NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

COURSE PLAN – PART I			
Name of the programme and specialization	**B. Tech - CSE**		
Course Title	**Natural Language Processing**		
Course Code	**CSPE73**	**No. of Credits**	**3**
Course Code of Pre-requisite subject(s)		**Semester**	**VII**
Session	**July 2022**	**Section (if, applicable)**	**B**
Name of Faculty	**Dr. Chandramani Chaudhary**	**Department**	**CSE**
Official Email	**chandramani@nitt.edu**	**Telephone No.**	**NIL**
Name of Course Coordinator(s) (if, applicable)	**NIL**		
Official E-mail	**NIL**	**Telephone No.**	**NIL**
Course Type (please tick appropriately)	**Elective course**		

Syllabus (approved in BoS)

UNIT I Lexical Analysis
Lexical Analysis - Regular expression and Automata for string matching - Words and Word Forms - Morphology fundamentals - Morphological Diversity of Indian Languages - Morphology Paradigms - Finite State Machine / Transducers Based Morphology - Automatic Morphology Learning - Parts of Speech - N-gram Models - Hidden Markov Models.*****

UNIT II Speech Processing
Biology of Speech Processing - Place and Manner of Articulation - Word Boundary Detection - Argmax based computations - HMM and Speech Recognition - Text to Speech Synthesis - Rule based-Concatenative based approach.*

UNIT III Parsing
Theories of Parsing - Parsing Algorithms – Earley Parser - CYK Parser - Probabilistic Parsing - CYK - Resolving attachment and structural ambiguity - Shallow Parsing - Dependency Parsing - Named Entity Recognition - Maximum Entropy Models - Conditional Random Fields.*

UNIT IV Lexical Knowledge Networks
Meaning: Lexical Knowledge Networks - Wordnet Theory - Indian Language Wordnets and Multilingual Dictionaries - Semantic Roles - Word Sense Disambiguation - WSD and Multilinguality - Metaphors - Coreference and Anaphora Resolution.*

UNIT V Applications
Applications: Sentiment Analysis - Text Entailment - Machine Translation - Question Answering System - Information Retrieval - Information Extraction - Cross Lingual Information Retrieval (CLIR).* *Programming Assignments are mandatory.

Text Books
1. Jurafsky Daniel, Martin James, “Speech and Language Processing”, Second Edition, Tenth Impression, Pearson Education, 2018.

2. Christopher Manning, Schutze Heinrich, "Foundations of Statistical Natural Language Processing", MIT Press, 1999.

COURSE OBJECTIVES

- To understand the steps involved in Natural language processing
- To learn about the lexical, syntactic and semantic analysis of natural language processing
- To explore the various parsing techniques for natural languages
- To understand the statistical models for Natural language processing
- To learn about the various applications involved in Natural language processing

MAPPING OF COs with POs

Course Outcomes	**Programme Outcomes (PO)** (Enter Numbers only)
1. Justify the various steps necessary for processing natural language	1,3,5,6,9,11
2. Suggest appropriate lexical and parsing techniques for a given natural language	2,5,9.10
3. Apply appropriate statistical models for a given natural language application	1,3,6,12
4. Modify existing algorithms to suit any natural language for processing	2,3,5,9,11
5. Suggest appropriate pre-processing steps essential for the various applications involving natural language processing	1,2,5,6,9,12

COURSE PLAN – PART II

COURSE OVERVIEW

In this course, students will gain a thorough introduction to cutting-edge research in Deep Learning for NLP. Through lectures and assignments, students will learn the necessary skills to design, implement, and understand their own neural network models.

COURSE TEACHING AND LEARNING ACTIVITIES (Add more rows)

S.No.	**Week/Contact Hours**	**Topic**	**Mode of Delivery**
1	10/08/2022 to 12/08/2022 1 hours	Introduction to the course	Chalk and Talk with PPT Presentation
2	15/08/2022 to 19/08/2022 2 hours	**UNIT I Lexical Analysis** : Lexical Analysis	Chalk and Talk with PPT Presentation
3	22/08/2022 to 26/08/2022 3 hours	Regular expression and Automata for string matching - Words and Word Forms - Morphology fundamentals - Morphological Diversity of Indian Languages	Chalk and Talk with PPT Presentation
4	29/08/2022 to 2/09/2022 3 hours	Morphology Paradigms - Finite State Machine / Transducers Based Morphology - Automatic Morphology Learning - Parts of Speech	Chalk and Talk with PPT Presentation

