DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

COURSE PLAN – PART I			
Course Title	NATURAL LANGUAGE PROCESSING		
Course Code	CSPE73	No. of Credits	03
Course Code of Pre-requisite subject(s)	CSPC62	Semester	VII
Session	JULY 2022	Section (if, applicable)	A
Name of Faculty	Dr. A. Santhanavijayan	Department	CSE
Email	vijayana@nitt.edu	Telephone No.	0431 - 2503217
Name of Course Coordinator(s) (if, applicable)	———		
E-mail	———	Telephone No.	———
Course Type	☐ Core course	☑ 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.

Reference Books

1. Allen James, "Natural Language Understanding", Second Edition, Benjamin Cumming,

   1995.
2. Charniack Eugene, "Statistical Language Learning", MIT Press, 1993.

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.

COURSE OUTCOMES (CO)

Upon completion of this course, the students will be able to:

• Justify the various steps necessary for processing natural language.

• Suggest appropriate lexical and parsing techniques for a given natural language.

• Apply appropriate statistical models for a given natural language application.

• Modify existing algorithms to suit any natural language for processing.

• Suggest appropriate pre-processing steps essential for the various applications involving natural language processing.

Mapping of Course Outcomes with Programme Outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	√		√		√	√			√		√	
CO2		√			√				√	√		
CO3	√		√			√						√
CO4		√	√		√				√		√	
CO5	√	√			√	√			√			√

COURSE PLAN – PART II

COURSE OVERVIEW

This course mainly describes about the Lexical Analysis, syntax, semantics, speech processing and various applications of natural language processing.

COURSE TEACHING AND LEARNING ACTIVITIES

S.No.	Week/Contact Hours	Topic	Mode of Delivery
1	1/3	Lexical Analysis - Regular expression and Automata for string matching - Words and Word Forms	Chalk and Talk
2	2/3	Morphology fundamentals - Morphological Diversity of Indian Languages - Morphology Paradigms - Finite State Machine /Transducers Based Morphology	Chalk and Talk
3	3/3	Automatic Morphology Learning - Parts of Speech - N-gram Models – Hidden Markov Models.*	Chalk and Talk

4	4/3	Biology of Speech Processing - Place and Manner of Articulation - Word Boundary Detection - Argmax based computations	Chalk and Talk
5	5/3	HMM and Speech Recognition - Text to Speech Synthesis - Rule based-Concatenative based approach.*	Chalk and Talk
6	6/3	Theories of Parsing - Parsing Algorithms – Earley Parser - CYK Parser - Probabilistic Parsing	Chalk and Talk, PPT
7	7/3	CYK – Resolving attachment and structural ambiguity - Shallow Parsing -	Chalk and Talk
8	8/3	Dependency Parsing - Named Entity Recognition - Maximum Entropy Models - Conditional Random Fields.*	Chalk and Talk
9	9/3	Meaning: Lexical Knowledge Networks - Wordnet Theory	Chalk and Talk , PPT
10	10/3	Indian Language Wordnets and Multilingual Dictionaries - Semantic Roles	Chalk and Talk
11	11/3	Word Sense Disambiguation - WSD and Multilinguality - Metaphors - Coreference and Anaphora Resolution.*	Chalk and Talk
12	12/3	Applications: Sentiment Analysis - Text Entailment - Machine Translation - Question Answering System	Chalk and Talk , PPT
13	13/3	Information Retrieval - Information Extraction - Cross Lingual Information Retrieval (CLIR).*	Chalk and Talk , PPT

COURSE ASSESSMENT METHODS

S.No.	Mode of Assessment	Week/Date	Duration	% Weightage
1	Assessment 1 (written Test)	SEP'22 3rd week	1 hour	20
2	Assessment 2 (written Test)	OCT'22 3rd week	1 hour	20
3	Assignment	AUG'22 3rd week	——	10
CPA	Compensation Assessment	NOV'22 4th week	1 hour	20
4	Final Assessment (written Test)	DEC'22 1st week	3 hours	50
			TOTAL	100

COURSE EXIT SURVEY

- Feed backs are collected after the end semester exam in the feedback forms.
- Suggestions from the students are incorporated for making the course more sympathetic and motivating.
- Students may give their feedback at any time, through their class representatives to the concerned faculty and also in the class committee meetings.

COURSE POLICY

MODE OF CORRESPONDENCE (email/ phone etc)
- Both email and phone

COMPENSATION ASSESSMENT

Compensation assessment (Retest) will be conducted for absentees in assessment 1 or assessment 2 only after the submission of medical or On-Duty certificates signed by competent authority. The portions for compensation assessment will be the portions of assessment 1 and assessment 2.

ATTENDANCE POLICY

- 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.

ADDITIONAL INFORMATION

The students can clarify their doubts at any time during working hours from the faculty with prior appointment.

FOR APPROVAL

Course Faculty	CC-Chairperson	HOD

(Dr. A. SANTHANAVIJAYAN)   (Dr.R. MOHAN)   (Dr.S. MARY SAIRA BHANU)