



RV College of Engineering\*

 Autonomous Institutions Affiliated to Vignansaraya Technological University, Belagavi  
 Approved by AICTE New Delhi

Academic Year 2023-2024 (Odd Semester)

ADM NO 1 R V 2 1 A I 0 5 2

## Department of Artificial Intelligence and Machine Learning

Course Code: 21AI53

Date: 22/03/2024

Semester: V Semester

Duration: 90 Minutes

Maximum Marks: 50

### Natural Language Processing and Transformers CIE 1 RE-TEST

Note: Answer all the Questions

SL. No	Questions	M	BT	CO
1	a) Explain the key stages in a typical NLP pipeline, and how the flow of data evolves through these stages.	5	2	1
	b) Discuss the various building blocks of a language with a neat diagram	5	2	1
2	a) Illustrate with neat diagrams various Machine Learning approaches used to solve an NLP task.	5	2	1
	b) Describe the various challenges associated with using deep learning for NLP task and the ways to mitigate them.	5	2	1
3	a) Explain stop-word filtration in a tokenized sentence with an example code	5	2	2
	b) Describe the concept of synset and ways to find synset similarity	5	2	2
4	a) Explain different ways to tokenize a text using wordnet with an example	5	2	2
	b) Illustrate with an example the need for spelling correction in the NLP pipeline and find the minimum distance between the words "INTENTION" and "EXECUTION".	5	3	2
5	a) Analyze different types of relation possible between two words with examples for each.	7	4	1
	b) Explain Lemmatization and stemming using Wordnet with pseudo code. Why is Lemmatization/stemming needed in NLP pipeline?	3	2	1

**Course Outcomes: After completing the course, the students will be able to**

CO1:	Discuss various concepts, architectures and frameworks of NLP
CO2:	Proficiency in utilizing the core and popular NLP libraries to provide solutions to real-world applications in Healthcare, Smart Cities, Agriculture, etc
CO3:	Design and Develop agents that use Transformers for natural language understanding and generation
CO4:	Demonstrate the use of modern tools in solving day-to-day problems by exhibiting teamwork through oral presentations and reports
CO5:	Collaborate in a group to build NLP solutions for the benefit of society

**M-Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes**

Marks Distribution	Particulars	CO1	CO2	CO3	CO4	CO5	L1	L2	L3	L4	L5	L6
	Max Marks CIE	30	20	-	-	-	-	38	5	7	-	-



# Department of Artificial Intelligence and Machine Learning

Date: 09/01/2024

Course Code: 21AI53  
Semester: V Semester  
Duration: 90 Minutes

Maximum Marks: 50

## Natural Language Processing and Transformers CIE 1

Note: Answer all the Questions

SL. No	Questions	M	BT	CO
1	a) Explain the key stages in a typical NLP pipeline, and how the flow of data evolves through these stages.	5	2	1
	b) Discuss the challenges associated with developing an NLP task	5	2	1
2	a) Illustrate with neat diagrams various Deep Learning approaches used to solve an NLP task.	5	2	1
	b) why are interpretable models and overfitting a challenge for using deep learning in NLP and describe the ways to mitigate them.	5	2	1
3	a) Apply the Decision-list algorithm and Walker's algorithm to determine the sense of the word "star" in the given sentence "He always wanted to be a Bollywood star".	5	3	1
	b) Applying the Leacock Chordorow algorithm, explain how to find Word similarity in Wordnet.	5	3	2
4	a) Illustrate different ways to tokenize a text using Wordnet with an example code for each.	5	3	2
	b) Illustrate with an example the need for spelling correction in the NLP pipeline and how is it done using the edit distance method. (eg: let the word entered be "behaf" its correct word can be either "behave or behalf". How to find it)	5	3	2
	a) Analyze different types of relation possible between two words with examples for each.	7	4	
	b) Explain Lemmatization and stemming using Wordnet with pseudo code. Why is Lemmatization/stemming needed in NLP pipeline?	3	2	

