The academic resources available in VIT –

VMIS (ERP)	V-Refer and V-Live	VIT Library	VAC & MOOC Courses
Institute & Department Vision and Mission	Former IA question papers and solutions (prepared by faculty)	Former IA question papers solutions - hardcopy	Value Added Courses (VAC) are conducted throughout the semester & in the semester break - Enrol for the VACs
Program Educational Objectives (PEO)	MU end semester examination question papers and solutions (prepared by faculty)	MU end semester exam question paper & solutions - by faculty, hardcopy	
Program Specific Outcome (PSO)	Class notes and Digital Content for the subject (scanned / typed by faculty)	All text books, reference books, e -books mentioned in the syllabus & AAP	Online courses from NPTEL, Coursera etc. are pursued throughout the semester - Register for the course & get certified
Program Outcome (PO)	Comprehensive question bank, EQ, GQ, PPT, Class Test papers	Technical journals and magazines for reference	
Departmental Knowledge Map	Academic Administration Plan & Beyond Syllabus Activity report	VIT library is member of IIT Bombay Library	Watch former lectures captured in LMS at VIT

1.a Course Objectives (Write in detail – as per NBA guidelines)

Cognitive	What do you want students to know?	Students should know the basic concepts of Natural Language Processing and algorithmic description of main language levels such as morphology, syntax, semantics, and pragmatics.
Affective	What do you want students to think / care about?	Implement various language Models.
Behavioural	What do you want students to be able to do?	Design and implement applications based on natural language processing.

Advice to Students:

Attend every class!!! Missing even one class can have a substantial effect on your ability to understand the course. Be prepared to think and concentrate, in the class and outside. I will try to make the class very interactive. Participate in the class discussions. Ask questions when you don't understand something. Keep up with the class readings. Start assignments and homework early. Meet me in office hour to discuss ideas, solutions or to check if, what you understand is correct.
The v-Refer Link http://vidyalankarlive.com/vrefer/index.php/apps/files/?dir=/vRefer/CMPN/SEM%20VII/2024-25/NLP/SJN&fileid=818550

Collaboration Policy:

We encourage discussion between students regarding the course material. However, no discussion of any sort is allowed with anyone on the assignment and homework for the class. If you find solution to some problems in a book or on the internet, you may use their idea for the solution; provided you acknowledge the source (name and page in the book or the website, if the idea is found on the internet). Even though you are allowed to use ideas from another source, you must write the solution in your own words. If you are unsure whether or not certain kinds of collaboration is possible please ask the teacher.

1.b **Course Outcome (CO) Statements and Module-Wise Mapping (follow NBA guideline)**

CO No.	Statements	Related Module/s
CO1	To describe the field of natural language processing.	1
CO2	To design language model for word level analysis for text processing.	2
CO3	To design various POS tagging techniques and parsers.	3
CO4	To design, implement and test algorithms for semantic and pragmatic analysis.	4
CO5	To formulate the discourse segmentation and anaphora resolution.	5
CO6	To apply NLP techniques to design real world NLP applications.	6

1.c **Mapping of COs with POs (mark S: Strong, M: Moderate, W: Weak, Dash '–': not mapped) (List of POs is available in V-refer)**

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12
CO 1	S	-	-	-	M	W	W	-	-	-	-	-
CO 2	M	S	S	M	M	-	-	-	-	-	-	-
CO 3	M	S	S	M	M	-	-	-	-	-	-	-
CO 4	W	S	S	S	M	-	-	-	-	-	-	-
CO 5	W	S	S	S	M	-	-	-	-	-	-	-
CO 6	M	S	S	-	S	-	-	W	S	M	M	M

1.d **Mapping of COs with PSOs (mark S: Strong, M: Moderate, W: Weak, Dash '–':not mapped)**

	PSO 1	PSO 2	PSO 3
CO 1	W	M	-
CO 2	S	W	-
CO 3	S	W	-
CO 4	S	W	-
CO 5	S	W	-
CO6	M	W	M

1.e Teaching and Examination Scheme (As specified by the University) for the Course

Categories	Humanities and Social Sciences	Basic Science	Engineering Science	Professional Core	General Education	Professional Elective	Project/ Internship	Open Elective
Tick suitable category						√		

Subject Code	Subject Name	Teaching Scheme			Credits Assigned			
		Theory	Practical	Tutorial	Theory	TW/Practical	Tutorial	Total
CSDC7013	Natural Language Processing	3	-	-	3	-	-	3
CSDL7013	Natural Language Processing Lab	-	2	-	-	1	-	1

