

GANPAT UNIVERSITY									
FACULTY OF ENGINEERING & TECHNOLOGY									
Programme		Bachelor of Technology				Branch/Spec.		Computer Science & Engineering	
Semester		VII				Version		1.0.0.0	
Effective from Academic Year					2026-27	Effective for the batch Admitted in			June 2023
Subject code		2CSE716			Subject Name	Large Language Model			
Teaching scheme					Examination scheme (Marks)				
(Per week)	Lecture(DT)		Practical(Lab.)		Total		CE	SEE	Total
	L	TU	P	TW					
Credit	2	0	1	0	3	Theory	40	60	100
Hours	2	0	2	0	4	Practical	60	40	100
Pre-requisites:									
Theory of Computation, Statistics, JAVA or PYTHON programming language									
Learning Outcome:									
Upon Completion of the course, the students will be able to									
<ul style="list-style-type: none"> • Introduce the mathematical foundations of computation, including Text Processing and language modeling • Differentiate and give examples for the different types of language models, Their Application, and the embedding system. • Correlate the different types of automata to real-world applications • Choose and design an appropriate Language Model for the different requirements outlined by theoretical Computer science and Engineering. 									
Theory syllabus									
Unit	Content								Hrs
1	Introduction Natural Language Processing: What is NLP? Why NLP is Difficult? History of NLP, Advantages of NLP, Disadvantages of NLP, Components of NLP, Applications of NLP, How to build an NLP pipeline? Phases of NLP, NLP APIs, NLP Libraries.								6
2	Language Modeling and Part of Speech Tagging: Unigram Language Model, Bigram, Trigram, N-gram, Advanced smoothing for language modeling, Empirical Comparison of Smoothing Techniques, Applications of Language Modeling, Natural Language Generation, Parts of Speech Tagging, Morphology, Named Entity Recognition								6
3	Words and Word Forms: Bag of words, skip-gram, Continuous Bag-Of-Words, Embedding representations for words Lexical Semantics, Word Sense Disambiguation, Knowledge Based and Supervised Word Sense Disambiguation.								6
4	Text Analysis, Summarization and Extraction: Sentiment Mining, Text Classification, Text Summarization, Information Extraction, Named Entity Recognition, Relation Extraction, Question Answering in Multilingual Setting; NLP in Information Retrieval, Cross-Lingual IR.								6
5	Machine Translation:								6

	Application of Machine Translation, Problems of Machine Translation, MT Approaches, Direct Machine Translations, Rule-Based Machine Translation, Knowledge Based MT System, Statistical Machine Translation (SMT), Parameter learning in SMT (IBM models) using EM), Encoder-decoder architecture, Neural Machine Translation. Working of Language Model. Case Study on one Language Model.	
--	---	--

Self-Study Topics

Sentiment Analysis and Semantics Introduction, Sentiment Lexicons, Semantics, Representing and understanding meaning, First order logic, knowledge representation, inference, semantic parsing

Practical content

Practicals will be based on various text processing activities (should be implemented using NLTK)

- Program on removing Non Texts
- Program on Sentence segmentation
- Program on Stemming
- Program on Part of speech Tagging
- Program on Named Entity Recognition
- Program on Information Extraction

Text Books

1	JSpeech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition Jurafsky, David, and James H. Martin, PEARSON
---	---

Reference Books

1	Foundations of Statistical Natural Language Processing, Manning, Christopher D., and Hinrich Schütze, Cambridge, MA: MIT Press
2	Natural Language Understanding, James Allen. The Benjamin/Cummings Publishing Company Inc..

Online Resources

1	NPTEL: <ol style="list-style-type: none"> 1. http://nptel.ac.in/courses/106105158/ 2. https://www.youtube.com/playlist?list=PLD392E2ACAEF0C689
2	Coursera: https://www.coursera.org/learn/natural-language-processing/home/welcome
3	NLTK: https://www.youtube.com/playlist?list=PLQVvva0QuDf2JswnfiGkliBlnZnIC4HL
4	Stanford (YouTube): https://www.youtube.com/playlist?list=PL6397E4B26D00A269

Course Outcomes:

COs	Description
C01	Introduce the mathematical foundations of computation, including Text Processing and language modeling
C02	Differentiate and give examples for the different types of language models, Their Application, and the embedding system.
C03	Correlate the different types of automata to real-world applications
C04	Choose and design an appropriate Language Model for the different requirements outlined by theoretical Computer science and Engineering.

Mapping of CO and PO:

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
C01	1	0	1	2	0	0	0	0	0	0	0	1
C02	2	1	1	2	1	0	0	0	0	0	0	1
C03	2	1	1	2	2	0	0	0	1	1	1	2
C04	2	3	2	2	2	0	0	0	1	1	1	2