOW model by considering trigram for the sentence -  
*“I like NLP and therefore, I want to learn NLP”.*
  - b. list the differences between GloVe and FastText, and Word-to-Vector. **(10 M)**
5. Given the sentence – “OpenAI CEO Sam Altman has dismissed rumours about the imminent deployment of Artificial General Intelligence (AGI). He started his company in late 2000’s and now clarified that the company is not close to achieving AGI and urged people to lower their expectations. Recently he met many CEOs’ and few of them are Bill Gates, who founded Microsoft in 1970’s, Sundar Pichai, Google which was founded in 1989.”
  - a. Implement python program to extract all the entities based on NER tagging.
  - b. Display the output with clear annotated entities. **(10 M)**