Subject Code	Subject Name	Examination Scheme					
		Theory			Practical/Oral		Total
		ISA	MSE	ESE	ISA	ESE	
CSDC7013	Natural Language Processing	20	30	50	-	-	100
CSDL7013	Natural Language Processing Lab	-	-	-	25	-	25

1.f Faculty-Wise Distribution of all Lecture-Practical-Tutorial Hours for the Course

Divisions	Lecture (Hrs.)	Practical (Hrs.)				Tutorial (Hrs.)			
		Batch 1	Batch 2	Batch 3	Batch 4	Batch 1	Batch 2	Batch 3	Batch 4
A	SJN(3)	SJN(2)	UMK(2)	SJN(2)	SJN(2)	-	-	-	-
B	SJN(3)	UMK(2)	UMK(2)	SJN(2)	SJN(2)	-	-	-	-

1.g Office Hours (Faculty will be available in office in this duration for solving students' query)

Division	Day	Time (at least 1 Hr. / Division)	Venue (Office Room No.)
A	Friday	03:45 PM to 04:45 PM	M209
B	Tuesday	03:45 PM to 4:45 PM	M209

2.a **Syllabus : Module Wise Teaching Hours and % Weightage in University Question Paper**

Module No.	Module Title and Brief Details	Teaching Hrs. for each module	% Weightage in University Question Papers
1	Introduction to NLP Origin & History of NLP; Language, Knowledge and Grammar in language processing; Stages in NLP; Ambiguities and its types in English and Indian Regional Languages; Challenges of NLP; Applications of NLP Self-Learning topics: Variety of types of tools for regional languages pre-processing and other functionalities	03	08
Learning Outcome	Understanding the stages of the NLP model and different types of ambiguity and challenges.		
2	Word Level Analysis Basic Terms: Tokenization, Stemming, Lemmatization; Survey of English Morphology, Inflectional Morphology, Derivational Morphology; Regular expression with types; Morphological Models: Dictionary lookup, finite state morphology; Morphological parsing with FST (Finite State Transducer);Lexicon free FST Porter Stemmer algorithm; Grams and its variation: Bigram, Trigram; Simple (Unsmoothed) N-grams; N-gram Sensitivity to the Training Corpus; Unknown Words: Open versus closed vocabulary tasks; Evaluating N-grams: Perplexity; Smoothing: Laplace Smoothing, Good-Turing Discounting; Self-Learning topics: Noisy channel models, various edit distances, Advanced Issues in Language Modelling	9	23
Learning Outcome	To design the language model for word-level analysis		
3	Syntax analysis Part-Of-Speech tagging(POS); Tag set for English (Upenn Treebank); Difficulties /Challenges in POS tagging; Rule-based, Stochastic and Transformation-based tagging; Generative Model: Hidden Markov Model (HMM Viterbi) for POS tagging; Issues in HMM POS tagging; Discriminative Model: Maximum Entropy model, Conditional random Field (CRF);Parsers: Top down and Bottom up; Modelling constituency; Bottom Up Parser: CYK, PCFG (Probabilistic Context Free Grammar), Shift Reduce Parser; Top Down Parser: Early Parser, Predictive Parser Self-Learning topics: Evaluating parsers, Parsers based language modelling, Regional languages POS tree banks	10	26
Learning Outcome	To design POS tagging models and perform syntax analysis		
4	Semantic Analysis Introduction, meaning representation; Lexical Semantics; Corpus study; Study of Various language dictionaries like WorldNet, Babelnet; Relations among lexemes & their senses –Homonymy, Polysemy, Synonymy, Hyponymy; Semantic Ambiguity; Word Sense Disambiguation (WSD); Knowledge based approach(Lesk's Algorithm), Supervised (Naïve Bayes, Decision List),Introduction to Semi-supervised method (Yarowsky) Unsupervised (Hyperlex) Self-Learning topics: Dictionaries for regional languages,	07	18

	Distributional Semantics, Topic Models		
Learning Outcome	To analyze how various ML models can be used to resolve word sense disambiguity		
5	Pragmatic & Discourse Processing Discourse: Reference Resolution, Reference Phenomena, Syntactic & Semantic constraint on coherence; Anaphora Resolution using Hobbs and Cantering Algorithm Self-Learning topics: Discourse segmentation, Conference resolution	05	13
Learning Outcome	To understand how pragmatics and anaphora resolution works		
6	Applications of NLP Case studies on (preferable in regional language):Machine translation; Text Summarization; Sentiment analysis; Information retrieval; Question Answering system Self-Learning topics: Applications based on Deep Neural Network with NLP such as LSTM network, Recurrent Neural network etc.	05	13
Learning Outcome	To design real life applications based on NLP techniques.		
* Insert rows for more modules in the Course	**Total**	**39**	**100**