Course Outcomes: After completing the course, the students will be able to

1:	Discuss various concepts, architectures and frameworks of NLP
2:	Proficiency in utilizing the core and popular NLP libraries to provide solutions to real-applications in Healthcare, Smart Cities, Agriculture, etc
3:	Design and Develop agents that use Transformers for natural language understanding and generation
4:	Demonstrate the use of modern tools in solving day-to-day problems by exhibiting team through oral presentations and reports
5:	Collaborate in a group to build NLP solutions for the benefit of society

Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes

Marks Distribution	Particulars	CO1	CO2	CO3	CO4	CO5	L1	L2	L3	L4	L5
	Max	30	20	-	-	-	-	23	7	20	-



# Department of Artificial Intelligence and Machine Learning

Course Code: 21AI53  
Semester: V Semester  
Duration: 90 Minutes

Date: 22/02/2024

Natural Language Processing and Transformers  
CIE 2

Maximum Marks: 50

Note: Answer all the Questions

SL. No	Questions	M	CO	BT
1	a) Explain the steps involved in creating a categorized text corpus in NLTK?	5	2	2
	b) Identify the need of sequential backoff tagger. Explain in detail with example	5	2	3
2	a) Explain the concept of part-of-speech tagging and discuss two different approaches to POS tagging.	5	1	2
	b) Develop a PoS tagger to tag the names of an individual as proper nouns.	5	2	3
3	a) Discuss in detail transfer learning and supervised learning with a neat diagram.	5	1	2
	b) Describe a scenario where Hugging Face Tokenizers' sub-word tokenization would be more beneficial than word tokenization for text processing tasks.	5	3	2
4	a) Explain the role of attention mechanisms in the encoder-decoder framework of neural machine translation.	5	1	2
	b) Discuss the main challenges faced when using Hugging face models for text generation tasks and propose potential solutions to address them.	5	1	2
5	a) Develop a pre-trained transformer model for a specific natural language processing task such as Text generation.	7	3	3
	b) How can Hugging Face Datasets be utilized to analyze the maximum context size for text classification tasks? Explain with the code snippet and an example dataset with some rows of text data.	3	2	1

Course Outcomes: After completing the course, the students will be able to	
C01:	Discuss various concepts, architectures and frameworks of NLP
C02:	Proficiency in utilizing the core and popular NLP libraries to provide solutions to real-world applications in Healthcare, Smart Cities, Agriculture, etc
C03:	Design and Develop agents that use Transformers for natural language understanding and generation
C04:	Demonstrate the use of modern tools in solving day-to-day problems by exhibiting teamwork through oral presentations and reports
C05:	Collaborate in a group to build NLP solutions for the benefit of society

M-Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes

Marks Distribution	Particulars	C01	C02	C03	C04	C05	L1	L2	L3	L4	L5	L6
	Max Marks CIE	20	18	12	-	-	3	30	17	-	-	-

Go, Change the World





RV College of  
Engineering®

Go, Change the World  
Academic Year 2023-2024 (Odd Semester)  
USN 1 R V 2 1 A I

## Department of Artificial Intelligence and Machine Learning

Course Code: 21AI53  
Semester: V Semester  
Duration: 90 Minutes

Date: 20/03/2024

Maximum Marks: 50

### Natural Language Processing and Transformers CIE 3

Note: Answer all the Questions

SL. No	Questions	M	CO	BT
1	a) Explain the concept of self-attention in the Transformer architecture and how it contributes to the model's ability to understand contextual relationships.	3+2	1	2
	b) Describe the role of positional embeddings in the Transformer model and how they help capture the sequential information in the input.	5	2	2
2	a) Explain greedy search decoding and beam search decoding in the context of text generation using Transformer models.	5	1	2
	b) Discuss the challenges associated with generating coherent text using language models like GPT-2 and how techniques like beam search decoding address these challenges.	5	2	2
3	a) What are some common methods used for text summarization, and how do models like T5, BART, and PEGASUS contribute to improving the state-of-the-art in summarization tasks?	10	2	1
	b) Compare and analyze top-k and nucleus sampling techniques in text generation tasks. How do they balance between diversity and coherence in generated text?	7+3	2	
5	a) Describe the key components of the Transformer's decoder branch and how they differ from those in the encoder branch in terms of functionality and purpose.	3+2	1	
	b) Explain the role of the classification head in the Transformer architecture. How does it enable the model to perform tasks beyond sequence generation, such as classification and summarization, effectively?	3+2	2	

