

FR. Conceicao Rodrigues College Of Engineering

Father Agnel Ashram, Bandstand, Bandra-west, Mumbai-50

Department of Electronics and Computer Science

Course Plan

B.E. (ECS) (Semester VIII) (2022-23)

Subject name: Natural Language Processing

Subject code: ECCDLO8012

Teacher-in-charge: Dipali Koshti

Academic Term: Jan 2023- June 2023

Module No.	Unit No.	Topics	Hrs.
1.0		Introduction to Natural Language Processing	06
	1.1	The need of NLP. Generic NLP system, Levels of NLP	02
	1.2	Stages in building a Natural Language Processing System. Challenges and ambiguities in NLP Design	04
2.0		Mathematical and Linguistic Preliminaries	06
	2.1	Probability Theory, Conditional Probability and Independence, Bayes Rule, Random Variables, Probability Distributions, Statistics, Counting, Frequency, Mean and Variance	04
	2.2	English Grammar, Parts of Speech, Phrase Structures	02
3.0		Word Level Analysis	06
	3.1	Tokenization, Segmentation, Lemmatization, Edit Distance, Collocations, Porter Stemmer, N-gram Language Model	04
	3.2	Morphological Analysis, Derivational and Reflectional Morphology	02
4.0		Syntax-Analysis	08
	4.1	Tag set for English, Penn Tree bank, Introduction to Parts of Speech Tagging (POST)	02
	4.2	Markov Processes, Hidden Markov Models (HMM)	02
	4.2	Parts of Speech Tagging using Hidden Markov Models, Viterbi Algorithm	04
5.0		Semantic Analysis	08
	4.1	Lexical Semantics, ambiguous words, word senses, Relations between senses: synonym, antonym, reversives, hyponym, hypernym, meronym, structured polysemy, metonymy, zeugma	04
	4.2	Introduction to WordNet, gloss, synset, sense relations in WordNet. Cosine distance between documents. Word sense disambiguation.	04

6.0		Pragmatics and applications of NLP	05
	6.1	Reference resolution: Discourse model, Reference Phenomenon, Syntactic and Semantic Constraints on co reference	03
	6.2	Applications of NLP: Categorization, Summarization, Sentiment Analysis, Named Entity Recognition, Machine Translation, Information Retrieval, Question Answer System	02
		Total	39

Text Books:

1. Daniel Jurafsky, James H. Martin, Speech and Language Processing| Second Edition, Prentice Hall.
2. Christopher D. Manning & Hinrich Schutze, Foundations of Statistical Natural Language Processing, MIT Press.

Reference Books:

1. Steven Bird, Ewan Klein, Natural Language Processing with Python, O'Reilly
2. Alexander Clark (Editor), Chris Fox (Editor), Shalom Lappin (Editor), The Handbook of Computational Linguistics and Natural Language Processing

Course Objectives:

1. To understand natural language processing and to learn how to apply basic algorithms in this field.
2. To get acquainted with the basic concepts and algorithmic description of the main language levels: morphology, syntax, semantics, and pragmatics.
3. To design and implement applications based on natural language processing

Course Outcomes:

At the end of the course student will be able to

ECCDLO8012.1: Describe the mathematical and linguistic preliminaries necessary for various processes in NLP.

ECCDLO8012.2: Perform Word-Level and Syntax-Level analysis on a text.

ECCDLO8012.3: Analyze the text at Semantic Level.

ECCDLO8012.4: Develop a basic understanding of Pragmatics in NLP

ECCDLO8012.5: Apply NLP techniques to design real-world NLP applications

BL Wise Description

Co No.	Course Outcome	BL	Target
ECCDLO8012.1	Understand the mathematical and linguistic preliminaries necessary for various processes in NLP.	2	2.5
ECCDLO8012.2	Perform Word-Level and Syntax-Level analysis on a text.	4	2.5
ECCDLO8012.3	Analyze the text at Semantic Level	4	2.5
ECCDLO8012.4	Develop a basic understanding of Pragmatics in NLP	2	2.5
ECCDLO8012.5	Apply NLP techniques to design real-world NLP applications	4	2.5

CO-PO-PSO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3	2												
CO2	2	3	2	2										
CO3	2	3	2	2										
CO4	2	3	2	2										
CO5	2	3	2	3										3

CO-PO Mapping:(BL – Blooms Taxonomy, C – Competency, PI – Performance Indicator)