2.b **Prerequisite Courses**

No.	Semester	Name of the Course	Topic/s
1	4	Analysis of Algorithms	Stack
2	5	Theory of Computer Science	Grammar
3	6	System Programming and Compiler Construction	Compiler Construction

2.c **Relevance to Future Courses**

No.	Semester	Name of the Course
1	7	Project-I
2	8	Social Media Analytics
3	8	Project-II

2.d **Identify real life scenarios/examples which uses the knowledge of the subject (Discussion on how to prepare examples and case studies e.g. "Boeing Plane": C Programming Language – Intro to Computer Science – Harvard's CS50 (2018) – Bing video)**

Real Life Scenario	Concept Used
Sentiment Analysis	Tokenization, Stop word filtering, Stemming
Machine Translation	Morphological analysis, Lexical analysis
Question Answering	Question Analysis, Document retrieval, Answer extraction

3 Past Results – Division-Wise

Details	Target – Dec 2024	Dec 2023	Dec 2022	Dec 2021
Course Passing % – Average of 2 Divisions	100	100	100	NA
Marks Obtained by Course Topper (mark/100)	95/100	94/100	93/100	NA

	Division A		Division B	
Year	Initials of Teacher	% Result	Initials of Teacher	% Result
Dec 2023	SJN	100%	SJN	100%
Dec 2022	RSR	100%	SJN	100%
Dec 2021	NA	NA	NA	NA

4 All the Learning Resources – Books and E-Resources

4.a List of Text Books (T – Symbol for Text Books) to be Referred by Students

Sr. No	Text Book Titles	Author/s	Publisher	Edition	Module Nos.
1	Speech and Language Processing	Daniel Jurafsky, James H. Martin	Prentice Hall	2	All
2	Foundations of Statistical Natural Language Processing	Christopher D.Manning and Hinrich Schutze	MIT Press	-	3, 4

4.b List of Reference Books (R – Symbol for Reference Books) to be Referred by Students

Sr. No	Reference Book Titles	Author/s	Publisher	Edition	Module Nos.
1	Natural Language Processing and Information Retrieval	Siddiqui and Tiwary U.S.,	Oxford University Press	2008	1
2	Multilingual natural language processing applications	Daniel M Bikel and Imed Zitouni	Pearson	2013	2, 5, 6
3	The Handbook of Computational Linguistics and Natural Language Processing	Alexander Clark, Chris Fox, Shalom Lappin	Wiley	2010	6
4	Handbook of Natural Language Processing	Nitin Indurkhya and Fred J. Damerau	Chapman and Hall/CRC Press	2010	All
5	Natural Language Processing with Python	Steven Bird, Ewan Klein	O'Reily	-	Lab
6	A step by step tutorial: An introduction into R application and programming	Niel J le Roux, Sugnet Lubbe	-	-	Lab

4.c List of E - Books (E – Symbol for E-Books) to be Referred by Students

Sr. No	E- Book Titles	Author/s	Publisher	Edition	Module Nos.
1	Natural Language Processing Succinctly	Joseph D. Booth	Syncfusion	-	2, 3

4.d **Reading latest / top rated research papers (at least 5 papers)**

Name of Paper	Name of Authors (Background)	Published in		Problem Statement
		Date	Journal	
A Multitask Framework to Detect Depression, Sentiment and Multi-label Emotion from Suicide Notes.	Soumitra Ghosh, Asif Ekbal, Pushpak Bhattacharyya:	2022	Cognitive Computation, Springer	Depression plays a major role in increasing suicide ideation among the individuals. This paper focuses on learning three closely related tasks, viz. depression detection, sentiment citation, and to investigate their impact in analysing the mental state of the victims.
XL-WSD: An Extra-Large and Cross-Lingual Evaluation Framework for Word Sense Disambiguation.	Tommaso Pasini, Alessandro Raganato, Roberto Navigli	2021	Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press	Development of new approaches has been encouraged by a well-framed evaluation suite for English, which has allowed their performances to be kept track of and compared fairly. However, other languages have remained largely unexplored, as testing data are available for a few languages only. In this paper, authors propose XL-WSD, a cross-lingual evaluation benchmark for the WSD task featuring sense-annotated development and test sets in 18 languages from six different linguistic families, together with language-specific silver training data.
Complex Temporal Question Answering on Knowledge Graphs.	Zhen Jia, Soumajit Pramanik, Rishiraj Saha Roy, Gerhard Weikum	2021	Proceedings of the 30th ACM International Conference on Information & Knowledge Management	Answering complex temporal questions that have multiple entities and predicates, and associated temporal conditions is difficult. EXAQT answers natural language questions over KGs in two stages, one geared towards high recall, the other towards precision at top ranks.
ProofWriter: Generating Implications, Proofs, and Abductive	Oyvind Tafjord, Bhavana Dalvi, Peter Clark	2021	Findings of the Association for Computational Linguistics: {ACL/IJCNLP}	In this paper a model, called ProofWriter, generates both implications of a theory and the natural language proof(s) that support them. In particular, iterating