**Course Outcomes: After completing the course, the students will be able to**

CO1:	Discuss various concepts, architectures and frameworks of NLP
CO2:	Proficiency in utilizing the core and popular NLP libraries to provide solutions to real-applications in Healthcare, Smart Cities, Agriculture, etc
CO3:	Design and Develop agents that use Transformers for natural language understanding generation
CO4:	Demonstrate the use of modern tools in solving day-to-day problems by exhibiting team through oral presentations and reports
CO5:	Collaborate in a group to build NLP solutions for the benefit of society

**M-Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes**

Particulars	CO1	CO2	CO3	CO4	CO5	L1	L2	L3	L4	L5
-------------	-----	-----	-----	-----	-----	----	----	----	----	----



21A153

USN

RV COLLEGE OF ENGINEERING®

(An Autonomous Institution Affiliated to VTU)

V Semester B. E. Examinations April/May-2024

Artificial Intelligence and Machine Learning

## NATURAL LANGUAGE PROCESSING AND TRANSFORMERS

Time: 03 Hours

*Maximum Marks: 100*

*Instructions to candidates:*

1. Answer all questions from Part A. Part A questions should be answered in first three pages of the answer book only.
2. Answer FIVE full questions from Part B. In Part B question number 2 is compulsory. Answer any one full question from 3 and 4, 5 and 6, 7 and 8, 9 and 10.

### PART-A

M CO BT

1.1	Briefly define two common NLP tasks.	02	1	2
1.2	Name two steps involved in text extraction and cleanup.	02	1	2
1.3	What is the purpose of stemming words?	02	1	3
1.4	How do you calculate WordNet Synset similarity?	02	1	2
1.5	What do you mean by lazy corpus loading?	02	1	2
1.6	What is affix tagging and how does it work?	02	1	2
1.7	Name all components of hugging face ecosystem.	02	1	2
1.8	What is the main purpose of the attention mechanism in transformers?	02	1	2
1.9	What is the purpose of positional embeddings in the Transformer model?	02	1	2
1.10	Explain cross-attention mechanism.	02	1	2

## PART-B

a	Examine and differentiate feature engineering techniques like TF/IDF, N-Grams and One Hot Encoding. Identify scenarios where each method is more advantageous over the others.	08	2	2
b	Describe the primary phases in a standard NLP pipeline and elucidate how data progresses through these phases.	08	2	2
a	Illustrate stemming and lemmatization using WordNet corpus with an example.	08	3	3
b	Analyse different types of relation possible between two words with an example for each.	08	2	3
<b>OR</b>				
a	Apply the Leacock Chodorow algorithm to find the Word similarity in WordNet with an example.	08	3	3
b	Apply Walker's algorithm to determine the sense of the word. Explain with an example.	08	2	3
a	Discuss sequential backoff tagger with an example.	08	3	2
b	Illustrate pre-trained transformer model for a specific natural language processing task such as text generation with a relevant example.	08	3	3



<b>OR</b>			
6	a b	Describe transfer learning in detail. List its advantages. How attention mechanisms are used in the encoder-decoder framework of neural machine translation? Give the steps in detail.	08 08
	a b	Discuss complete pipeline using Hugging Face Models for text classification using feature extraction. Illustrate pre-trained transformer model for a specific natural language processing task such as text generation and discuss the steps in detail.	08 08
<b>OR</b>			
	a b	How attention mechanisms are used in the encoder-decoder framework of neural machine translation. Give an example. Describe transfer learning in detail.	08 08
	a b	Explore the complexities inherent in producing cohesive text through language models such as GPT-2 and elucidate how methodologies like beam search decoding endeavor to mitigate these obstacles. The key components of the transformer's decoder branch are different from encoder branch. Justify your answer.	08 08
<b>OR</b>			
	a b	Identify the difference in text summarization pipeline like GPT and BART with neat a diagram. Describe the use of positional embeddings in capturing the sequential information in the input.	08 08