CO	BL	C	PI	PO
ECCDLO8012.1	2	1.1	1.1.2	PO1
		1.4	1.4.1	
		2.1	2.1.3	PO2
ECCDLO8012.2	4	1.1	1.1.2	PO1
		1.4	1.4.1	
		2.1	2.1.3	PO2
		2.2	2.2.3 2.2.4	
		3.2	3.2.1 3.2.2	PO3
ECCDLO8012.3	4	1.1	1.1.2	PO1
		1.4	1.4.1	
		2.1	2.1.3	PO2
		2.2	2.2.3 2.2.4	
		3.2	3.2.1 3.2.2	PO3
ECCDLO8012.4	2	1.1	1.1.2	PO1
		1.4	1.4.1	
		2.2	2.2.3 2.2.4	PO2
		3.2	3.2.1 3.2.2	PO3
		4.2 4.3	4.2.1 4.3.2	PO4
ECCDLO8012.5	4	1.1	1.1.2	PO1
		1.4	1.4.1	

		2.1 2.2 2.4	2.1.3 2.2.3 2.2.4 2.4.1 2.4.3	PO2
		3.2	3.2.1 3.2.2	PO3
		4.2 4.3	4.2.1 4.3.2	PO4

CO Assessment Tools:

Course Outcome	Assessment Method									
	Direct Method (80 %)									Indirect Method (20%)
	Unit Tests		Assignments			Quiz		Technical paper presentation	University results (Th)	Course exit survey
	1	2	1	2	3	1	2		THEORY	
ECCDLO8012.1	20	-	-	-	-	20	-	-	6	100
ECCDLO8012.2	10	10	20		-	-	-	-	60	100
ECCDLO8012.3	-	20	-	20	-	-	-	-	60	100
ECCDLO8012.4	-	20	-	-	-	-	20	-	60	100
ECCDLO8012.5	-	20	-	-	-	-	-	20	60	100

CO calculation= (0.8 *Direct method + 0.2*Indirect method)

Rubrics for assessing Course Outcome with each assessment tool:

Rubrics for Assignment:

Indicator	Poor	Average	Good	Excellent
Timeline (2)	More than two session late (0.5)	Two sessions late (1)	One session late (1.5)	Early or on time (2)
Organization (2)	poor readability and not structured (0.5)	Poor readability but somewhat structured (1)	Readable with one or two mistakes and structured (1.5)	Very well written and structured without any mistakes (2)
Level of content (4)	Major points are omitted or addressed minimally (1)	All major topics are covered, the information is accurate.(2)	Most major and some minor criteria are included. Information is Accurate (3)	All major and minor criteria are covered and are accurate. (4)
Depth and breadth discussion (2)	None in evidence; superficial at most (0.5)	Minor points/information may be missing and discussion is minimal (1)	Discussion centers on some of the points and covers them adequately (1.5)	Information is presented in depth and is accurate (2)

Rubrics Mini project evaluation: (20M)

Indicator	Poor	Average	Good	Excellent
Timeline -Maintains project deadline (2)	More than two session late (0.5)	Two sessions late (1)	One session late (1.5)	Early or on time (2)
Complexity of the chosen problem (4)	Simple (1)	Moderate (2)	Complex(3)	Too Complex(4)
Completeness (6)	< 40% complete (1)	~ 60% complete (2)	~ 80% complete (3-4)	100% complete (5-6)
Project-specific Technical Features (4)	60-65% of features (1)	65-70%of features (2)	70-80% of features(3)	Most of the features taught(4)
Project Report (4)	Poor organization, Major content missing, Not as per guidelines.	Good organization , Few of the project aspects missing (2)	Well organized, Major aspects of the project covered, as per guide lines (3)	Very well organized , covering major and minute details of the project , as per guidelines (4)

Rubrics for Research paper presentation:

Rubrics	Poor	Average	Good	Excellent
Time Line (2)	Not presented (0)	One session late(1)	Two sessions late(1.5)	On-time(2)
Presentation Skills (2)	Does not seem to be prepared, and does not speak clearly (0)	Somewhat prepared (1)	Overall Prepared, and speaks clearly (1.5)	Well prepared, Enthusiastic, excellent communication(2)
Depth and breadth of Discussion (4)	None in evidence; superficial at most (0.5)	Minor points/information may be missing and discussion is minimal (1)	Discussion centers on some of the points and covers them adequately (1.5)	Information is presented in depth and is accurate (2)
Level of content (4)	Major points are omitted or addressed minimally (1)	All major topics are covered, the information is accurate.(2)	Most major and some minor criteria are included. Information is Accurate (3)	All major and minor criteria are covered and are accurate. (4)

Content beyond syllabus:

1. Study five latest research papers from reputed journals/conferences related to advanced NLP applications and present the findings and summary in front of the class. This activity will be a group activity and help students to explore the research work already done in the field of NLP.
2. Natural language processing using Transformers: Transformers have changed the way we process longer text. Nowadays, most of the NLP applications such as Alexa use Transformers for encoding and decoding text. To bridge this gap, industry expert from the relevant field would be invited to deliver a guest session.