Statements over Natural Language.				a 1-step implication generator results in proofs that are highly reliable, and represent actual model decisions (rather than post-hoc rationalizations). Paper also shows that generative techniques can perform a type of abduction with high precision.
Semi-supervised Stance Detection of Tweets Via Distant Network Supervision.	Subhabrata Dutta, Samiya Caur, Soumen Chakrabarti, Tanmoy Chakraborty	2022	CoRR	Detecting and labeling stance in social media text is strongly motivated by hate speech detection, poll prediction, engagement forecasting, and concerted propaganda detection. Today's best neural stance detectors need large volumes of training data, which is difficult to curate. In this paper, authors present SANDS, a semi-supervised stance detector. SANDS starts from very few labeled tweets. It builds multiple deep feature views of tweets. It also uses a distant supervision signal from the social network to provide a surrogate loss signal to the component learners.

4.e **Based on research paper an identify the current Problem statement**

Problem Statement			Used in				
	Quiz	Assignment	Lab	Mini Project	Poster Presentation	Test	Any Other
Depression plays a major role in increasing suicide ideation among the individuals. This paper focuses on learning three closely related tasks, viz. depression detection, sentiment citation, and to investigate their impact in analysing the mental state of the victims.				√			
In the current era of social				√			

media, the popularity of smartphones and social media platforms has increased exponentially. Through these electronic media, fake news has been rising rapidly with the advent of new sources of information, which are highly unreliable. Checking off a particular news article is genuine or fake is not easy for any end user. Search engines like Google are also not capable of telling about the fakeness of any news article due to its restriction with limited query keywords. Design an efficient deep learning model to detect the degree of fakeness in a news statement.							

4.f **Identify Companies / Industries which use the knowledge of the subject and thus may provide Internships and final Placements**

Name of the Company	To be / Contacted for		
	Student Internship	Student Final Placement	Faculty Internship
TCS		√	

4.g **Identify suitable relevant TOP Guest Speakers from Industry (CS50 Lecture by Mark Zuckerberg - 7 December 2005 - YouTube)**

Name of the Identified Guest Speaker	Designation	Name of the Company
Abhishek Thakur	Chief Data Scientist	boost.ai

Rushabh Mishra	Software Engineer	Moveworks

4.h **Identify relevant Technical competitions to participate [Competitions -Paper Presentations, Projects, Hackathons, IVs etc..]**

Name of the Relevant Technical Competition Identified to participate	Organized by	Date of the Event
NLP Summit https://www.nlpsummit.org/nlp-summit-2024/	Johnsnowlabs	24-26 September, 2024

4.i **Identify faculty in TOP schools / Universities who are teaching same / similar subject and develop rapport e.g. Exchange Lecture Material (Assignments / Tests / Project etc..), Joint Paper Publication**

University	Name of the Course	Name of Faculty	Type of Collaboration		
			Exchange of Lecture Material	Joint Publication/ Research	Other
IIT Kharagpur	Natural Language Processing	Prof. Pawan Goyal	√		

4.j **Module Best Available in - Title best resource [from *4.a* to *4.d* in this AAP] & give details**

Module No.	Title of the Module	Web Link	Mention the Tile			
			Journal	E-Journal	Magazine	Other Resource
1	Introduction to NLP	https://www.atlantis-press.com/journals/nlpr https://www.quantamagazine.org/tag/natural-language-processing/	Natural Language Processing Research		WHAT'S UP IN natural language processing	
2	Word Level Analysis	https://www.atlantis-press.com/journals/nlpr https://www.quantamagazine.org/tag/natural-language-processing/	Natural Language Processing Research		WHAT'S UP IN natural language processing	
3	Syntax analysis	https://www.atlantis-press.com/journals/nlpr https://www.quantamagazine.org/tag/natural-language-processing/	Natural Language Processing Research		WHAT'S UP IN natural language processing	
4	Semantic Analysis	https://www.atlantis-press.com/journals/nlpr https://www.quantamagazine.org/tag/natural-language-processing/	Natural Language Processing Research		WHAT'S UP IN natural language processing	