5	5/09/2022 to 9/09/2022 3 hours	N-gram Models - Hidden Markov Models. **UNIT II Speech Processing** Biology of Speech Processing - Place and Manner of Articulation	Chalk and Talk with PPT Presentation
6	12/09/2022 to 16/09/2022 3 hours	Word Boundary Detection - Argmax based computations - HMM and Speech Recognition - Text to Speech Synthesis - Rule based-Concatenative based approach.	Chalk and Talk with PPT Presentation
7	19/09/2022 to 22/09/2022 2 hours	CT	
8	23/09/2022 1 hour	**UNIT III Parsing** Theories of Parsing - Parsing Algorithms – Earley Parser	Chalk and Talk with PPT Presentation
9	26/09/2022 to 30/09/2022 3 hours	CYK Parser - Probabilistic Parsing - CYK - Resolving attachment and structural ambiguity - Shallow Parsing - Dependency Parsing	Chalk and Talk with PPT Presentation
10	3/10/2022 to 7/10/2022 3 hours	Named Entity Recognition - Maximum Entropy Models - Conditional Random Fields.*	Chalk and Talk with PPT Presentation
11	10/10/2022 to 14/10/2022 3 hours	**UNIT IV Lexical Knowledge Networks** Meaning: Lexical Knowledge Networks - Wordnet Theory - Indian Language	Chalk and Talk with PPT Presentation
12	17/10/2022 to 20/10/2022 2 hours	**CT**	
13	21/10/2022 1 hour	Wordnets and Multilingual Dictionaries - Semantic Roles	Chalk and Talk with PPT Presentation
14	24/10/2022 to 28/10/2022 2 hours	Word Sense Disambiguation - WSD and Multilinguality - Metaphors - Coreference and Anaphora Resolution.*	Chalk and Talk with PPT Presentation
15	31/10/2022 to 4/11/2022 3 hours	**UNIT V Applications** Applications: Sentiment Analysis	Chalk and Talk with PPT Presentation
16	7/11/2022 to 11/11/2022 2 hours	Text Entailment	Chalk and Talk with PPT Presentation

17	14/11/2022 to 18/11/2022 3 hours	Machine Translation	Chalk and Talk with PPT Presentation
18	21/11/2022 to 25/11/2022 3 hours	Question Answering Systems	Chalk and Talk with PPT Presentation
19	28/11/2022 to 30/11/2022 2 hours	Information Retrieval - Information Extraction - Cross Lingual Information Retrieval (CLIR).*	Chalk and Talk with PPT Presentation

COURSE ASSESSMENT METHODS (shall range from 4 to 6**)**

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Cycle Test 1	19/09/2022 to 22/09/2022	1 hour	15
2	Cycle Test 2	17/10/2022 to 20/10/2022	1 hour	15
3	Assignments	5/09/2022 to 30/11/2022	4 hours	30
CPA	Compensation Assessment*	As per academic schedule	1 hour	15
4	Final Assessment *	As per academic schedule	3 hours	40

***mandatory; refer to guidelines on page 4**

COURSE EXIT SURVEY (mention the ways in which the feedback about the course shall be assessed)

1. Students' feedback through PAC meetings
2. Feedbacks are collected before final examination through MIS or any other standard format followed by the institute
3. Students, through their Class Representatives, may give their feedback at any time to the course faculty which will be duly addressed.

COURSE POLICY (including compensation assessment to be specified**)**

MODE OF CORRESPONDENCE (email/ phone etc)
Email and Phone

COMPENSATION ASSESSMENT POLICY
1. One compensation assessment will be given after completion of Cycle Test 1 and 2 for the students those who are absent for any assessment due to genuine reason.
2. Compensatory assessments would cover the syllabus of Cycle tests 1 & 2
3. The prior permission and required documents must be submitted for absence signed by HoD/CSE.

ATTENDANCE POLICY (A uniform attendance policy as specified below shall be followed)

- At least 75% attendance in each course is mandatory.
- A maximum of 10% shall be allowed under On Duty (OD) category.
- Students with less than 65% of attendance shall be prevented from writing the final assessment and shall be awarded 'V' grade.

ACADEMIC DISHONESTY & PLAGIARISM

- Possessing a mobile phone, carrying bits of paper, talking to other students, copying from others during an assessment will be treated as punishable dishonesty.
- Zero mark to be awarded for the offenders. For copying from another student, both students get the same penalty of zero mark.
- The departmental disciplinary committee including the course faculty member, PAC chairperson and the HoD, as members shall verify the facts of the malpractice and award the punishment if the student is found guilty. The report shall be submitted to the Academic office.
- The above policy against academic dishonesty shall be applicable for all the programmes.

ADDITIONAL INFORMATION, IF ANY

1. The Course Coordinator is available for consultation during the time intimated to the students
2. Relative grading adhering to the instructions from the office of the dean (Academic) will be adopted for the course.

FOR APPROVAL

Course Faculty **CC- Chairperson** **HOD**

Guidelines

a) The number of assessments for any theory course shall range from 4 to 6.

b) Every theory course shall have a final assessment on the entire syllabus with at least 30% weightage.

c) One compensation assessment for absentees in assessments (other than final assessment) is mandatory. Only genuine cases of absence shall be considered.

d) The passing minimum shall be as per the regulations.

B.Tech. Admitted in				P.G.
2018	2017	2016	2015	
35% or (Class average/2) whichever is greater.		(Peak/3) or (Class Average/2) whichever is lower		40%

e) Attendance policy and the policy on academic dishonesty & plagiarism by students are uniform for all the courses.

f) Absolute grading policy shall be incorporated if the number of students per course is less than 10.

g) Necessary care shall be taken to ensure that the course plan is reasonable and is objective.