3. Debate activity: The topics covered in the debate will address societal issues related to NLP and AI.

Curriculum gap: No gap

Modes of content delivery

Modes of Delivery	Brief description of content delivered	Attained CO	Attained PO
Classroom lecture, PPT	Introduction to NLP, Mathematical and Linguistic Preliminaries Word-level analysis Syntax Analysis Semantic analysis Pragmatics and Applications of NLP	-	-
Assignments	Assignment 1: Covering the word level and syntax level analysis Assignment 2: Covering Semantic Level analysis	-	-
Quizzes	Quiz1: covering mathematical and linguistic preliminaries Quiz 2: covering pragmatics	-	-
Review and present technical papers	Covering Module 6 (Applications of NLP)	-	2,4,9,10,12
Informative videos	Video 1: Minimum edit Distance: Minimum Edit distance (Dynamic Programming) for converting one string to another string - YouTube Video 2: collocations: Lecture - 13 - Collocations in NLP - YouTube Video 3: N-gram Model : NPTEL: Co-occurrence matrix, n-grams - YouTube Video 4: Viterbi Algorithm for POST: Part of Speech (POS) Tagging, Viterbi Algorithm, Solved Problem, Natural Language Processing - YouTube Natural Language Processing (Prof. Pawan Goyal, IIT Kharagpur): Lecture 16 - Viterbi Decoding for Hidden Markov Models, Parameter Learning (infocobuild.com)	-	-
Group Activity (Debate Competition)	Advanced topics – address societal issues related to NLP and AI (Content beyond syllabus)	--	3,6,7,8,9,10, 13

Examination Scheme

Module		Lecture Hours	Marks distribution in Test (For internal assessment/TW)		Approximate Marks distribution in Sem. End Examination
			Test 1	Test 2	
1	Introduction	06	05	-	-
2	Mathematical and Linguistic Preliminaries	06	05	-	-
3	Word – level analysis	06	05	-	-
4	Syntax level analysis	08	05	-	-
5	Semantic analysis	08	-	06	-
6	Pragmatics and applications	05	-	14	-

Syllabus Scheme

Course Code	Course Name	Teaching Scheme (Contact Hours)			Credits Assigned			
		Theory	Practical	Tutorial	Theory	Practical	Tutorial	Total
ECCDO 801	Natural Language Processing	03	--	--	03	--	--	03

Course Code	Course Name	Examination Scheme								
		Theory Marks					Exam Duration (Hrs.)	Term Work	Practical and Oral	Total
		Internal Assessment			End Sem. Exam.					
		Test1	Test2	Avg.						
ECCDO 801	Natural Language Processing	20	20	20	80	03	--	--	100	

Lesson Plan
(Academic Year 2022-23)

Lecture No.	Dates		Topic	Content delivery	References	Remarks
	Planned	Actual				
Module 1: Introduction						
1	10/1/2023		The need of NLP, a Generic NLP system	Classroom teaching, PPT	1,2	
2	11/1/2023		Levels of NLP	Classroom teaching, PPT	1,2	
3	12/1/2023		Stages in building a Natural Language Processing System.	Classroom teaching, PPT	1,2	
4	17/1/2023		Challenges and ambiguities in NLP	Classroom teaching, PPT	1,2	
Module 2: Mathematical and Linguistic Preliminaries						
5	18/1/2023		Probability Theory	Classroom teaching, PPT	1,2	
6	19/1/2023		Conditional Probability and Independence, Bayes Rule	Classroom teaching, PPT	1,2	
7	24/1/2023		Random Variables, Probability Distributions,	Classroom teaching, PPT	1,2	
8	25/1/2023		Statistics, Counting, Frequency, Mean and Variance	Classroom teaching, PPT	1,2	
9	31/1/2023		English Grammar, Parts of Speech,	Classroom teaching, PPT	1,2	
10	1 / 2/2023		Phrase Structures	Classroom teaching, PPT	1,2	Quiz1
Module 3: Word Level Analysis						
11	2/2/2023		tokenization, Segmentation	Classroom teaching, PPT	1,2	
12	3/2/2023		Lemmatization	Classroom teaching, PPT	1,2	
13	7/2/2023		Edit Distance, Collocations	PPT, Video	1,2,8,9,	
14	8/2/2023		Porter Stemmer	PPT	1,2,10	
15	9/2/2023		N-gram Language Model	PPT, Video	1,2	
16	14/2/2023		Morphological Analysis, Derivational Morphology	Classroom teaching, PPT	1,2	
17	15/2/2023		and Reflectional Morphology	Classroom teaching, PPT	1,2	Assignment1
Module 4: Syntax-Analysis:						
18	16/2/2023		Tag set for English, Penn Tree bank,	Classroom teaching, PPT	1,2	
19	21/2/2023		Introduction to Parts of Speech Tagging (POST)	Classroom teaching, PPT	1,2	
20	22/2/2023		Markov Processes	Classroom teaching, PPT	1,2	
21	23/2/2023		Hidden Markov Models (HMM)	Classroom teaching, PPT	1,2	