5	Pragmatic & Discourse Processing	https://www.atlantis-press.com/journals/nlpr https://www.quantamagazine.org/tag/natural-language-processing/	Natural Language Processing Research		WHAT'S UP IN natural language processing	
6	Applications of NLP	https://www.atlantis-press.com/journals/nlpr https://www.quantamagazine.org/tag/natural-language-processing/	Natural Language Processing Research		WHAT'S UP IN natural language processing	

4.k **Referred to any top-rated university in that subject for content**

University	Name of the Course	Name of Faculty	Date of Delivery of the Course	Remarks
Stanford University	CS224n: Natural Language Processing with Deep Learning	Christopher D Manning	Winter 2024	

4.l **Faculty received any certification related to this subject. List of Certifications Identified / Done**

Course	Certifying Agency	No. of Hours	Level of the Course		Certification		Remarks
			Introductory	Advance Skill Development	Done on	Proposed to be on	
Natural Language Processing	Great Learning	2 months	√		November 2022	-	
Natural Language Processing with Projects	Great Learning	2 months		√	November 2023		

4.m **Completed subject wise/cluster wise training with cluster mentor.**
List of relevant Refresher Course Identified / Done

Course	Certifying Agency (As suggested by DAB/Cluster Mentor/Industry/ University other than MU)	Certification		Remarks
		Done on	Proposed to be on	
Pedagogy	Coursera	April 2020		https://www.coursera.org/learn/teach-online

PBL	Great Learning	November 2023		https://www.mygreatlearning.com/academy/learn-for-free/courses/natural-language-processing-projects

4.n Best Practices Identified and adopted

No.	Item	Best Practices Identified		
		Univ. 1	Univ. 2	Univ. 3
1	Microsite	https://www.cse.iitb.ac.in/~cs626-449/	https://www.mit.edu/~jda/teaching/6.864/	http://courses.washington.edu/ling567/
2	Video Lectures	√		
3	Assignments	√		
4	Mini Project		√	
5	Assessment Metric	√		
6	Quizzes			√
7	Labs/ Practical (PBL)		√	
8	Tests	√		
9	Peer Assessment		√	
10	Any Other			

4.o Web Links for Online Notes/YouTube/VIT Digital Content/VIT Lecture Capture/NPTEL Videos

Students can view lectures by VIT professors, captured through LMS 'Lecture Capture' in VIT campus for previous years.

No.	Websites / Links	Module Nos.
1	https://www.kaggle.com/c/word2vec-nlp-tutorial/	6
2	https://www.cse.iitb.ac.in/~cs626-449/	All

4.p Recommended MOOC Courses like Coursera / NPTEL / MIT-OCW / edX/VAC etc.

Sr. No.	MOOC Course Link	Course conducted by – Person / University / Institute / Industry	Course Duration	Certificate (Y / N)
1	https://www.kaggle.com/learn/natural-language-processing	Kaggle	Self-paced	Y
2	https://www.coursera.org/specializations/natural-language-processing	Deeplearning.ai	Self-paced	Y
3	https://www.coursera.org/projects/deep-learning-nlp-gpt-2	Coursera	Self-paced	Y

5 Consolidated Course Lesson Plan

	From (date/month/year)	From (date/month/year)	Total Number of Weeks
Semester Duration	8/7/2024	20/10/2024	14

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	COs Mapped	Recommended Prior Viewing / Reading	
						Lecture No. (on LMS)	Chapter No./ Books/ Web Site
1	1	1	Origin & History of NLP; Language, Knowledge and Grammar in language processing; Stages in NLP;		CO1	NA	R1/Chp1
	2, 3	1	Ambiguities and its types in English and Indian Regional Languages; Challenges of NLP; Applications of NLP		CO1	NA	R1/Chp1
2	4	2	Basic Terms: Tokenization, Stemming, Lemmatization; Survey of English Morphology, Inflectional Morphology, Derivational Morphology;		CO2	NA	T1/Chp 3/Pg.- 85-104
	5, 6	2	Regular expression with types; Morphological Models: Dictionary lookup, finite state morphology;		CO2	NA	T1/Chp 3/Pg.- 85-104 T1/Chp 3/Pg.- 47-76
3	7	2	Morphological parsing with FST (Finite State Transducer);		CO2	NA	T1/Chp 3/Pg.- 47-76
	8, 9	2	Lexicon free FST Porter Stemmer algorithm; Grams and its variation: Bigram, Trigram; Simple (Unsmoothed) N-grams;		CO2	NA	T1/Chp 2/Pg.- 219
4	10	2	N-gram Sensitivity to the Training Corpus; Unknown Words: Open versus closed vocabulary tasks;		CO2	NA	T1/Chp 2/Pg.- 246
	11, 12	2	Evaluating N-grams: Perplexity; Smoothing: Laplace Smoothing, Good-Turing Discounting;		CO2	NA	T1/Chp 2/Pg.- 246

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	COs Mapped	Recommended Prior Viewing / Reading	
						Lecture No. (on LMS)	Chapter No./ Books/ Web Site
5	13	3	Part-Of-Speech tagging(POS); Tag set for English (Upenn Treebank);		CO3	NA	T1/Chp 8/Pg.-315-334, 338-341
	14, 15	3	Difficulties /Challenges in POS tagging; Rule-based, Stochastic and Transformation-based tagging;		CO3	NA	T1/Chp 8/Pg.-315-334, 338-341
6	16	3	Generative Model: Hidden Markov Model (HMM Viterbi) for POS tagging;		CO3	NA	T2/Chp10,16 /Pg.-356, 589-596
	17, 18	3	Issues in HMM POS tagging; Discriminative Model:		CO3	NA	T2/Chp10,16 /Pg.-356, 589-596
7	19	3	Maximum Entropy model, Conditional random Field (CRF);Parsers: Top down and Bottom up; Modelling constituency;		CO3	NA	T2/Chp10,16 /Pg.-356, 589-596
	20, 21	3	Bottom Up Parser: CYK, PCFG (Probabilistic Context Free Grammar), Shift Reduce Parser;		CO3	NA	T2/Chp10,16 /Pg.-356, 589-596
8	22	3	Top Down Parser: Early Parser, Predictive Parser		CO3	NA	T2/Chp10,16 /Pg.-356, 589-596
	23, 24	4	Introduction, meaning representation; Lexical Semantics; Corpus study;		CO4	NA	T1/Chp 15,16,17/Pg. -571-594,618-630,662-671
9	25	4	Study of Various language dictionaries like WorldNet, Babelnet;		CO4	NA	T1/Chp 15,16,17/Pg. -571-

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	COs Mapped	Recommended Prior Viewing / Reading	
						Lecture No. (on LMS)	Chapter No./ Books/ Web Site
							594,618-630,662-671
	26, 27	4	Relations among lexemes & their senses Homonymy, Polysemy, Synonymy, Hyponymy; Semantic Ambiguity;		CO4	NA	T1/Chp 15,16,17/Pg.-571-594,618-630,662-671
10	28	4	Word Sense Disambiguation (WSD); Knowledge based approach (List),		CO4	NA	T1/Chp18/Pg.-697-706
	29, 30	4	Introduction to Semi-supervised method (Yarowsky), Unsupervised (Hyperlex)		CO4	NA	T1/Chp18/Pg.-697-706
11	31	5	Discourse: Reference Resolution, Reference Phenomena,		CO5	NA	T1/Chp18/Pg.-697-706
	32, 33	5	Syntactic & Semantic constraint on coherence;		CO5	NA	T1/Chp18/Pg.-697-706
12	34	5	Anaphora Resolution using Hobbs and Cantering Algorithm		CO5	NA	1/Chp18/Pg.-697-706
	35, 36	6	Case studies on (preferable in regional language):Machine translation;		CO6	NA	T1/Chp 21
13	37	6	Text Summarization; Sentiment analysis;		CO6	NA	R1/Chp 9
	38,39	6	Information retrieval; Question Answering system		CO6	NA	R1/Chp 11/ Chp 10
14			Revision				

Week	Lecture no.	Module No.	Lecture Topics / MSE / BSA planned to be covered	Actual date of Completion (Handwritten)	COs Mapped	Recommended Prior Viewing / Reading	
						Lecture No. (on LMS)	Chapter No./ Books/ Web Site
			Revision				

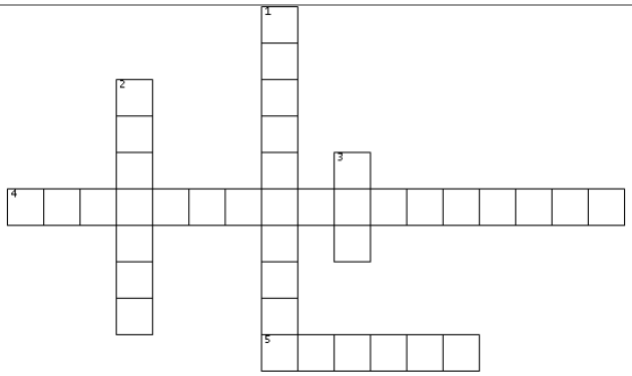
6 Rubric for Grading and Marking of Term Work (inform students at the beginning of semester)