22	1/3/2023		Parts of Speech Tagging using Hidden Markov Models,	Classroom teaching, PPT	1,2	
23	2/3/2023		Viterbi Algorithm	PPT,Video	1,2	
24	7/3/2023		Problems based on Viterbi Algorithm	Classroom teaching, PPT	1,2	Assignment1
Module 5: Semantic Analysis						
25	9/3/2023		Lexical Semantics, ambiguous words	Classroom teaching, PPT	1,2	
26	14/3/2023		word senses	Classroom teaching, PPT	1,2	
27	15/3/2023		Introduction to WordNet, gloss, synset, sense relations in WordNet. Cosine distance between documents. Word sense disambiguation	Classroom teaching, PPT	1,2	
28	16/3/2023		Relations between senses: structured polysemy, metonymy, zeugma	Classroom teaching, PPT	1,2	
29	21/3/2023		Introduction to WordNet	Classroom teaching, PPT	1,2	
30	23/3/2023		gloss, synset, sense relations in WordNet.	Classroom teaching, PPT	1,2	
31	5/4/2023		Cosine distance between documents.	Classroom teaching, PPT	1,2	
32	6/4/2023		Word sense disambiguation	Classroom teaching, PPT	1,2	Assignment2
Module 6: Pragmatics and applications of NLP						
33	11/4/2023		Reference resolution: Discourse model, Reference Phenomenon	Classroom teaching, PPT	1,2	
34	12/4/2023		Syntactic and Semantic Constraints on co reference	Classroom teaching, PPT	1,2	
35	13/4/2023		Applications of NLP: Categorization, Summarization	PPT/Research paper		
36	18/4/2023		Sentiment Analysis, Named Entity Recognition,	PPT/ research paper		
37	19/4/2023		Machine Translation, Information Retrieval	PPT/ research paper		Quiz2
38	20/4/2023		Question Answer System	PPT/ Research paper		

Text Books:

1. Daniel Jurafsky, James H. Martin, Speech and Language Processing || Second Edition, Prentice Hall.
2. Christopher D. Manning & Hinrich Schutze, Foundations of Statistical Natural Language Processing, MIT Press.

Reference Books:

3. Steven Bird, Ewan Klein, Natural Language Processing with Python, O 'Reilly
4. Alexander Clark (Editor), Chris Fox (Editor), Shalom Lappin (Editor), The Handbook of Computational Linguistics and Natural Language Processing.

Online Resources:

5. Course: Natural Language Processing By Prof. Pawan Goyal, IIT Kharagpur
https://onlinecourses.nptel.ac.in/noc21_cs102/preview

6. Course: Applied Natural Language Processing By Prof. Ramaseshan R, CMI
https://onlinecourses.nptel.ac.in/noc20_cs87/preview
7. IIT Vlabs: Welcome to Virtual Labs - A MHRD Govt of India Initiative (vlabs.ac.in)
8. Word Collocations in Natural Language Processing – CodeSpeedy
9. Collocations — identifying phrases that act like single words in Natural Language Processing | by Nicha Ruchirawat | Medium
10. Porter Stemmer - Porter Stemming Algorithm – Basic Intro | Vijini Mallawaarachchi

Videos:

Video 1: Minimum edit Distance: [Minimum Edit distance \(Dynamic Programming\) for converting one string to another string - YouTube](#)

Video 2: collocations: [Lecture - 13 - Collocations in NLP - YouTube](#)

Video 3: N-gram Model : NPTEL: [Co-occurrence matrix, n-grams - YouTube](#)

Video 4: Viterbi Algorithm for POST: [Part of Speech \(POS\) Tagging, Viterbi Algorithm, Solved Problem, Natural Language Processing - YouTube](#)

[Natural Language Processing \(Prof. Pawan Goyal, IIT Kharagpur\): Lecture 16 - Viterbi Decoding for Hidden Markov Models, Parameter Learning \(infocobuild.com\)](#)

Technical papers:

[1] A Review towards the Sentiment Analysis Techniques for the Analysis of Twitter Data by Priyanka Tyagi, R.C. Tripathi :: SSRN (2019)

[2] Recent Trends in Named Entity Recognition (NER). (2021)

[3] Recent Trends in Deep Learning Based Open-Domain Textual Question Answering Systems, IEE access (2020)

Submitted By	Approved By
Sign:	Sign:
Date of Submission:	Date of Approval:
Remarks by PAC (if any):	