ISA/TW	Activity	Total
ISA	11 Assignments will be given out of which best 10 will be graded	20 marks
TW	Lab work:10 marks Mini Project:10 marks Attendance :5 marks	25 marks

7 Assignments / Tutorials Details

Assignment / Tutorial No.	Title of the Assignments / Tutorials	CO Map	Assignment / Tutorials given to Students on	Week of Submission
1	What is the role of NLP in the following application? Explain anyone. Batch1: ChatGPT/Siri/Alexa Batch 2: QuillBot/Naukri.com/Resume Classifier Batch 3: Email Spam Filtering/Grammarly/Auto-correction Batch 4: Recommendation Engine of any E-commerce website/Opinion Mining Prepare a PPT for the same and present on 31st July 2024	CO1,6	2nd week	3rd
2	Compare Top 5 Python Libraries for Natural Language Processing with reference to its usage.	CO2	3rd	4th
3	PopQuiz-1	CO1,2	4th	5th
4	1. Design a Finite State Automata (FSA) for the words of English numbers 1-99. Write FSA for noun, verb, and adjective. 2. What is a language model? Write a note on evaluation metrics for the language model. 3. Evaluate the performance of the developed question answering system in terms of providing contextually relevant answers for Marathi educational content.	CO4	5th	6th

<table>
<tr>
<td>5</td>
<td>1.Compare different Stemming techniques- Porter Stemmer, Snowball Stemmer, Lancaster, Regexp Stemmer
2. Identify the unique challenges in sentiment analysis when applied to customer reviews in Bengali, compared to other languages.</td>
<td>CO2</td>
<td>6th</td>
<td>7th</td>
</tr>
<tr>
<td>6</td>
<td>Take Home Test
Based on the given state transition and emission probability matrix, assign POS to the statement: Time flies like an arrow.
Emission Probability Matrix
<table>
<tr><td></td><td>Time</td><td>flies</td><td>like</td><td>an</td><td>arrow</td></tr>
<tr><td>VB</td><td>0.1</td><td>0.2</td><td>0.2</td><td>0</td><td>0</td></tr>
<tr><td>NN</td><td>0.1</td><td>0.1</td><td>0</td><td>0</td><td>0.1</td></tr>
<tr><td>IN</td><td>0</td><td>0</td><td>0.25</td><td>0</td><td>0</td></tr>
<tr><td>DT</td><td>0</td><td>0</td><td>0</td><td>0.5</td><td>0</td></tr>
</table>
State Transition Matrix
<table>
<tr><td></td><td>VB</td><td>NN</td><td>IN</td><td>DT</td><td></S></td></tr>
<tr><td><S></td><td>0.2</td><td>0.8</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>VB</td><td>0</td><td>0.3</td><td>0.2</td><td>0.5</td><td>0</td></tr>
<tr><td>NN</td><td>0.4</td><td>0.5</td><td>0.1</td><td>0</td><td>0</td></tr>
<tr><td>IN</td><td>0</td><td>0.75</td><td>0</td><td>0.25</td><td>0</td></tr>
<tr><td>DT</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
</table>
</td>
<td>CO3</td>
<td>7th</td>
<td>8th</td>
</tr>
<tr>
<td>7</td>
<td>Discuss the applications of WordNet and BabelNet Lexical Database</td>
<td>CO5</td>
<td>8th</td>
<td>9th</td>
</tr>
<tr>
<td>8</td>
<td>Write a one-page summary report of any recent paper(2020 -2024) in the domain of NLP, published in Elsevier/Springer Journal</td>
<td>CO6</td>
<td>9th</td>
<td>10th</td>
</tr>
<tr>
<td>9</td>
<td>Crossword</td>
<td>CO1,2,3,4, 5</td>
<td>10th</td>
<td>11th</td>
</tr>
</table>

	 ACROSS 4. Probability based technique for Part of Speech tagging 5. study of the structural relationship between words DOWN 1. study of the relation between language and context -of-use 2. Lexical database for semantic analysis in NLP 3. Bottom up parsing technique			
10	Pop Quiz on Module-3,4,5,6	CO3.4.5.6	11th	12th
11	Discuss the role of Prompt engineering in Generative AI and ChatGPT	CO6	12th	13th

Analysis of Assignment / Tutorial Questions and Related Resources

Assignment / Tutorial No.	Week No.	Type* (√)			Module No.	Based on #			Question Type (√)	
		R	PQ	OBT		Text Book	Reference Book	Other Learning Resource	MU EQ	Thought Provoking
1	2		√		1,6	1			√	
2	3	√			2	1			√	
3	4		√		1,2		1			√
4	5	√			4		1		√	
5	6	√			2		1			√
6	7			√	3	1			√	
7	8			√	5	1				√
8	9			√	6	1				√
9	10		√		1,2,3,4,5		1			√
10	11	√			3,4,5,6		1		√	
11	12			√	6			1	√	

* Tick (√) the Type of the Assignment: Regular (R); Pop Quiz (PQ) ; Open Book Test for TE/BE/ME (OBT)
# Write number for text book, reference book, other learning resource from this AAP – *from Points 4.a to 4.d*

8 In Semester Assessment (ISE) / Other Class Test / Open Book Test (OBT)/Take Home Test (THT) Details

Tests	Test Dates	Module No.	CO Map	MSE Question Paper Pattern	Policy
MSE-1	TBD	1,2	CO1, CO2	Q1 – Short Questions - 10 Marks Q2 – 1 numerical 5 Marks Q3 – 1 numerical 5 Marks 30 marks each for MSE1, MSE2	Average of two MSE as MSE marks.
MSE-2	TBD	3,4	CO3,CO4		
Pop Quiz	4th week	1, 2	CO1, CO2	10 Question 1 mark each	One time submission
Pop Quiz	11th week	3,4,5,6	CO,4,5,6		

9.a Practical Activities

Practical No.	Module No.	Title of the **Experiments**	Type of Experiment		Topics to be highlighted	CO Map
			PBL	Newly Added		
1	1	Study various applications of NLP and Formulate the Problem Statement for Mini Project based on chosen real world NLP applications			Applications of NLP	CO6
2	1	Apply various text preprocessing techniques for any given text : Tokenization and Filtration & Script Validation.			Word Analysis	CO2
3	1	Apply various other text preprocessing techniques for any given text : Stop Word Removal, Lemmatization / Stemming.			Word Analysis	CO2
4	2	Perform morphological analysis and word generation for any given text.		√	Word Analysis	CO2
5	2	Implement Porter Stemmer Algorithm		√	Stemming	CO2
6	2	Implement N-Gram model for the given text input.			N Gram Model	CO2
7	3	To calculate emission and transition matrix for tagging Parts of Speech using Hidden Markov Model.			POS Tagging	CO3

8	6	Implement Named Entity Recognizer for the given text input.			Named Entity Recognition	CO4
9	3	Exploratory data analysis of a given text (Word Cloud)		√	Topic Modelling	CO3
10	All	Mini Project Report: For any one chosen real world NLP application. Implementation and Presentation of Mini Project	√		Applications of NLP	CO6
11	3	To import Brown Corpus, extract all the word tokens from the Science Fiction Category with their tags, and calculate a Frequency Distribution of the tags.	√		Text Measure	CO3

10 Beyond Syllabus Activities for Gap Mitigation

No.	Type of the Activity	Activities	Number of beneficiaries	Other Details – guest profile, feedback, mark sheet, report
1	**Experiential learning/Interaction with Outside World**	1- Guest Lectures by Industry Expert	150	Mr. Lakshmikant Tiwari, Lead Data Scientist, IGP.Com
		2- Workshops	150	Mr. Lakshmikant Tiwari, AVP,IGP.Com
		3- Mini Project	150	Group wise
		4- Industrial Visit	-	-
		5- Any other activity	-	-
2	**Collaborative & Group Activity**	6- Poster Presentation	150	Group wise
		7- Minute Papers	150	In class activity
		8- Students Seminars	150	Group wise
		9- Students Debates	-	-
		10- Panel Discussion / Mock GD	-	-
		11- Mock Interview	-	-
		12- Any other activity	-	-
3	**Co-Curricular Activity**	13- Informative videos (NPTEL/Youtube /TEDx/ MIT OW/edX)	150	NPTEL videos

		14- Lecture Capture Usage	-	-
		15- Any other activity	-	-
4	**Tests & Assessments**	16- Class Tests/ Weekly Tests	-	-
		17- Pop Quiz	150	4th,11th week
		18- Mobile App Based Quiz	-	-
		19- Open Book Test	-	-
		20- Take Home Test	150	7th week
		21- Any other activity	-	-

11 AAP/ Lecture Guide

No.	Programme	Course	Uploaded on V-refer	Date
1	CMPN	NLP	Yes	29/7/2024

*** Do not delete any activity. Give details for planned events. Write 'NA' for activity Not Planned.**

Consolidated Academic Administration Plan Prepared by (mention all theory teaching faculty names with signature)

Please write below your name and sign with date of the external cluster mentor meeting

Faculty 1	Faculty 2	Faculty 3

External Industry Mentor	External Academic Mentor	VIT Cluster Mentor	Program HOD